The Co-Evolution of Capabilities and Transaction Costs: Explaining the Institutional Structure of Production

Michael G. Jacobides

Sidney G. Winter
University of Pennsylvania

The Co-Evolution of Capabilities and Transaction Costs: Explaining the Institutional Structure of Production

Abstract

This paper proposes that transaction costs and capabilities are fundamentally intertwined in the determination of vertical scope, and identifies the key mechanisms of their co-evolution. Specifically, we argue that capability differences are a necessary condition for vertical specialization; and that transaction cost reductions only lead to specialization when capabilities along the value chain are heterogeneous. Furthermore, we argue that there are four evolutionary mechanisms that shape vertical scope over time. First, the selection process, itself driven by capability differences, dynamically shapes vertical scope; second, transaction costs are endogenously changed by firms that try to reshape the transactional environment to increase their profit and market share; third, changes in vertical scope affect the nature of the capability development process, i.e., the way in which firms improve their operations over time; and finally, the changes in the capability development process reshape the capability pool in the industry, changing the roster of qualified participants. These dynamics of capability and transaction cost co-evolution are illustrated through two contrasting examples: the mortgage banking industry in the United States, which shows the shift from integrated to disintegrated production; and the Swiss watch-manufacturing industry, which went from disintegration to integration.

Keywords

capabilities, transaction costs, evolution, institutional structure of production

Disciplines

Business Administration, Management, and Operations

Capabilities, Transaction Costs, and Evolution:

Understanding the Institutional Structure of Production

Michael G. Jacobides
Assistant Professor of Strategic & International Management
London Business School
Sussex Place, Regent's Park, London NW1 4SA, United Kingdom
mjacobides@london.edu; Tel (+ 44 20) 7706 6725; fax (+ 44 20) 7724 7875

Sidney G. Winter
Deloitte and Touche Professor of Management
The Wharton School of the University of Pennsylvania
Suite 2000, SH-DH, 3620 Locust Walk, Philadelphia PA 19104, USA
Winter@wharton.upenn.edu; Tel (215) 898 4140; fax (215) 898 4041

Working Paper

July 2003

CAPABILITIES, TRANSACTION COSTS, AND EVOLUTION: UNDERSTANDING THE INSTITUTIONAL STRUCTURE OF PRODUCTION

Abstract

Analyzing the "institutional structure of production" as proposed by Coase requires a dynamic analysis of the business system. We focus on the evolution of vertical scope, and propose a framework where capability differences interact with changing transaction costs. Such aggregate forces shape the choice context of individual firms. Specifically, dis-integration tends to happen when firms differ in relative competence along the value chain, creating latent gains from trade that then motivate efforts to reduce transaction costs. Changes in scope, in turn, shape the process of capability formation and productivity improvement. These insights are partly formalized in an extension of Fisher's Selection Theorem.

Over the last twenty years, much progress has been made in the analysis of vertical scope, and in understanding what drives the governance structures observed in practice. A key figure in that development was Oliver Williamson (1975, 1985, 1999) who elaborated and, crucially, operationalized the concept of transaction costs, initially formulated by Coase (1937). This research has focused on a particular strand of the Coasean inquiry, examining the conditions under which firms choose to abandon markets in favor of integration. The potential for hold-ups and opportunistic behavior, this theory suggests, is the main determinant of vertical scope.

Independently, another stream of literature has come to have a defining impact on strategy as a field: the resource- and capability- based view of the firm. This approach, which has its roots in Penrose (1959) and more recently Wernerfelt (1984) and Barney (1991) emphasizes the importance of resources in guiding firm action, and the management of a firm's resource and capability portfolio as the basic principle in strategy. Of late, this research has used principles suggested by evolutionary economists (Nelson and Winter, 1982) and the focus has shifted to dynamic capabilities (Teece, Pisano and Shuen, 1997). That theory suggests that the scope of the firm could be explained as a result of the dynamics of resource management and the selection environment (Teece, Rumelt, Dosi and Winter, 1994).

In the last few years, a convergence between these two fields has started. Transactions cost economists, in particular, now accept that we cannot fully understand choices of scope without assessing the resource bases of firms. Williamson himself recognizes that the transaction cost and internal firm perspectives "deal with partly overlapping phenomena, often in complementary ways" (1999: 1098) and points out that a firm's history and capability endowments matter to boundary choices, a theme developed by Argyres (1996) and Argyres and Liebeskind (1999). Williamson also recommends that the traditional TCE query "'What is the best generic mode (market, hybrid, firm) to organize X' be replaced by the question 'How should firm A -- which has pre-existing strengths and weaknesses (core competences and

1

disabilities) -- organize X?' " (1999: 1003). This question has been recently pursued by Madhok (2002), who suggested that an individual firm's choice must depend not only on the characteristics of the transactional conditions, but also on its strategic objectives, the attributes of its own capabilities, and the governance context it has created. There is by now substantial empirical support for the proposition that considerations of transaction governance trade off against capability considerations when firms choose component suppliers (Walker and Weber, 1984; Poppo and Zenger, 1998; Schilling and Stensmaa, 2001; Afuah, 2001; Jacobides and Hitt, 2001; Hoetker, 2003). These contributions consider the complementary roles of transactional considerations and capability considerations in the micro-analysis of firm decisions.

This paper proposes a qualitatively different step in the direction of bringing the analysis of transaction costs together with the examination of capabilities. Specifically, we argue that capabilities play a pivotal role in the dynamic evolution of transaction costs; and that transaction costs, in turn, shape the distribution of capabilities over time. We identify the specific causal mechanisms involved in this co-evolution of transaction costs and capabilities; to explore these mechanisms, we have to shift the focus to the level of the industry. To understand the menu of choices a firm faces at any point in time (based on its relative capability and transactional environment), we have to understand the generative process at the level of the industry, which shapes the menu of choices for each and every firm.

The contribution of this study, then, is the analysis of the evolutionary dynamics shaping the firm's environment, and of the mechanisms through which capabilities and transaction costs co-evolve. Concretely, we identify five causal mechanisms to explicate the interrelationship between capabilities and transaction costs. First, we consider *the intra-organizational formation of the vertical divide*, i.e., the progress of internal organization in firms, which creates internal boundaries that may subsequently become firm boundaries. Second, we examine the impact of *capability differences and selection effects on vertical scope and industry structure*. The extent

of intra-firm differences in relative competence along the value chain (the unevenness of the distribution of capabilities up- vs. down-stream) is the key source of incentives to create and use an intermediate market. Simply put, a firm that finds itself relatively weak in one vertical stage is likely to look outside for help. Third, we consider the *endogenous transaction cost reduction* mechanism. Specifically, we suggest that the extent of actual reduction of TC in an industry is endogenous since a) forces of selection, imitation and scale economies reduce the effective heterogeneity of firms, thus affecting incentives for TC reduction, and b) individual firms have incentives to invest in reducing TC when the increased "extent of the market" will increase the rents they expect to capture from the increased "division of labor" (Smith, 1776; Stigler, 1951). Fourth, we look at the role of *knowledge accumulation*, and argue that vertical specialization may facilitate (and come about because of) the accumulation of specialized knowledge. We then show how these four mechanisms interact and work in self-reinforcing ways, determining the industry's scope on the basis of the process of capability development. We then consider a final mechanism, which suggests that *changes in the transactional environment affect the process of knowledge accumulation and capability development* in ways that reach beyond the industry boundaries of a particular time; thus TC shape the nature and origin of the capabilities in an industry.

Our analysis extends the systemic analysis proposed by Silver (1984), Langlois (1992; 2003) and Langlois and Robertson (1989; 1995). It thus provides some insight on what Coase (1991) termed the "Institutional Structure of Production" (ISP) – a set of systemic relationships and structures extending beyond the make-vs.-buy choices, as we elaborate in the discussion.

The paper is organized as follows: The next section lays out our analytical premises. We then consider the specific co-evolutionary mechanisms that affect vertical scope, the distribution of capabilities, and the transactional environment. A portion of the verbal analysis of this co-evolutionary process is formalized in a simple dynamic model that features an extension of R.A.

Fisher's "Fundamental Theorem of Natural Selection." We then illustrate the working of these mechanisms with an account of the recent evolution of the mortgage banking industry. The final section reconsiders Coase's legacy, noting that he has repeatedly indicated the need for a fuller analysis of the ISP, and finally concludes with implications of this research program for research and practice in strategic management.

Explaining Vertical Structure: Premises and Definitions

Level of Analysis: From the Firm to the Firm and the Industry. While our interest is in the analysis of the evolutionary mechanisms that co-determine scope and the division of labor between companies, we focus our discussion around a narrower problem. This choice helps us make our point more clearly, and illustrates the links with existing research. We focus on the question of vertical integration, which has also been used as the "main case" for TCE (Williamson 1985, 1991). However, unlike much organizational economics analysis, we examine the vertical structure of the *industry* rather than the choices of the individual firm. Most importantly, we concentrate on the dynamics rather than the statics of the determination of vertical structure. For analytical simplicity, we posit the existence of a single, significant vertical interface in the production process.[1] We ask what determines whether this interface corresponds to an intermediate product market or is organized within firms; and we consider how vertical specialization shifts from within to across firm boundaries over time.

Behavioral Assumptions: Profit Seeking. We do not assume that firms are guided by comprehensive accurate foresight regarding the implications of the dynamic process and of their own choices. Rather, we treat firm behavior as "profit seeking" – responsive to the prevailing incentives at each point of time and occasionally involving efforts to transform the developing situation through creative strategic action, but not informed by a clear view of where the system

---

[1] In reality, of course, there are typically a multitude of such interfaces, at a series of vertical stages in the value chain, and their importance to the focal firm varies widely; see Baldwin and Clark (2003) on the development of discontinuities in the value chain, and Jacobides (2003) on the emergence of intermediate markets.

is headed. In this respect, our approach reflects the perspective on firm behavior taken by Nelson and Winter (1982, 2002).[2]

Transaction Costs as Market Friction. The Coasean tradition emphasizes that there are costs of using the market (or "the price system"). These costs are obviously highly relevant for an individual firm confronting the "make or buy" decision. In that context, different types of costs can usefully be distinguished. Coase himself emphasized the "frictional" costs, such as those of identifying a potential supplier, negotiating, drafting a contract and monitoring it, etc. Williamson (1975, 1985) transformed the subject by shifting attention to the costs of transactional hazards and of governance arrangements to limit such hazards. His focus is on the tendency of transactions to run into difficulty for reasons associated with bounded rationality and opportunism, whereas the frictional costs are present even when things go well -- a feature of economic reality more like physical friction, being independent of human calculation and motivation. Alternatively, transaction costs may arise from difficulties in measuring and monitoring performance (Alchian and Demsetz, 1972; Barzel 1982), or the inability to specify the goods and services needed (Jacobides and Croson, 2001). While the distinctions among these types of costs are clearly important for some purposes, such as the micro-analysis of governance arrangements, they are all quite similar when viewed in a systemic perspective. They all represent burdens or obstacles to market transactions, and they are all potentially subject to reduction, at least in the long run, through some combination of managerial ingenuity and appropriate investments. In their systemic consequences, these costs are all akin to a tax on market transactions.

