Transaction-cost Economics in Real Time

RICHARD N. LANGLOIS
(The University of Connecticut U63, Storrs, CT 06269-1063, USA)

This paper attempts to place the theory of the boundaries of the firm within the context of the passage of time. More precisely, it resurrects and places in a modern frame some of the insights of the classical and Marshallian theories of organization. The modern reinterpretation of those theories centers around the 'capabilities' view of the firm. Taken together with governance costs, the capabilities of firm and market determine the boundaries of the firm in the short run. Over time, capabilities change as firms and markets learn, which implies a kind of information or knowledge cost—the cost of transferring the firm's capability to the market or vice versa. These 'dynamic' governance costs are the costs of persuading, negotiating and coordinating with, and teaching others. They arise in the face of change, notably technological and organizational innovation. In effect, they are the costs of not having the capabilities you need when you need them. Dynamic transaction costs provide an explanation for vertical integration that is arguably more general than those dominant in the literature. In the face of uncertainty and divergent views of the future, common ownership of multiple stages of production is a superior institutional arrangement for coordinating systemic change. Asset-specificity is neither necessary nor sufficient for this to be true. Dynamic governance costs may also afflict internal organization. It may sometimes be costly—in terms of persuasion, negotiation and teaching—to create within the firm capabilities readily available on the market. Indeed, in cases in which systemic coordination is not the issue, the market may turn out to be the superior institution of coordination. In general, the capabilities view of the firm suggests that we look at firm and market as alternative—and sometimes overlapping—institutions of learning.

1. *Transaction costs in the long run and the short*

Classical and neoclassical perspectives

One of the crucial ways in which classical economics differed from neoclassical was in its preoccupation with costs of production. In value theory, the interpretation runs along the following lines. The classicals were interested in the long run. And in the long run, all factors are variable, implying production

Industrial and Corporate Change Volume 1 Number 1 1992

at constant returns to scale. In such a world, supply-side factors—and not demand—do indeed determine value. Although not widely remarked on, there is also, I believe, an organizational corollary to this interpretation of the classicals. Because it takes a long-run perspective, the classical theory of organization is preoccupied almost entirely by production costs and largely ignores transaction costs. As a result, the classical theory tells us much about the organization of production in the economy. But it also tells us less than we want to know about the boundaries of the firm, that is, about the ownership of the various stages of production and the nature of the contractual relationships among them.

The fountainhead of the classical theory of the organization of production is, of course, Adam Smith's discussion of the benefits of the division of labor.[1] One can imagine the economies attendant upon that organizational innovation as taking place within the boundaries of the firm. But it is also possible for a subdivided stage to be spun off to become what Smith would have called a 'peculiar trade' of its own. Consider, as a historical example, the development of the American machine-tool industry (Rosenberg, 1963). Before 1840, textile firms made their own machine tools as needed: in effect, the manufacture of such tools was a stage in the production of cloth into which the textile firms were integrated vertically. As the demand for final products grew, the demand for machine tools—from the textile industry and elsewhere—increased to an extent that the textile machine shops could spin off and become independent machine-tool firms. There is, however, nothing in the classical story to tell us whether such spinning off will occur or whether the division of labor will proceed under the umbrella of a single firm.

Since Coase (1937), economists have begun to explain observed patterns of ownership and contract by their ability to minimize the sum of production costs *and* transaction costs.[2] If my corollary is right, however, this modern—shall I call it neoclassical?—theory of the boundaries of the firm is necessarily a short-run theory. Transaction costs are essentially short-run phenomena. This does not by any means make such costs unimportant. One cannot explain ownership and contracting structures without them. But the modern focus on transaction costs, salutary as it has been, has nonetheless put into the background the richness of the classical cost-of-production theory.

As I phrased it above, the long run is the period over which all costs are variable costs. A couple of points are worthy of note. First of all, the standard

[1] For excellent modern discussions of the classical theory, see Stigler (1951), Ames and Rosenberg (1965) and Leijonhufvud (1986).

[2] The methodological issues surrounding this assertion are in fact somewhat complex. For an introduction, see Langlois (1984, 1986). For present purposes, however, I will not poke directly into its explanatory merits.

concept of the runs is—almost paradoxically—a timeless notion. That is, the time that passes between the short run and the long run is what Mark Blaug (1987, p. 371) calls 'operational time' rather than real time.[3] The length of the run is defined entirely in terms of the variability of factors, not in terms of the external standard of a clock. The long run may come about in a week in some industries and a century in others.