Transaction Cost Reduction: The Role of Agency. In the firm-level and short-run analyses of transaction cost economics or (incomplete) contract theory (cf. Hart, 1995), the problem is basically one of choice from a menu of governance alternatives. We argue that the

---

[2] Our case example below illustrates the empirical relevance of this low-foresight perspective.

choice menu from which a firm picks is determined by the conditions of the industry as a whole, at each point of time, so that even an optimal choice is constrained by an institutional environment that is *fixed* in the short run. That environment is determined by such slow-evolving things as prevailing contracting norms, firm reputations and transactions technology, as well as the existence of "interfaces" that can support market exchange to a greater or lesser extent. Thus, we qualify the emphasis of transaction cost economics on the ability of a firm to choose (minimize) its transaction costs in the short run, by recognizing the constraints in its menu of choices. At the same time, however, we stress that a firm can influence the transactional environment in the medium term – shaping the "TC context" within which the short-term choices are made. Indeed, we suggest that there is significant firm agency in shaping the TC context of the entire industry (and not strictly for itself, for better or for worse) by its own actions. For instance, an individual firm may come up with a particular way to organize its production by, say, creating a new way to measure and assess an intermediate good, or a new way to coordinate the up-stream and down-stream operations. This reduces TC and enables specialization, thus reshaping its institutional context for the next period. And, by and large, the benefits of that action affect the economy as a whole – at least as far as such transactional solutions proliferate through imitation.[3]

Heterogeneous Capabilities. We also adopt the premise that capabilities are quite heterogeneous across firms and across stages in the production process.[4] The term "capabilities" embraces the underlying determinants of the efficiency with which firms manage to carry out their productive activities. Capabilities rest on the firm's general and specific knowledge of how to do things (Richardson, 1972; Teece, Pisano and Shuen, 1997), but also involve the specific

---

[3] The term "capabilities" should not be reserved to the sphere of production, for transacting itself involves capabilities, built through experience, learning and investment (Winter 1988). To the extent that such a firm-specific advantage does not facilitate potential transactions across firm boundaries for *other* firms, we consider it to be "capability enhancing" rather than "TC reducing" for the purpose of this paper.

[4] For striking evidence, see Lieberman and Dhawan (2001). Their charts display time series for various performance measures for U.S. and Japanese auto producers, and show a wide variation in both levels and trends.

investments in equipment, training and retention of key personnel, etc., required to put that knowledge to work. Heterogeneity is typical because the capability to carry out a complex activity is typically developed in an organization through a long, path-dependent learning process, in which there is abundant opportunity for various contingencies to shape the way of doing things that ultimately emerges (cf. Levinthal, 1997). Particularly important contingencies are the different "bets" that actors make in the face of great uncertainty as to what will prove to be the most effective way of doing things.[5] Thus, even in environments where primary resources are quite homogeneous, different organizations are likely to display significantly different ways of accomplishing approximately the same thing, displaying different efficiencies as a result. A corollary observation, important in the present context, is that we should not a priori expect strong similarities between the capabilities a single organization displays at different vertical stages.

Starting Point: In the Beginning, There Were … Firms.[6] The last piece of the frame of our analysis deals with the initial condition of the industry. Following the tradition of Smith (1776) and Stigler (1951), our "main case" is where the industry is "born integrated," and dis-integration may subsequently emerge in the evolutionary process. The rationale and character of this assumption need explication. The birth of an industry occurs when some innovative product or service crosses the threshold where it begins to be producible at a cost, and with quality attributes, such that a substantial number of customers are willing to pay a price that yields a profit to the seller.[7] Novelty in what we identify as the "downstream" segment is therefore a

---

[5] Even if the correct recipes become clear, their diffusion is limited by complexity, often due to interactions among activities (Porter 1996, Rivkin 2001, Siggelkow 2001). The force of imitation is also weakened by the path-dependence associated with the fact that investments in capabilities are so often durable and/or sunk. Correction of past mistakes is not necessarily economic at the relevant margin.

[6] We think of this as a (qualified) historical proposition, running counter to Williamson's converse proposition favoring markets (Williamson 1985, p. 87). Williamson seems to be offering more an analytical starting point than a stylized historical one.

[7] There has been a substantial amount of work detailing the processes at work in the early stages of an industry and describing important examples in detail. See, for example, Abernathy and Utterback (1978) , Utterback (1984), Klepper (1998) or Langlois (2003). For a simple theoretical account of industry birth, see Winter (1984).

factor in the initial situation by definition – regardless of whether this product is sold to producers or consumers. This need not be true of the upstream segment. If there is an established market for an intermediate product needed by the new industry, early reliance on that market is virtually a foregone conclusion for reasons of both efficiency and minimizing capital requirements for the new producers, and that initial pattern may never be challenged. By contrast, if the upstream product is itself novel and its appearance incidental to the appearance of the downstream product, we expect to see integration at the start.

This is true in the first place because the new downstream producers may have no effective and profitable way to trigger the appearance of a supplying industry to meet their needs. The question of whether the required product is available on the market may have a sharp "no" answer, or it may have a more or less ambiguous "yes" answer. The suitability of existing products for the new role is often a matter of degree – some adaptation or improvement may be required. This is the second reason for integration early on -- the need to customize the intermediate product to novel ends, and also to learn to manage the interface between the stages. In the early stages of an industry, this management task is itself changing along with design details upstream and downstream, making the coordination task particularly challenging. Langlois (1992: 116) summarizes, "Ultimately, the costs that lead to vertical integration are the (dynamic) transaction costs of persuading, negotiating with and coordinating among, and teaching outside suppliers in the face of economic change or innovation." [8]

Sometimes the balance of short-term considerations at an early life stage is such as to defer integration for a relatively short interval. An industry may start off as a vertically-specialized assembly sector, employing inputs made originally for different uses. But integration is often quick to arrive, as firms seek out new ways for organizing their own production

---

[8] An example of this concerns the meat packing innovations of Gustavus Swift, a tale recounted by Chandler (1977), Porter and Livesay (1971) and Langlois (2003) among others.

process.[9] We do not examine this case explicitly, but consider it to be adequately approximated by our assumption that industries are "born integrated," and that markets emerge only later.

Co-evolution of Capabilities and Transaction Costs

This section provides the analysis of the principal causal mechanisms that operate on the vertical structure as the industry evolves. Although all of these mechanisms continue to play a role throughout the industry's history, some are more important in the early stages. In our exposition, we place at the top of our list the mechanisms that are more important early on. Also, the relative importance of the mechanisms changes as an endogenous feature of the process; we flag this point in the important cases. After examining the first four mechanisms individually, we consider how they interact, driving vertical scope. We then briefly discuss how scope, in turn, affects the capability development process. Table 1 provides a brief summary of our argument.

Insert Table 1 about here

First Mechanism: The Formation of the Vertical Divide. As integrated firms develop better methods through processes of learning and search, they typically display increasingly well-defined and stable internal structures. Such a process of organizational unbundling increases efficiency of operations and, critically, enables scaling up a business from an integrated crafts-based production system to a larger-scale yet manageable unit (Fayol, 1921; Chandler, 1962). To the extent that the issues addressed by these structures are significant and inherent in the activity, there are likely to be broadly similar responses (see Gould, 1980, chapter 2, on the analogous process of "convergent evolution" in biology).

For example, a need for accountability for work-in-progress inventories will require clarity about when transfers of control take place; and this pushes for vertical structure inside

---

[9] An example is the early days of automobile production, where producers typically started off as engine assemblers who relied on the existing market for horse-drawn carriage parts to acquire inputs for automobile bodies. Soon

firm boundaries. These vertical discontinuities are more likely to be set at points where the condition or quality of the product is relatively easily estimated (Barzel, 1982, 1997; Baldwin and Clark, 2003). Such factors shape the embryonic form of the vertical interface as it appears within the integrated firm, and the similarity of the managerial challenges of the different firms in the industry will tend to produce similar solutions among firms.

In addition to the coordnation challenges, a number of managerial challenges drive the vertical division of labor within firm boundaries. The parts of the process upstream and downstream from the interface may, for example, call for different managerial styles and incentive mechanisms (Ghemawat and Ricart I Costa, 1993), or different "dominant logics" (Prahalad and Bettis, 1986). They may draw on different knowledge bases and hence require different types of specialized personnel, and demand familiarity with different technical languages (Argyres, 1999). Any such managerial diseconomy of scope provides a reason to revert to separate organizational units (Lawrence and Lorsch, 1967; Monteverde, 1995). Likewise, as firm size increases, the need to restrain the growth of managerial load at the top provides pressure for sharper demarcation of managerial responsibilities below, and these pressures may lead, for example, to the identification of units as cost or profit centers. These changes make parts of the production process more autonomous; and as such, there is an endogenous pressure to vertically dis-integrate, so as to maximize the relative strengths of a vertically specialized unit. The development of an organizationally un-bundled structure within the integrated firm lays the foundation for subsequent dis-integration (Jacobides, 2003). It defines the "dotted lines" along which the market process of the future will cut the firm apart.

Second Mechanism: Capability Differences and Selection Effects. As noted above, we posit that the initial situation in the industry is characterized by integrated firms with heterogeneous capabilities. Our focus will be on two aspects of this heterogeneity, diversity of

---

enough, this "flexible specialization" gave way to integration, as the horse-carriage producers could not serve well the needs of the new industry (Langlois and Robertson, 1989; 1995).

unit production costs and diversity in internal organization at the vertical interface. It is important to note, however, that the initial heterogeneity is likely to extend beyond these aspects, into differences in product attributes and quality, and even to the deeper question of what customer needs are fundamental to the definition of the product itself. A sharpening of the industry definition, and of its self-perceived "identity," is an important aspect of the broad process of competitive struggle and sorting-out that is characteristic of the early life stages of an industry (Utterback and Abernathy, 1975; Utterback, 1994; Klepper, 1998.)

In those early stages, the cost differences that matter are the unit production cost differences of the integrated firms. Those firms with the lower overall costs are more profitable and are likely to grow – the more so because of the attractive opportunities in the marketplace, the threat from ambitious rivals and the desire to seize advantageous positions before others do. As time passes, however, there are likely to be occasional inter-firm transactions across the vertical interface, then more frequent transactions, and finally a process of evolution toward an institutionalized intermediate market (as we detail below). This implies a change in the relevant cost differences, since in the later stages the upstream and downstream costs of a firm play independent roles whereas only the total mattered earlier. Our formal analysis below lays out the implications of this change for industry structure and performance. Briefly: in every period, the economic selection process of profit-driven growth implies that the scale of the most successful firms, or vertical stage units, will increase -- in many cases very dramatically. Simply from a statistical viewpoint, this means that the effective heterogeneity in the industry is decreased; the weight on the choices made by the few most successful firms has increased. The same observation applies to internal arrangements at the vertical interface. To the extent that successful firms tend to replicate those arrangements as they extend their capacity, selection forces produce a decrease in the effective variety of such arrangements across the industry.

As a consequence, success and failure become more apparent as time passes. This increases the power of a related mechanism that is also operating to decrease heterogeneity -- imitation of the more successful firms by others, including the particularly important case of imitation in new firms founded by managers who began their careers in those more successful firms and know a good deal about what is done there (Klepper 2002).[10] These considerations apply to firm practices in general, and hence to production costs, and they apply in particular to arrangements at the vertical interface. Both at the level of firms and at the levels of production units or individual practices, the tendency of selection is to move the industry toward greater homogeneity of method and uniformity in the division of labor. In general this trend is accompanied by an improvement in productivity and other performance measures over the course of the industry's history. However, even in the absence of hard evidence on the efficiency of conducting business in any particular way, pressures of institutional isomorphism tend to develop and soon restrict the menus of intra-firm organization of activity (Meyer and Rowan, 1977; DiMaggio and Powell, 1983). For all these reasons, vertical structures within the industry become increasingly homogeneous.

Third Mechanism: Endogenous Reduction of Transaction Costs. Although the TC context is fixed in the short run, it is subject to change in the medium term as firms make incremental adaptations and explore new directions in the quest for profit. In taking this view, we adopt the evolutionary economics stance that business practices are akin to fixed plant and equipment, and hence usefully analyzed by distinguishing the causal forces operating at different time horizons (Nelson and Winter 1982; Marshall 1948).[11]

---

[10] More generally, the founding of new firms is obviously an important feature of industry evolution, as is the failure and disappearance of unsuccessful firms. Our strong emphasis on differential growth, as opposed to the demographic patterns, is characteristic of the tradition of evolutionary economics as distinguished from organizational ecology (Winter 1990; cf. Hannan and Freeman, 1989; Carroll and Hannan, 1995).

[11] In many cases, including many of particular interest in the Information Age, there is clearly more than "kinship" involved: the practices and the equipment supporting those practices are not really separable. In such cases, the actual business decisions relate to the acquisition of capabilities (whether production or transactional) that combine specific types of knowledge, human capital and physical capital in effective ways.

The efforts to prepare the ground for dis-integration are made under a wide range of conditions of risk and uncertainty. But there are generally incentives to promote a TC-reducing change, and the extent of these incentives is affected by the selection process just described. Such effects arise, for example, in the context of a "standards war," like IBM vs. Apple in personal computers, or VHS vs. Betamax in VCRs (Cusamano, Mylonadis and Rosenbloom, 1992). As one contender starts to prevail in such a contest, the incentives of firms at various points in the vertical chain are affected. The tilt is in the direction of making adjustments to support their attempts to share in the increasing flow of business stimulated by the emerging winners, e.g., as their co-specialized suppliers. Much of the expense of such an adaptation is likely to be a one-time cost in the sense that the benefits would continue without significant further expenditure, until another major design change comes along.

Such investments by individual firms also affect the situation with respect to asset specificity. Asset specificity is particularly characteristic when the industry is largely integrated and the upstream and downstream segments are idiosyncratically specialized to each other, but it fades as the industry converges on designs and standards for the intermediate product. In this perspective, the market is seen to be most hazardous when it is an organizational frontier (cf. Langlois 1992). Finally, investments in the creation of market interfaces are often imperfectly appropriable. Firms make them when they anticipate that their private benefits will exceed their costs, but even firms that had not been concerned with funding the creation of the market might, once the fixed costs are sunk by their competitors, adopt their ways of transacting and thus fuel further specialization.

This leads to an important behavioral observation – namely, that the efforts of firms in terms of shaping the boundaries of their industry are quite often not driven by a grand design and a well-articulated sense of the structure of the industry. Rather, firms are directed to areas of short- to medium- term profit potential, even if this, in the long run, undermines their own

strategic position. This point has been well made with respect to the personal computer industry (Baldwin and Clark, 2001); we illustrate it again in Section 5. An important reason for the apparent paradox is, of course, the non-cooperative logic of the competitive game. True, some of the industry participants might be tempted to not invest in, or even discourage, vertical dis-integration, to protect their interests. But other industry participants or new entrants will, and the holdouts will be brought into line. This is why the important endogenous mechanism in the course of industry evolution is TC *reduction,* even though an increase in TC might sometimes appear as a tempting rent-seeking move for some participants.

Fourth Mechanism: Knowledge Accumulation and Capability Development. Different activities within the firm build on different knowledge bases. This means that as firms try to improve, specialization might help firms create more effective improvement methods. For instance, a unit that focuses exclusively on acquiring customers may improve customer acquisition practices more effectively, not only because of greater accountability and clarity of objectives of a focused unit (Lawrence and Lorsch, 1967), but also because independent divisions can consider solutions that depart more radically from existing practices. The scope of references of a specialized department within a firm is deeper and better chosen in its own domain than for the firm as a whole. The customer acquisition unit may consider techniques not only in its own field, but also in other sectors, and it will try to integrate the experience from a better calibrated and broader set of related projects and similar cases; the capacity to absorb new productive knowledge arising externally (Cohen and Levinthal, 1990) depends on scope. Latent gains from specialization thus push firms to build intra-firm interfaces, to reap both the static and dynamic benefits. So the capability development process itself is a driver of dis-integration, provided that the separation does not create "silos" that inhibit systemic business improvement.

To the extent that the specialized production leads to faster knowledge accumulation, vertically specialized firms may be able to improve more quickly than the integrated ones. So

even if the initial capability endowment favors some larger, integrated entities, the knowledge accumulation process may ultimately lead to gains from specialization, and thence to gains from trade. This means that the emergence of gains from specialization may depend (a) on the extent to which the knowledge bases of the different segment are divergent; and (b) on the extent to which focusing on one of the two improves the ability of any given firm to enhance its productive capability (Argyres, 1996; Christensen, Verlinden and Westerman, 2002.) Whether integration or dis-integration provides the stronger basis for capability improvements is an empirical issue; the answer tends to vary with industry life-stage.[12]

Interactions and Feedback Loops. The four mechanisms just described do not operate independently as the industry evolves. They interact, and as the interactions occur over time, the general effect is to amplify the effect of the mechanisms individually. The structure of these interactions shapes the co-evolution of capabilities, transaction costs and the division of labor in an industry. The first and most crucial interaction is between transaction cost reduction and the selection process. Given the TC context of any particular time, there will in general be a mix of integrated, partially integrated and specialized firms. Market transactions will occur where the differences in productive capabilities between the relevant firms are at least as large as the transaction costs incidental to the transaction. So while a firm does take transaction costs into consideration in determining its decision to produce or buy, it also looks at the comparative efficiencies, i.e., gains from trade for so doing. And these gains from trade are determined by the heterogeneity of capabilities. Transaction cost reductions are consequential only if there is underlying heterogeneity in the production structures up- vs. down-stream in the industry (a point we treat formally in Section 4). Only to the extent that firms are unevenly efficient in the upstream or the downstream side of their business, a reduction of transaction costs may directly

---

[12] This may help explain why, when knowledge bases change, industries often shift from dis-integration back to integration: In the new status quo, knowledge accumulation in an integrated firms may be superior (cf. Fine, 1998).

increase vertical specialization, thereby shifting the locus of selection from paired segments in integrated structures to independent segments.

The second interaction is between the operation of the selection forces and the individual choice of a firm with regard to its scope, and its potential TC-reducing investments. Firms start trading (and even invest to make that trade possible) when they see that alternative mechanisms can profitably complement or supplant internal production in a part of the value chain. But why should this happen? Why would a firm want to abandon what might be a profitable operation, sourcing it outside? Simply put, expansion and profitable growth will have to focus on the areas where a firm has a competitive advantage when compared to the rest of the industry. Even if a firm has two segments where it can make profit, it may choose to drop one of them to concentrate on its core strength. This, however, will only be the case if the balance of competencies in the rest of the industry implies a match between what the focal firm wants to outsource and what the rest of the industry wants, on net balance, to supply. Dis-integration requires gains from trade, not just gains from specialization, to emerge. It also requires the selection environment to be stringent enough to prompt the abandonment of the competitively weaker market / value chain position, which may explain why vertical reorganization is often prompted by an economic crisis (North, 1986; Silver, 1984).

Other important interactions between the four mechanisms we discussed can arise to the extent that there are scale effects on learning and other investments. The immediate consequence of market selection is to increase the scale of some firms while shrinking, or eliminating, others. Larger scale in turn motivates investments to improve productivity; and this further increases scale; and as such, specialization (driven by the motive to become large enough in one segment so as to afford the investments to be most productive) dominates. A related self-reinforcing cycle happens, as we have said, through the investments on methods of transacting. By this

channel, transaction cost reductions beget more of the same, and knowledge accumulation begets more of the same, and both intensify selection pressures and beget more specialization.

The Role of Historical Contingencies. A variety of contingencies can lead to episodes in which firms are led for the first time to transact with each other across the vertical interface. Accidents, natural disasters, labor disputes, etc. may impair one vertical segment in a firm, leaving it out of balance and with a strong incentive to overcome barriers to use of the market (e.g., North, 1986; Langlois, 2003). Attempts at rapid growth may encounter much different levels of friction/ adjustment cost in the two segments, similarly producing imbalance. Demand changes in downstream sub-markets -- including the appearance of a new use for the intermediate product -- can produce upstream-downstream imbalance. Opportunities from importing capabilities developed in other industries may enhance gains from specialization. Such contingencies and the resulting imbalances provide specific stimuli for firm actions to reduce transaction costs. This sporadic and exploratory use of the market produces initial solutions to some of the problems of using the markets, perhaps including recognition of the desirability of some standards both for products and ways of contracting. As noted above, these adjustments are largely one-time costs. The convergence on a standard to coordinate market exchange does not have to be re-invented, and the sorting out of how quality issues are to be handled between buyer and seller is a process which does not start anew with every deal; it can improve, but builds on the solutions that have been found so far. A cumulative process gets under way, as transaction cost reductions improve access to gains from specialization, expand the market, and thus shift incentives toward further transaction cost reductions. Thus, these one-time investments irreversibly dispose the industry toward greater use of the market, and also set the stage for further incremental investments to the same end. The individual attempts of a few firms to shape the transactional environment for their own benefit leads to changes in the transactional environment other firms face, thus facilitating further vertical dis-integration.

17

The picture discussed above is, we would argue, a typical one provided there is sufficient underlying heterogeneity in the capability distribution or the knowledge base. Otherwise, there will be no reason for this co-evolutionary sequence to start. Yet we should also note that these are necessary, rather than sufficient, conditions. The size distribution of industry participants may be such that no entity is large enough to be able to afford the investments that would lead to an effective TC reduction; so absent government intervention or effective collective action through industry associations, the potentially inefficient integrated status quo may remain, until historical accident jump-starts the process (Polanyi, 1957; North, 1986).

Finally, this exposition assumes that the selection environment is relatively tight. This, however, may not be the case in industries with a high degree of concentration and control, where the incentives of the dominant, strong players are to maintain, say, an integrated structure in the industry. In such cases, firms may attempt to thwart the efforts of a more dis-integrated mode of production, as this may eventually reduce the profitability of the industry. Thus, in some settings, long-term strategic considerations may forestall competitive pressures, making for near-term efficiency or rent-seeking.[13]

Fifth Mechanism: The Impact of Changing Scope on Capability Development. Our analysis so far has explained how organizational, knowledge- and capability- related factors interact in shaping vertical scope. The final mechanism we briefly consider relates to the effect of changes in vertical scope on the process of capability development, the knowledge accumulation process and the roster of actors qualified to compete. In particular, when specialization breaks an industry into pieces, some pieces may correspond closely to activities in other sectors. For instance, the vertical specialization in banking and the creation of specialist data-handling units highlighted the promise of drawing from data-management practices outside

---

[13] We are inclined to the view that such cases are relatively rare. Some authors clearly disagree, and see it as more typical that individual firms exert substantial control in distorting and reshaping the competitive environment to their advantage (Bowles and Gintis, 1988; Lazonick, 1991). The empirical issue here merits more attention.

of banking. It also allowed companies previously not associated with banking to export their capabilities to the newly specialized sector. Even the presumed boundaries of the industry, as viewed at a particular time, may thus be called into question by the adjustment of intra-firm knowledge boundaries. The more a specialized unit draws on knowledge sources outside the industry, the more it draws attention to its particular function as one that is performed well outside the industry. The result may be, for example, that a "model" functional specialist from outside the industry sees an attractive niche within the industry and tries to get the business – and if successful, this may ultimately become a "toehold" entry. This sort of process is prominent today in business process outsourcing and other areas involving substantial IT applications. So changes in firm scope affect both the dynamics of capability development, and the relevant actors who bring these capabilities to bear.

Summary and Implications. To summarize, we propose that the evolution of the ISP is driven by the selection mechanisms that operate within an industry; that these selection mechanisms will tend to reduce variety in the ways labor is organized within firms (though not necessarily yielding maximally efficient ways of organizing production). Once a clear division of labor emerges within firms, across-firm comparisons of segment performance may begin, and these may lead to the creation of a market. Whether this happens – or the extent to which it happens at each point of time – depends on the balance between the force of heterogeneity of capabilities and the friction of transaction costs. The intensity of the TC reduction effort depends on the latent gains from so doing; and the latent gains from trade depend, in turn, on the distribution of capabilities along the value chain. Finally, the TC context further feeds back into the process of capability development: Specialization opens up the industry to a new knowledge base, and possibly new industry participants. Figure 1 summarizes our argument.

Insert Figure 1 about here

19

Selection and Vertical Integration: The Fisher Theorem, Extended

In the preceding section, we explained how differences in capabilities lead to changes in scope; this section explains how different conditions in terms of scope (integration vs. specialization) affect the economic selection process. In particular, we extend R.A. Fisher's "Fundamental Theorem of Natural Selection" to show that, when capabilities are heterogeneous, a dis-integrated vertical structure can give rise to more rapid progress than an integrated one – if transaction costs permit. And this occurs for selection reasons alone, even in the absence of improvements triggered at the firm level by specialization.

We present the full exposition, including the mathematical calculations, in an appendix. Here, we describe the model, state the result and discuss it in an informal way. The model portrays the evolution of industry average cost in an industry with vertical segments. When unit production costs are constant but differ across firms, profit-driven differential growth shifts output shares in favor of the most efficient firms, thus reducing the industry weighted-average cost. The rate at which cost is reduced by this mechanism obviously depends on the cross-sectional dispersion, or "variety," of efficiency levels displayed by firms. That variety is the only thing a pure selection process has to work with. The economic analogue of Fisher's Theorem (Fisher 1958)[14] makes this point more precise: it says that when firm growth is determined by the re-investment of net profits, the rate of reduction of the industry average cost is proportional at every point of time to the cross-sectional, share-weighted variance of unit cost. This means that there is a lot of potential for progress when firms with non-trivial shares have widely dispersed efficiency levels; the progress comes when the strong performers grow relative

---

[14] R.A. Fisher, the famous statistician, was also a pioneer of the mathematical modeling of evolutionary processes. His theorem states that the rate of increase of (average) fitness in a population is equal to the genetic variance of fitness; the term "fitness" refers to the rate of increase of the type, generally normalized in some way – e.g., relative to the average or to the most fit type. Various results of this kind can be proved; our treatment in the appendix illustrates the common mechanism of them all, that relative frequencies or shares change as a result of the differential growth of the different types. Because the variance in question is a share-weighted variance, it can be low for different reasons – e.g., that all fitness/average cost values are closely packed, or that the values are quite dispersed but the *weight* is highly concentrated in a narrow range. The rate of advance is small in either case. The

to the weak ones. By contrast, progress will be slow (in the near term) when one firm is much more efficient than the others but also very small. It will grow larger, and when it does its growth will improve the industry average rapidly – but while it is still small there is little impact either on the level or rate of change of industry cost.

 We first reproduce the standard result for the case where all firms are integrated, and then compare the case where the two segments are independent, i.e., separated by a market. Cost reduction is more rapid in the latter, to a degree that depends on the cross-sectional correlation between upstream and downstream costs. That is, we demonstrate that the degree of improvement in the capability pool of an industry is higher under specialization, but only when the up-stream and down-stream capabilities do *not* co-vary systematically. We illustrate this result with charts based on numerical integration of the differential equation system, and then discuss some of the broader implications of the exercise.

Our simple model assumes that each firm has a single technique upstream and another downstream. Per our discussion in the previous section, we consider capacity fixed in each short run period; from one period to the next a firm can grow as a function of its profits. For simplicity, we analyze the case where growth is funded by the profits (which go to buy the capital stock needed for the expansion).[15] Firms produce to capacity at each point of time and the aggregate output determines price when it encounters a demand curve in the marketplace. On the basis of these calculations, we derive the industry average cost, which depends on the individual cost of production of each firm and varies over time as shares change.

If firms have to be integrated (because the TC are so high that they more than offset the potential gains from trade), then the relevant cost is that for the paired units, upstream and down (since each firm produces in both segments). The industry cost is the weighted average of

logic of Fisher's theorem has previously been applied to various problems in evolutionary economics, see Nelson and Winter (1982), Iwai (1984), Metcalfe (1998), for example.

individual firm costs, the weights being the shares of production. We then consider how these shares in the production process change over time; clearly, the firms with the highest profitability will grow the most, as profits are re-invested to fuel capacity growth. The insight is that a successful integrated firm must re-invest its profits to fund growth not only in its efficient part of the value chain (say, upstream) but also the inefficient one (say, downstream) as it needs to remain vertically balanced. On the basis of this calculation, we can compute the rate of change of industry average costs $\frac{dc}{dt} = \sum \frac{ds_j}{dt} \cdot c_j$, i.e., the sum of firm unit costs multiplied by the changes in their share of industry output. The change in costs depends on how much more volume the efficient firms take up. This last equation, after appropriate manipulation, gives the Fisher theorem result, that is

$$(0.1) \qquad\qquad \frac{dc}{dt} = -(v^U + v^D)^{-1} \cdot Var(\mathbf{c})$$

where $Var(\mathbf{c})$ is the share-weighted cross-sectional variance of the vector $\mathbf{c} = (c_1, ..., c_N)$, that is, the weighted variance of unit costs. This means that the change of costs in the industry is proportional at each point of time to the variance of the *overall* costs of different firms (where the upstream and the downstream segment are lumped together). The v's are the values of capital required per unit output, upstream and down. Higher values mean slower firm growth and hence slower progress.

Computing the Results for the Dis-integrated Case. We next perform the thought experiment of an instantaneous dis-integration of the industry, the result of a notional drop of TC from prohibitive levels to zero. The basic calculations we perform in the model are the same, with the notable exception of the fact that we now have two sectors -- an upstream and a downstream, both of which have "products" (an intermediate and a final product respectively).

---

[15] The qualitative results hold if growth tends to be roughly proportional to profits across firms. Progress will be slower if a portion of profits is not-reinvested, faster if external funding supplements the retained earnings.

Whereas before we had introduced the need of each firm to be vertically balanced, we now let the two segments respond to their individual profitabilities. In the moving equilibrium, segment capacities and outputs will balance overall at each point of time, through the endogenous setting of the intermediate price. We then compute the evolution of the industry-wide costs. These costs have to be considered separately now, as each firm has in general a different share in the upstream and the downstream side of the industry. Letting $c^U(t)$ and $c^D(t)$ respectively be the upstream and downstream share-weighted industry costs, we can compute the changes in the upstream and downstream share-weighted industry cost, which is

$$(0.2) \qquad \frac{dc^U}{dt} = (v^U)^{-1} \cdot Var(\mathbf{c^U})$$

$$(0.3) \qquad \frac{dc^D}{dt} = (v^D)^{-1} \cdot Var(\mathbf{c^D})$$

where $Var(\mathbf{c}^U)$ is the share-weighted cross-sectional variance of the vector $\mathbf{c}^U = (c_1^U, ..., c_N^U)$, that is, the weighted variance of unit costs in the upstream sector alone. Similarly, $Var(\mathbf{c}^D)$ is the share-weighted cross-sectional variance of the vector $\mathbf{c}^D = (c_1^D, ..., c_N^D)$. This means that the way the industry's cost changes under the dis-integrated case is $c(t) \overset{!}{=} c^U(t) + c^D(t)$, as defined by (0.2) and (0.3). This is *not* the same as the integrated case, which was $c(t)$ characterized by (0.8). This is because the upstream and downstream shares are allowed to diverge in the dis-integrated case, with the ending of cross-subsidization of the relatively inefficient segment by the relatively efficient one as a corollary. Industry costs do *not* generally change the same way in the integrated and the dis-integrated cases. Because of the different dynamics under the integrated and market-mediated conditions (i.e., the different share increases), the causal mechanisms are different.

Comparison of Integrated vs. Specialized Production. Having derived the Fisher theorem for both the integrated and the dis-integrated case, we can now evaluate how dis-

integration can affect the changes in industry costs. What principally obtains from our analysis is that we need to consider not only the variances, but also the co-variance between the upstream and the downstream sectors.

Consider first the case where the costs upstream and downstream have the same variance and are perfectly correlated. In that situation, we can write

$$(0.4) \qquad Var(\mathbf{c}) = Var(\mathbf{c^U}) + Var(\mathbf{c^D}) + 2 \cdot Cov(\mathbf{c^U}, \mathbf{c^D}) = 4 \cdot h$$

Here, h is the common value of the two variances and the covariance. Suppose also that the prices v^U and v^D have the common value v. Using (0.1) with (0.4) we get

$$(0.5) \qquad \frac{dc}{dt} = 2 \cdot \frac{h}{v} = \frac{d\overset{!}{c}}{dt}$$

Here, the second equality follows from adding (0.2) and (0.3). So the perfect correlation simply means that the dis-integrated case reproduces the result of the integrated case.

Compare this result with the case where the two variances are the same (=h), but the covariance is zero. Then, by comparing (0.4) with (0.2) and (0.3), we see that the rate of cost-reduction in the dis-integrated industry is twice the rate in the integrated one. The reason for this lost efficiency under integration is that integration ties together the fates of the upstream and downstream segments, although the competence of those segments varies independently; and firms which are good in one segment are "tied down" by the fact that they have to grow in both segments. If only a market existed, they would specialize -- and the industry would improve, too.

Thus the lack of covariance between costs in the two segments not only *motivates* the creation of a market, as we suggested in the previous section; it also affects the *extent and speed* of improvement at the industry level. This may be illustrated by displaying the results of a numerical integration of the differential equations of the model (see Figures 2 and 3). Each of the figures compares the path of industry average cost under full integration and full dis-

24

integration. There are fifteen firms, with equal shares at the start. Cost parameters are drawn from identical distributions upstream and downstream, with mean one in each segment. In Figure 2, units costs upstream and down are virtually uncorrelated; the R^2 is .004.[16] This figure suggests that in the presence of a limited correlation between the segments' capabilities, specialization both ensues, and intensifies the capability improvement process of a sector. By contrast, the inter-segment R^2 in the calculation underlying Figure 3, is large, about .85. (In other respects the two figures are broadly the same.) The result is a barely discernible discrepancy between the integrated and dis-integrated tracks, and, as expected, both are quite similar to the integrated track of Figure 2.

Insert Figures 2 and 3 about here

Of course, it is a law of evolutionary processes that the action of selection alone is to destroy variety. As shares shift systematically in favor of the more efficient units, the share-weighted variance goes to zero and progress from this source alone tends to disappear; this is represented by the gradually diminishing downward slope in both cases in our graph. Specialization, however, intensifies this competition, and speeds up the process of natural selection; this is represented by the steeper slope of the dis-integrated case. Finally, the magnitude of the initial difference between these two slopes is determined by the degree of correlation between the cost levels in the two segments.[17]

Interpreting the Evolution of Capabilities under Specialization and Integration. Our model indicates that the evolution of capabilities in an industry depends on the degree of integration. Stepping back from the formalism, we offer three remarks about its implications.

---

[16] The lowest cost of an integrated firm is 1.14, and the share of that firm is 50.5% at the end of the calculation. The lowest cost in an upstream unit is .42, and the final share of that unit when an independent firm is 43.4%. The corresponding downstream figures are .50 and 38.0%. The initial capital-output ratio in value terms, on an annual basis, is 2.67, and half of that capital is in each segment. Thus the capital-output ratios are 1.33 in each segment and overall. Each period corresponds to a calendar quarter.

[17] This means that even while firms try to push towards specialization to capitalize on their capabilities, competition in the medium term will lead to a reduction in capability heterogeneity – which, ironically enough, will lead to a reduction of rents and profits (Winter, 1995; Jacobides, 2001).

First, it captures the point that in the context of the problem of vertical structure, the gains from vertical specialization are higher when there is independent variation of competence, upstream and down. Although the "gains" in our formalism are not cast as incentives it is easy to see that the mechanism of this simple evolutionary process implies time-varying incentives to improve the efficiency of the transacting process, and that these incentives present themselves in particular to the more competent production units. Second, to understand the evolution of scope, we should turn to aggregate measures of correlation in the capabilities of different sectors (cf. Richardson, 1972; Langlois and Robertson, 1995). As Argyres remarks (1996: 148), if one asks when competence is likely to vary independently across the vertical divide, one possible answer is in terms of the knowledge bases involved. If those knowledge bases are very similar, competence is likely to be highly correlated. For example, consider the case when both segments involve metal-working, as contrasted with the case where one involves metal-working and the other involves chemical processes. In this contrast, the knowledge correlation argument aligns with an argument based on the incentives to invest in, and leverage knowledge. When a downstream firm heavily involved in metal-working acquires a metal-working component supplier, it is making a related diversification move, perhaps leveraging its excellence in metal-working. It could happen that such a firm could prove to be much worse, or much better in component production than in its original business, but it seems unlikely – less likely than if the upstream firm's technology were based on organic chemistry. Finally, even in this simplified example, it is clear that while capability differences shape scope, the aggregate evolution of capabilities also depend on scope, and also that the choices on the individual level at any one period modify (and often restrain) the menu of choices any individual firm faces in future periods.

An Illustration of the Evolution of the ISP: Mortgage Banking in the US

We now examine an historical example -- the evolution of the US mortgage banking sector. In the early years following World War II, mortgage financing was a very localized and vertically integrated business. Integrated firms, especially savings and loan associations, used to produce (i.e., originate), hold, and service mortgage loans. In the late 1970s, this model of integrated housing finance started giving way to a vertically specialized one. The process as a whole included a number of discrete episodes that illustrate the mechanisms and causal paths discussed above. We focus on the development of vertical specialization between the two first parts of the mortgage loan origination process by mortgage banks: Retail production (identifying the lender, selecting the right customer, assessing the risk, approving the customer), and wholesale loan management (closing and funding the loan, and then keeping the loan in the bank's portfolio, for subsequent sale of the capital claim to the secondary market for loans).

In the early 1980s, several mortgage banks shifted from localized, small players to larger, national entities. As a result, the organization of work within them became increasingly modular. In the early days, mortgage banking executives were involved in all aspects of loan origination and servicing, as well as the process of selling the capital claim to the secondary market. As time progressed, however, firms found the need to create separate functions. The separation of activities in the mortgage banks was due to the perceived need for organizational sub-units that could manage their operations in a more autonomous way. The success of the firms that adopted this structure (as well as the imitation of these structures) led its proliferation in the industry. This illustrates our first evolutionary mechanism, as the logic of internal division of labor gave rise to the possibility for inter-firm specialization: With the separation of activities, the integrated mortgage banks increasingly became collections of neatly divided cost and profit centers. And with the creation of these centers, it became increasingly clear that integration had drawbacks. The divisions that over-performed the market ensured that their contribution got

27

recognized; and they tried to find a way to break free of the restrictions imposed by (what they saw as) their relatively inadequate counter-parts. So in the short run, a firm would identify that it was more profitable in, and could productively expand in, only *some* parts of the value chain, such as origination *or* warehousing.

These industry dynamics also illustrate our second mechanism, captured by the Fisher Theorem extension, at work. The increasing separation of the up-and down-stream, driven by the heterogeneity of capabilities of firms in the industry, led to a very drastic reduction of the variety in cost structures and to an intensification of competitive pressures, a pattern visible in industry statistics (Duncan, 1998). After the intermediate market was instituted, the inefficient origination divisions of large integrated firms were simply axed, and in-house loan production was substituted with purchased loan production, made by more efficient firms. This led to an improvement of productivity and reduction in cost, both in the up- and the down-stream segments, as the division of labor became more effective. In the relatively scalable warehousing and servicing segments, significant concentration came about, as the efficient firms were no longer held hostage to their slow origination growth, and as they competed out their less productive competitors.[18] This not only reduced costs, but also reasonably soon put pressure on margins too (Duncan, 1998). The firms that had pushed for the institution of dis-integration found themselves hurt by the departure of the least effective players in their segment, as they had inadvertently raised the bar of competition: Short-run profit-seeking undermined strategic viability.

Our third evolutionary mechanism, endogenous TC reduction, was also in operation in parallel to these competitive developments – indeed, it precipitated them. Given the short-term conditions of relative efficiency in the production process, the quest for market-mediated solutions had begun, with those firms with the highest productivity differentials pushing the

hardest. And it is clear that if there were firms that found themselves to be better than average in origination and worse than average in warehousing or servicing, other firms would realize the inverse; and this imbalance of capabilities led to the gradual increase of across-firm-boundaries transactions. They thus actively tried to find ways to make such trade possible, through efforts to standardize the requisite information (Jacobides, 2001, 2003).

Yet these underlying forces were necessary, not sufficient, drivers of the changes in scope. It was a "historical accident," which facilitated the learning process, that enabled the creation of specialization. As Garrett (1989) explains:

> In 1981 the first rumblings of change began, … the unraveling and unbundling of the housing finance delivery system... With no money with which to make mortgages, the S&L's started round after massive round of layoffs in their loan origination staffs. Even though many, if not most, originators had compensation schedules geared toward commissions, the nation's thrift industry could not afford even modest salaries.
>
> But as loan officers cleaned out their desks and headed for the door, they were often told something like the following: "Listen, Bob, you know we can't afford to keep you, you know that there's almost no business and that we have almost no money to lend. But we've always thought you put together high quality loans and we'd like to stay in touch. You have good contacts with Realtors, so if you can go out on your own and put together some loans, bring them to us and we'll fund them. And we'll figure out some way to split the loan fee." The seeds of a new world had been planted. The mortgage broker became even more useful with his increased number of sources; and then origination fees started to shrink for institutional lenders. Before long, our broker was approved to place loans with a great many lenders. Calling themselves wholesalers, lending institutions set up whole departments not to originate loans at the borrower level, but to take them in from originating brokers. (Garrett, 1989: 30-32)

The effect of such a contingency still depends, however, on the underlying conditions, as described in our fourth mechanism. In this case, differences in culture, incentives and knowledge bases militated against the combination of the activities in a single firm. In mortgage origination, integration into the local community and links with the builders (who help individuals find a package that finances their purchases) are the key success drivers; the mentality in running the business revolves around sales, and around client relationship management. The appropriate incentives are high-powered; agents are paid in commissions

---

[18] To wit, between 1989 and 1999, the top 10 servicers' share of the market rose from 11% to 41%, while the top

more than in salaries. The structure should be flexible and responsive, with a knowledge of the local markets, customers and products, and, most important, of how to convince. Wholesale, on the other hand, is an entirely different business. It consists of managing the pipeline of loans; of being able to guess how many loans that are made in a falling interest rate environment will fall off and never close; knowing when to sell these loans to the secondary market; and ensuring there is a good management of the different types of risks, including underwriting / default risk early on. Compensation is more stable, based on salaries; skills revolve around finance and risk management, and the types of successful executives are vastly different. Even more important than the differences in the appropriate managerial and incentive structures, improvements in wholesale build on an entirely different set of skills and knowledge than prevail in origination. While there still were significant transaction costs from dis-integrated production (as evidenced by the higher propensity of brokered loans as opposed to the in-house produced ones to default), specialization was preferred on the basis of reaping the dynamic benefits from the focused accumulation of knowledge that the specialized firm structure permits. Institutional separation enabled the specialized firms to improve and expand *faster* than integrated production, leading to the benefits that pushed toward specialization in the sector.

Overall, the industry witnessed an increasing separation between the strong originators, who would be good at the customer acquisition and loan closing, and the good wholesalers, who would focus on purchasing loans that were almost closed by others. This example also illustrates the structure of the causal mechanism and the interactions we proposed in section 3. Rather than considering the transactional environment as fixed, firms try to change it to capitalize (at least in the short run) on their advantage. So the menu of choices available and fixed in the short run (including transaction costs, or the set of potential competitors and hence the nature of the profitable activities and the amount of rent to be had) depend on the structure of production (capability differences).

---

10 originators' share went from 17% to 36% (*Inside Mortgage Finance*, 2000).

This case also shows how specialization affects the process of capability development: The creation of a distinct segment of brokers led to a change in the tools of marketing. Thus new entrants, such as real estate agencies, came to brokerage on the basis of their capabilities. Likewise, retail banks migrated into loan warehousing, which they could both understand and perform profitably on their basis of their experience, something that was not true of the previously integrated mortgage banking. So vertical specialization changed the "points of reference," the relevant knowledge base, as well as the types of industry participants who could, on the basis of their capabilities, now compete.

Discussion and Conclusion

We have set forth a dynamic view of vertical scope, in which the choices that firms make about their boundaries are conditioned by an evolving industry context. Five causal mechanisms of this co-evolutionary process were identified, and a key consideration brought to light – the extent to which relative competencies are different upstream and down. This dynamic analysis reveals important links between transaction costs and capabilities that go far beyond simply using transaction costs and capabilities in an additive function when assessing an individual firm's decision to make or buy (Williamson 1999). More broadly, we argue for a systemic view – a dynamic, evolutionary analysis of the institutional structure of production.

Our contribution has its roots in the work of Ronald Coase. Before moving into the implications of our analysis for strategy research, we first consider this intellectual lineage and clarify our debt to the Coasean heritage.

The Unexplored Legacy of Ronald Coase. This paper follows a part of Coase's legacy which has not been explored to date. Coase himself, at many points in his writings over the past several decades, has signaled an ambition for economic analysis that extends well beyond the scope of contemporary TCE. While the micro-analytics of transaction costs are a key tool, the broad purpose is to understand the determination of the overall level and composition of output

31

and the efficacy of the system's functioning. Coase has signaled a degree of frustration with the

narrow scope of the transaction cost economics tradition that derives from his 1937 paper:

> "But if one is concerned with the further development of the analysis, 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." (Coase, 1991: 65)

This resonates with his earlier observations on the nature and the structure of the economic

system, which have largely been unnoticed:

> "We would not expect firms to be similar in the range of activities that they
> embrace; but, so far as I am aware, the distribution of activities between firms is
> not something on which we have much to say." (1972: 65)
>
> "In my view, what is wanted in industrial organization is a direct approach to the
> problem. This would concentrate on what activities firms undertake, and would
> endeavor to discover the characteristics of the groupings of activities within
> firms. Which activities tend to be associated and which not?" (1972: 73)

On the basis of these and other comments, we can reasonably surmise that Coase's idea

of the appropriate domain of inquiry extends well beyond the logic of the governance of a

particular (type of) transaction, and even beyond the full delineation of the boundaries of the

individual firm, to the question of the boundaries of firms collectively and the allocation of

activities among them. It is the question of the institutional "packaging" of activity, which in the

Coasean perspective appears as inseparable from the question of how diverse types of exchanges

are conducted across the boundaries of the packages. These are the questions that an

explanation of the "institutional structure of production" has to address. But these two are also

inseparable from the question of *what actually goes on inside the packages* as well as the deeper

question of the ultimate sources of gains from specialization; these interrelated questions are all

aspects of the determination of the level and composition of output in the economic system.

As Coase suggested in his 1991 Nobel lecture, understanding the Institutional Structure

of Production, and its determinants, is a far broader task than the microanalysis of governance at

the level of the transaction. Our effort to illustrate the analysis of the ISP focused on the

determination of the vertical structure of an industry in the course of its evolution. This analysis vividly illustrates our major point about the challenge presented by the ISP: it demands that capabilities considerations and transaction cost considerations be placed side by side, and shown in interaction with each other in a historical process.

Implications for Strategy Research. There are implications for research in the capability- and resource- based view of the firm. Our argument is that transaction costs shape the trajectories of capability development; they determine the nature of the knowledge acquisition process, and quite possibly the type of competitors that can emerge in an industry, and thus the nature of the underlying resources and capabilities that can be leveraged for competitive advantage (Dierickx and Cool, 1989). Thus vertical specialization enables new types of references, and new competitors, to invade previously integrated sectors; it also calls upon incumbent firms to reconsider the appropriateness of their knowledge bases.

This analysis also clarifies that a drastic reduction of TC is not just a "disruptive technology," which will lead to the possible demise of incumbents (Christensen and Bower, 1996; Christensen et al, 2001, 2002). It suggests that the potential success of incumbents hinges on their relative capabilities in each part of the value chain. In a specialized world, incumbents may well thrive by retreating to the segment where they are competitively stronger -- a pattern we observe in a number of industries, including mortgage banking. This calls into question the generic characterization of "disruption," at least as far as value chain changes go. It also suggests that changes in integration may dramatically affect the patterns of competition inasmuch as they open up a sector to entirely new capability bases, which have been developed in different contexts. The interplay of changes in TC / scope and competitive dynamics / competitor identification are a promising area for future research.

On the level of strategy and policy, our analysis suggests that we must shift our focus from the individual decisions firms make, to the dynamics of industry evolution and the

mechanisms that change the business landscape. Executives and policymakers alike are increasingly interested in understanding the evolutionary dynamics of their business environment, in the interest of extending their limited foresight in dynamic settings. A focus on evolutionary mechanisms can be a welcome addition to our analytical arsenal. It is clear that some of the mechanisms we have identified deserve deeper exploration than we have been able to provide, while the accuracy of the general picture requires further, and broader assessment. Related research on the impacts of the co-evolution of scope and capabilities on rents and profits can complement this.

Finally, on the methodological level, our analysis suggests that while the factors driving individual decisions to integrate or not are important, an exclusive focus on them may block the way to understanding the evolution of the ISP. The analysis of the ISP evolution should complement our understanding of the individual transaction, much like the study of evolution should complement our knowledge of botany and zoology. Although great progress has been made in the micro-analytic understanding of firm choices of scope, Coase has pointed the way not merely in the analysis of that aspect of transaction costs, but in proposing the problem of understanding the systemic implications. Likewise, the analysis of capabilities should be done with reference to the conditions of vertical scope. We believe that we have made a useful start down the path he suggested, but there is clearly much more to be done.

References

Afuah, A. 2001. Dynamic boundaries of the firm: Are firms better off being vertically integrated in the face of a technological change? *Academy of Management Journal*, 44(6): 1211-1229.

Alchian, A.A. and H. Demsetz. 1972. Production, Information Costs and Economic Organization. *American Econmic Review* 62(5): 777-795.

Argyres, N.S. 1996. Evidence on the Role of Capabilities in Vertical integration Decisions, *Strategic Management Journal*, 17(2): 129-150.

Argyres, N. S. 1999. The impact of information technology on coordination: Evidence from the B-2 'Stealth' bomber. *Organization Science*,10(2): 162-181.

Argyres, N.S. and J.Liebeskind. 1998. Contractual Commitments, Bargaining Power and Governance Inseparability: Incorporating History into Transaction Cost Theory. *Academy of Management Review.*

Baldwin, C.Y. and K.B. Clark. 2001. *Design Rules: Modularity and Product Architecture.* MIT Press: Cambridge, MA.

Baldwin, C.Y. and K.B. Clark. 2003. *Where do Transactions Come From?* Working Paper, Harvard Business School, February 11.

Barzel, Y. 1982. Measurement Cost and the Organization of Markets. *Journal of Law and Economics* 25(1): 27-48.

Barney, J.B. 1991. Firm Resources and Sustained Competitive Advantage. *Journal of Management.* 17(1): 99-120.

Bowles S. and H. Gintis. 1988. Contested Exchange: Political Economy and Modern Economic Theory, *American Economic Review,* 78(2): 145-150.

Carroll, G.R., and M.T. Hannan. 1995. *Organizations in Industry: Strategy, Structure and Selection.* Oxford University Press: New York, NY.

Chandler, A. D. 1962. *Strategy and Structure*, MIT Press: Cambridge, MA.

Chandler, A.D. 1977. *The Visible Hand.* Harvard University Press: Cambridge, MA.

Christensen, C., and J. Bower. 1996. Customer Power, Strategic Investment, and the Failure of Established Firms. *Strategic Management Journal.* 17(3): 197 – 199

Christensen, C., M Raynor and M. Verlinden. 2001. Skate to Where the Money Will Be. *Harvard Business Review.* (11): 72-82.

Christensen, C., M. Verlinden and G. Westerman. 2002. Disruption, Disintegration and the Dissipation of Differentiability. *Industrial & Corporate Change.* 11(5): 955 – 994.

Coase, R.H. 1937. The Nature of the Firm. *Economica* 4:386-405.

Coase, R.H. 1972. Industrial Organization: A Proposal for Research, in V.R. Fuchs (Ed.), *Policy Issues and Research Opportunities in Industrial Organization.* NBER: New York, NY.

Coase, R.H. 1992. The Institutional Structure of Production. *The American Economic Review* 82(4): 713-719.

Cusumano, M. , Y. Mylonadis and R. Rosenbloom. 1992. Strategic Maneuvering and Mass - Market Dynamics: The Triumph of VHS Over Beta. *Business History Review*.

Dierickx, I., and K. Cool, 1989. Asset Stock Accumulation and Sustainability of Competitive Advantage. *Management Science* 35(12): 1504-1514.

DiMaggio, P. and Powell, W. 1983. The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. *American Sociological Review* 48: 147-160.

Duncan, D.G. 1998. Relative Profitability of the Mortgage Industry. *Mortgage Banking*. June: 95.

Fayol, H. 1949. [1919]. *General and Industrial Management.* Pitman: London.

Fine, C. 1998. *Clockspeed.* Proteus Books.

Fisher, R.A. 1958 [1929]. *The Genetical Theory of Natural Selection.* Dover: New York.

Garrett, J. 1989. A New Paradigm for Lenders. *Mortgage Banking* May: 30-41.

Gould, S.J. 1980. *The Panda's Thumb: More Reflections in Natural History.* Norton: New York, NY.

Gould, S.J. and Vrba, E.S. 1981. Exaptation, A Missing Term in the Science of Form. *Paleobiology* 8: 415.

Hannan, M.T. and Freeman, J. *Organizational Ecology.* Harvard University Press: Cambridge, MA.

Hart, O. 1995. *Firms, Contracts and Financial Structure.* Oxford University Press: New York, NY.

Hoetker, G. 2003. *How much you know versus how well I know you: Selecting a supplier for a technically innovative component.* Working Paper, University of Illinois.

Iwai, K. 1984. Schumpeterian Dynamics, Part II: Technological Progress, Firm Growth and Economic Selection. *Journal of Economic Behavior and Organization* 5: 321-351.

Jacobides, M.G. 2001. Mortgage Banking Unbundling. *Mortgage Banking* 61: 28-39.

Jacobides, M.G. 2002. *Modeling Strategic Dynamics: Transaction Costs, Capabilities and Scope* Working Paper 01-16, Financial Institutions Center, The Wharton School of the University of Pennsylvania.

Jacobides, M.G. 2003. *How do Markets Emerge? Organizational Unbundling and Vertical Dis-Integration in Mortgage Banking,* Working Paper, Centre for the Network Economy, London Business School, May.

Jacobides, M.G., and D.C. Croson. 2001. Information Policy: Shaping the Value of Agency Relations. *Academy of Management Review* April.

Jacobides, M.G. and L. M. Hitt. 2001. *Vertical Scope, Revisited: Transaction Costs vs Capabilities and Profit Opportunities in Mortgage Banking* Working Paper 01-17, Financial Institutions Center, Wharton School.

Klepper, S. 1997. Industry Life Cycles. *Industrial and Corporate Change* 6(1): 145-182.

Klepper, S., 2002, Firm Survival and the Evolution of Oligopoly, Vol.33, No.1, Spring 2002, pp. 37-61.

Kogut, B. and U. Zander. 1996. What Firms Do? Coordination, Identity, and Learning. *Organization Science* 7(5): 502-518

Langlois, R.N. 1992. Transaction Cost Economics in Real Time. *Industrial and Corporate Change* 1: 99-127

Langlois, R. N. 2003*. The vanishing hand*: *the changing dynamics of industrial capitalism.* Industrial and Corporate Change 12(2):351-385

Langlois, R.N., and P.L. Robertson. 1989. Explaining Vertical Integration: Lessons from the American Automobile Industry. *Journal of Economic History* 49(2): 361-375.

Langlois, R.N., and P.L. Robertson, 1995. *Firms, Markets, and Economic Change: A Dynamic Theory of Business Institutions.* Routledge: London.

Lawrence, P. and J. Lorsch. 1967. *Organization and Environment.* Harvard University Press: Cambridge MA.

Lazonick, W. 1991. *Business Organization and the Myth of the Market Economy.* Cambridge University Press: New York, NY.

Levinthal, D.A. 1997. Adaptation in Rugged Landscapes. *Management Science* 43(7): 934-950.

Lieberman, M.B., and R. Dhawan. 2001. *Assessing the Resource Base of Japanese and U.S. Auto Producers: A Stochastic Frontier Production Function Approach.* UCLA Working Paper. Los Angeles, CA.

Madhok, A. 2002. Reassessing the Fundamentals and Beyond: Ronald Coase, the Transaction Cost and Resource-Based Theories of the Firm and the Institutional Structure of Production. *Strategic Management Journal* 23: 535-550.

Marshall, A. 1948. *Principles of Economics*, 8th ed. Macmillan: New York, NY.

Metcalfe, J. S. 1998. *Evolutionary Economics and Creative Destruction.* Routledge: London.

Meyer, J. and B. Rowan. 1977. Institutionalized Organizations: Formal Structure as Myth and Ceremony. *The American Journal of Sociology* 83: 340-363.

Nelson, R. R., and S. G. Winter. 1982. *An Evolutionary Theory of Economic Change.* The Belknap Press of Harvard University Press: Cambridge, MA.

Nelson, R.R. and S. G. Winter. 2002. Evolutionary Theorizing in Economics. *Journal of Economic Perspectives* 16: 23-46.

North, D.C. 1986.The New Institutional Economics. *Journal of Institutional and Theoretical Economics* 142: 230-327.

Polanyi, K. 1957. *The Great Transformation.* Boston: Beacon.

Poppo, L. and T.R. Zenger. 1998. Testing alternative theories of the firm: Transaction cost, knowledge-based, and measurement. *Strategic Management Journal*, 19(9): 853-878.

Porter, G. and H.C. Livesay. 1971. *Merchants and Manufacturers.* John Hopkins: Baltimore, MD.

Porter, M.E. 1996. What Is Strategy. *Harvard Business Review* 74: 61-78.

Prahalad, C.K. and R. Bettis. 1986. The Dominant Logic: A New Linkage between Diversity and Performance. *Strategic Management Journal.* 7: 485-501.

Rivkin, J. 2001. Reproducing Knowledge: Replication Without Imitation at Moderate Complexity. *Organization Science* 12: 274-293.

Schilling, M.A. and H. Steensma. 2001. The Use of Modular organizational Forms: An Industry-Level Analysis. *Academy of Management Journal* 44 (6): 1149-1169

Siggelkow, N. 2001. Change in the Presence of Fit: The Rise, the Fall, and the Renaissance of Liz Claiborne. *Academy of Management Journal* 44: 837-858.

Silver, M. 1984. *Enterprise and the Scope of the Firm*. Martin Robertson: London.

Smith, A. 1776. *The Wealth of Nations.* Glasgow.

Stigler, G. 1951. The Division of Labor is Limited by the Extent of the Market. *Journal of Political Economy* 59 (June): 185-93.

Teece, D.J., R. Rumelt, G. Dosi and S. Winter. 1994. Understanding Corporate Coherence: Theory and Evidence. *Journal of Economic Behavior and Organization* 23: 1-30.

Teece, D.J., G. Pisano, and A. Shuen. 1997. Dynamic Capabilities and Strategic Management. *Strategic Management Journal* 18(7): 509-533.

Utterback, J.M and W. Abernathy. 1975. A Dynamic Model of Process and Product Innovation. *OMEGA.* 3: 639-655.

Utterback, J.M. 1994. *Mastering the Dynamics of Innovation*. Harvard Business School: Cambridge MA

Walker, G., and D. Weber. 1984. A Transaction Cost Approach to Make-or-Buy Decisions, *Administrative Science Quarterly,* 29(3): 373-392.

Williamson, O. E. 1975. *Markets and Hierarchies: Analysis and Antitrust Implications.* Free Press: New York, NY.

Williamson, O. E. 1985. *The Economic Institutions of Capitalism.* Free Press: New York, NY.

Williamson, O.E. 1999. Strategy Research: Governance and Competence Perspectives. *Strategic Management Journal* 20(12): 1087-1108.

Winter, S.G. 1984. Schumpeterian Competition in Alternative Technological Regimes. *Journal of Economic Behavior and Organization* 5(Sept-Dec): 287-320.

Winter, S.G. 1988. On Coase, Competence, and the Corporation. *Journal of Law Economics and Organization* Vol. 1.

Winter, S.G. 1990. Survival, Selection and Inheritance in Evolutionary Theories of Organization, in J.V. Singh, Ed., *Organizational Evolution: New Directions.* Sage: Newbury Park, CA.

Winter, S.G. 1995. The Four R's of Profitability, in C. Montgomery, Ed., *Evolutionary and Capability-Based Views: Towards a Synthesis.* Kluwer Academic Publishers: Boston, MA.

Table 1:
The Evolutionary Mechanisms of the Institutional Structure of Production: How Capabilities and Transaction Costs Co-evolve

Mechanism	Causal Process	Result
1. *Intra-organizational Creation of a Vertical Divide*	**Clearer administrative separations along the value chain within a firm make vertical break-up more plausible**	**Clear and similar vertical separations within firms prompt consideration of using the market**
2. *Capability Differences and Selection Effects on Scope*	**Differential profitability arising from heterogeneous capabilities shapes firm growth and imitation processes.**	**Selection increases scales of firms, or specific units, prompts imitation of practices, and reduces heterogeneity in all things, including vertical structure**
3. *Endogenous Transaction Cost Reduction*	**Capability differences between firms up- vs. down-stream provide incentives to reduce TC, turning latent into realized gains from trade**	**Reduction of transaction costs and initiation of contracting, *provided* capability differences provide incentives**
4. *Impact of Knowledge Accumulation and Capability Dynamics on Scope*	**Differences in management styles or structures, as well as differences in knowledge base, enable better capability development for specialists and make vertical separation desirable**	**Centrifugal forces to specialize *to the extent* that there are better opportunities to improve or draw upon specialized knowledge when vertically specialized**
5. *Impact of Changing Vertical Scope on the Capability Development Process*	**Specialization means that new knowledge bases become appropriate; relevant expertise is outside the industry as currently defined**	**Changes in relevant capability pools; knowledge bases; or players** **Changes in industry definitions and competitor / capability profile**

Note that the above mechanisms interact in specific ways – Figure 1 provides an overview of the systemic nature of these interactions

Figure 1:
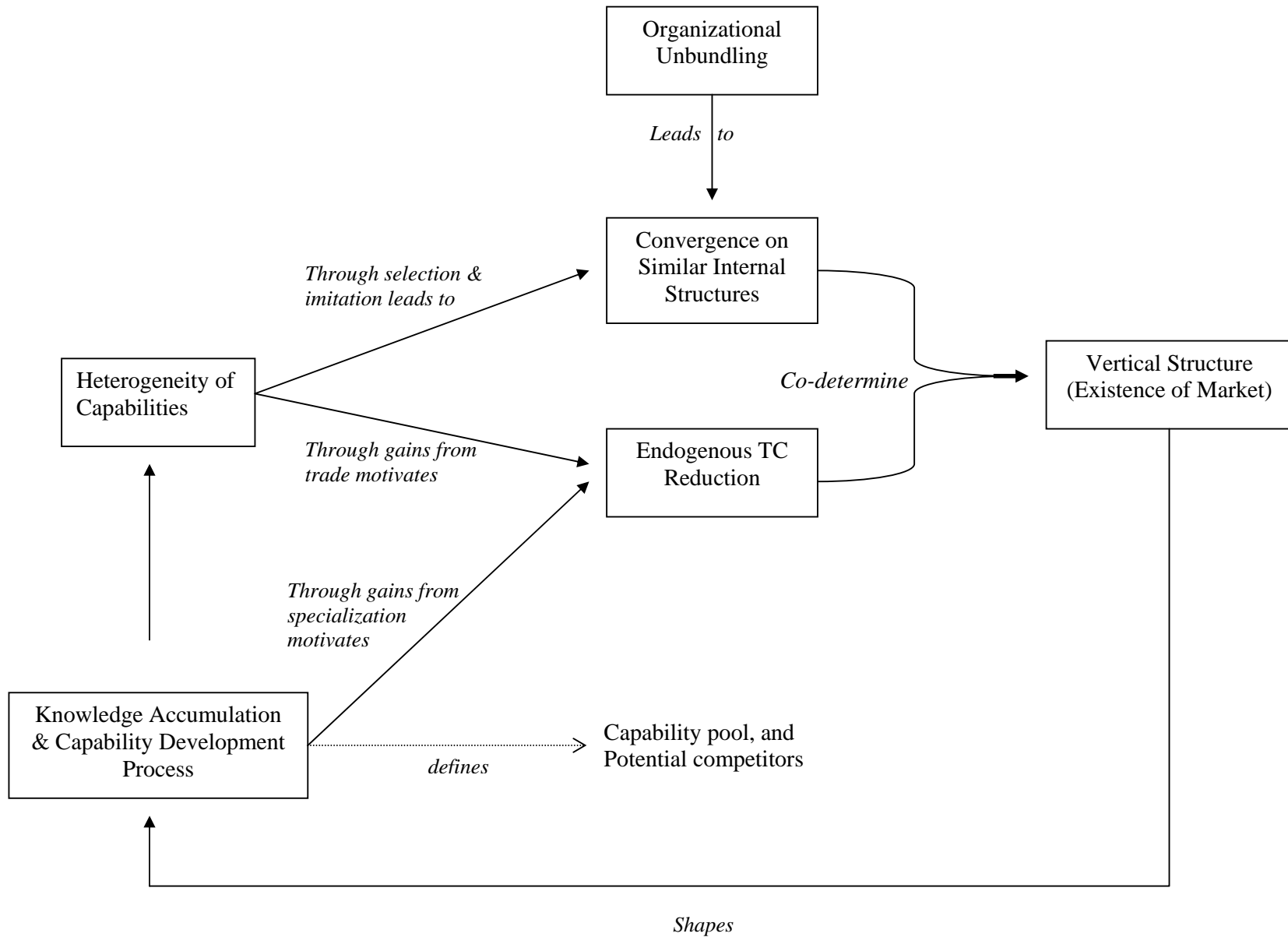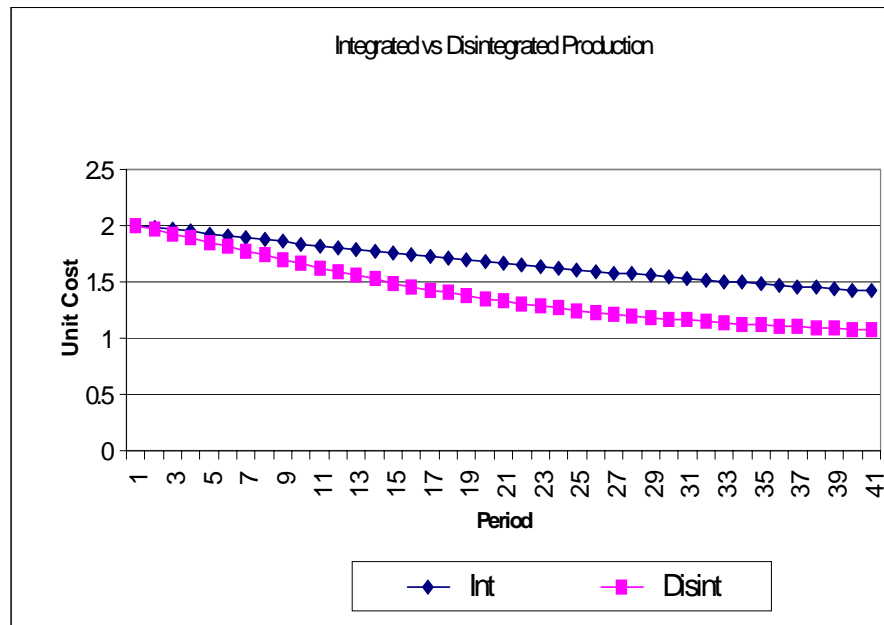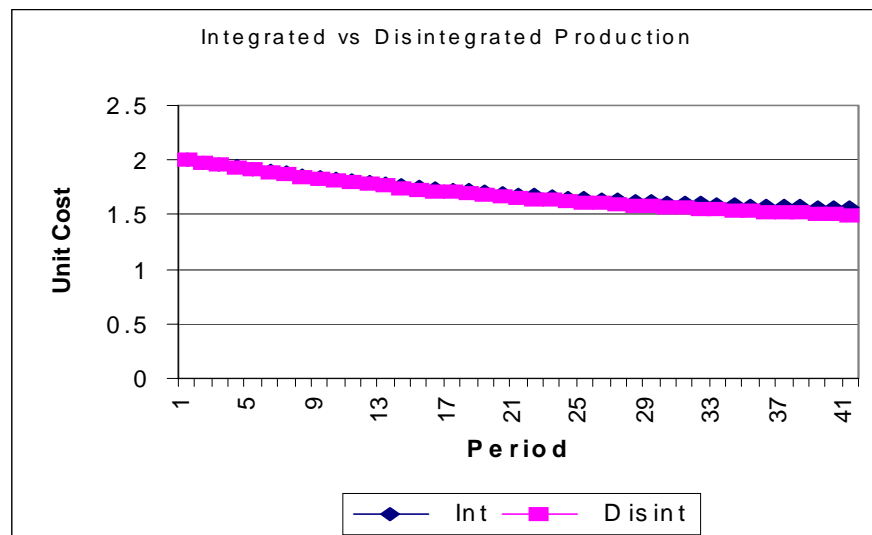**The Simplified Model of Capability and Transaction Cost Co-evolution**

Figure 2:
Evolution of Industry Cost When Capabilities (Up- vs Down-stream) are Independent

Correlation upstream- downstream: .004

Figure 3:
Evolution of Industry Cost When Capabilities (Up- vs Down-stream) are Correlated

Correlation upstream- downstream: .84

<u>**Appendix**</u>

Extending the Fisher Fundamental Theorem

-- Comparing Specialization and Integration as they Shape Capability Evolution --

We present a simple model of the evolution of industry average cost in an industry with vertical segments.[i] We initially consider the results in the case where firms have to be integrated, and then consider how these compare with the dis-integrated case.

Production, Cost and Growth Assumptions in the Integrated Case

We assume that each firm has a single technique upstream and another downstream. It is important to take some care with the units of measurement here, partly because the effects we are concerned with depend on the relative importance of the upstream and downstream segments, and also to facilitate interpretation of the numerical examples that follow. We take as given the unit of the final product, the currency unit (a dollar), and the time unit (a year). Choose units for the product of the upstream segment, the intermediate good, such that production of one unit of the final good requires one unit of the intermediate good. Similarly, choose units for capital equipment such that one unit of upstream capacity supports production of one unit of final product per year (via one unit of intermediate product), and similarly downstream a unit of capacity supports production of a unit of final product per year. We assume that these types of capacity are subject to proportional depreciation at instantaneous rates of d^U and d^D, per year, respectively. The prices of the two types of capacity are v^U and v^D respectively, and r represents the cost of capital (as an instantaneous interest rate). Let b denote the service price of capital services, per unit final product. The service price is the sum of interest and depreciation on the value of the capital required per unit, upstream and down. That is,

(0.6) $$b = b^U + b^D = (r + d^U) \cdot v^U + (r + d^{D)} \cdot v^D$$

I

Our key simplifying assumption is that firms are identical with respect to their capital costs per unit – signaled by the absence of subscript j in (0.6). Variable costs differ across firms; let

(0.7) $$c_j = c_j^U + c_j^D$$

be the unit cost of variable inputs for firm j, the sum of the upstream and downstream costs. Variable input flows are matched to output rates, which are determined by capacity levels. We assume that there are N firms, N>1, with different costs. The price P of the final product is determined at each point of time by a downward-sloping industry demand curve (which, however, turns out to play no significant role). Denoting the net profit of firm j by π_j, we have

(0.8) $$\pi_j = (P - b - c_j) \cdot K_j$$

where K_j denotes the capital stock of firm j, consisting of pairs of upstream and downstream units.

We further assume that firm profitability drives firm growth by providing the necessary capital to engage in re-investment, and this provides us with the simple expression

(0.9) $$\frac{dK_j}{dt} = \pi_j / (v^U + v^D)$$

Evolution of Industry Production: Cost, Shares, Growth under Integration

We then define the industry total capital stock, $K(t) = \sum K_j(t)$, the

shares $s_j(t) = K_j(t) / K(t)$, and the industry average cost $c(t) = \sum s_j(t) \cdot c_j$ Then the

evolution of the relative shares in the production will be

$$(0.10) \qquad \frac{ds_j(t)}{dt} / s_j = (\frac{dK_j}{dt}) / K_j - \frac{dK}{dt} / K \,,$$

which, on the base of the previous equations then becomes

$$(0.11) \qquad \begin{aligned} \frac{ds_j}{dt} / s_j &= (v^U + v^D)^{-1} \cdot [(P - b - c_j) - \sum s_j(t) \cdot (P - b - c_j)] \\ &= (v^U + v^D)^{-1} \cdot [c(t) - c_j] \end{aligned}$$

Since P and b drop out of this expression, the evolution of the shares is independent of the evolution of industry output, price and capital costs. This simplification reflects the assumption of equal capital intensities across firms. When that assumption is abandoned, two firms with identical unit costs need not have identical growth rates. Specifically, positive profitability induces higher growth in the less capital intensive firm, although at equilibrium the growth rates are equated at zero.

Equation (0.11) provides the basis for computing the rate of change of industry

average cost as $\dfrac{dc}{dt} = \sum \dfrac{ds_j}{dt} \cdot c_j \,,$

$$(0.12) \qquad \frac{dc}{dt} = (v^U + v^D)^{-1} \cdot \sum s_j \cdot (c - c_j) \cdot c_j = (v^U + v^D)^{-1} \cdot [\sum s_{j.}(c - c_j) \cdot (c_j - c)]$$

The last equality reflects the fact that the sum of share-weighted deviations from the share-weighted mean is zero. This gives the Fisher theorem result

$$(0.13) \qquad \frac{dc}{dt} = -(v^U + v^D)^{-1} \cdot Var(\mathbf{c})$$

The notation $Var(\mathbf{c})$ is here to be understood as the share-weighted cross-sectional variance of the vector $\mathbf{c} = (c_1, ..., c_N)$, the descriptive statistic computed with the share weights of that point of time.[ii] It is straightforward that replacing the

assumption that net investment equals economic profit by an assumption that investment

is some fraction of α of economic profit, $0 \leq \alpha \leq 1$, the result changes to

$\dfrac{dc}{dt} = -\alpha \cdot (v^U + v^D)^{-1} \cdot Var(\mathbf{c})$. The effect of lowering α corresponds to the effect of raising

the price of capital equipment, it increases the amount of profit a firm must earn

to finance a given increase of capital. Allowing α to vary across firms produces results

equivalent to those of different capital intensities.

Comparing Integrated and Specialized Industry Results

We now perform the thought experiment of an instantaneous dis-integration of the industry

and consider how the Fisher expression changes at that point of time. Market shares

upstream and downstream are initially equal to each other and to the share that prevailed

before dis-intergration. In general, the upstream and downstream segments may differ in

terms of both the prices of capital goods and the capital service prices. If we retain the

assumption the investment equals net economic profit in each segment, the price in the

market for intermediate product plays the role of bringing about the profit rates that equate

the growth rates of the upstream and downstream segments so that capacity remains

balanced. With the price P of the final product determined from the demand curve at the

prevailing capacity levels, the price of the intermediate product $P*$ is determined by the

growth rate condition,

(0.14)
$$\frac{dK^U}{dt} = \frac{dK^D}{dt}$$

Manipulations analogous to those above, applied to the two segments in turn, reveal this

condition to be

(0.15)
$$(v^U)^{-1} \cdot [(P*-b^U) - c^U(t)] = (v^D)^{-1} \cdot [(P - P* - b^D) - c^D(t)]$$

Here, $c^U(t)$ and $c^D(t)$ are respectively the upstream and downstream share-weighted

industry average variable costs; in general, the shares are different upstream and down.

As in the analysis of the integrated condition above, the equality of capital costs

across firms implies that that the industry average cost evolves independently of output

and price – here, independent of $(P*-b^U)$ upstream and of $(P-P*-b^D)$ downstream.

Thus the path of the analysis is the same as before, and we can thus compute

(0.16) $$\frac{dc^U}{dt} = (v^U)^{-1} \cdot Var(\mathbf{c^U})$$

(0.17) $$\frac{dc^D}{dt} = (v^D)^{-1} \cdot Var(\mathbf{c^D})$$

Finally, note that function $c(t)$ characterized by (1.8) is not the same as the function $\overset{!}{c}(t)$ based on (1.11) and (1.12) and defined by $\overset{!}{c}(t) = c^U(t) + c^D(t)$. Because of the different dynamics under the integrated and market- mediated conditions, the causal mechanisms are different.

---

[i] For a more comprehensive modeling analysis, which is based on computational dynamic general equilibrium of an economy with two sectors, different resources, and explicit re-ivestment modeling, see Jacobides (2002). That paper considers the micro-economic drivers and the competitive implications of changing vertical scope, and examines more comprehensively how TC and capabilities interact dynamically to affect both. The model in this paper formalizes a more limited, specific set of insights.

[ii] To confirm that the above expression survives a units check, it is necessary to allow for the existence of an invisible numeral one in the denominator , a physical capital-output ratio with the units "equipment units/(output units/year)" – implied by the units choices described above.