Although this mechanical conception of the run is normally described as Marshallian, it was not in fact the way Marshall himself understood the concept (Currie and Steedman, 1990, pp. 22–28). As he tells us in the preface to the eighth edition of the *Principles*, his use of static models is a matter of convenience rather than conviction, something appropriate to a textbook introduction. 'The Mecca of the economist,' he says, 'lies in economic biology rather than in economic dynamics.' As Brian Loasby (1989, 1990) has argued, Marshall's vision of economic progress was basically a Smithian one, overlain with this biological metaphor. The Smithian process of progressive specialization is not an economic process merely but a process characteristic of nature in its broadest. It is, Marshall says, a

> general rule, to which there are not very many exceptions, that the development of the organism, whether social or physical, involves an increasing subdivision of functions between its separate parts on the one hand, and on the other a more intimate connection between them. Each part gets to be less and less self-sufficient, to depend for its wellbeing more and more on other parts . . . This increased subdivision of functions, or "differentiation," as it is called, manifests itself with regard to industry in such forms as the division of labour, and the development of specialized skill, knowledge and machinery: while "integration," that is, a growing intimacy and firmness of the connections between the separate parts of the industrial organism, shows itself in such forms as the increase of security of commercial credit, and of the means and habits of communication by sea and road, by railway and telegraph, by post and printing press. (Marshall, 1961, IV.viii.1, p. 241).

Economic progress, then, is for Marshall a matter of improvements in knowledge and organization as much as a matter of scale economies in the neoclassical sense. We can see this clearly in his 'law of increasing return,' which is distinctly not a law of increasing returns to scale: 'An increase of labour and capital leads generally to improved organization, which increases the efficiency of the work of labour and capital' (Marshall, 1961, IV.xiii.2, p. 318). And, in arguing that long-run marginal cost is falling with increases in output, he suggests that we 'exclude from view any economies that may result from substantive new inventions; but we include those which may be

[3] My use of the phrase 'real time' is inspired by O'Driscoll and Rizzo (1985).

expected to arise naturally out of adaptations of existing ideas' (Marshall, 1961, V.xii.3, p. 460).

To say that a movement to the long run involves progressive changes in organization and knowledge is really to suggest an interpretation quite different from the standard neoclassical conception, in which substitution is supposed to take place with knowledge held constant. Adopting this learning-and-organization view, I argue, implies a shift to a real-time conception of the long run. In some sense, the long run is the period over which enough learning has taken place that adjustments are small and come only in response to foreseeable changes in exogenous conditions.[4]

Transaction costs in the short run

My contention is that transaction costs lose their importance in *this* kind of long run. To the extent that transaction costs are 'frictions'—a term one often hears applied—then such costs are bound to diminish over time with learning, all other things equal. In order to make this case, however, we need to examine the nature of transaction costs in more detail.

Alchian and Woodward (1988) have recently argued that there are two distinct traditions in transaction-cost analysis. 'One emphasizes the administering, directing, negotiating, and monitoring of the joint productive teamwork in a firm. The other emphasizes assuring the quality or performance of contractual agreements' (Alchian and Woodward, 1988, p. 66). The former is what we might call the measurement-cost view. The latter we may call the asset-specificity view. Looked at in the right way, however, these two traditions yield strikingly similar conclusions.

The basic notion of the measurement-cost approach is that it is often costly to measure the quality and sometimes even the quantity of the output of a stage of production (Barzel, 1982; Cheung, 1983). In the best-known example of this approach (Alchian and Demsetz, 1972), indivisibilities in team production lead to shirking that is costly to detect, suggesting a rationale for a residual claimant to hire and monitor the team members. More recently, Barzel has provided a more general theory of how measurement costs affect organizational form. He suggests that 'among factors contributing to the value of common effort, the greater the difficulty in measuring one factor's contribution vis-à-vis that of others, the more likely is the owner of that factor to assume the position of the residual claimant' (Barzel, 1987, p. 105). Since the factor least easily measured is most tempted

[4] This conception of the long run is similar to what Schumpeter (1934) called 'the circular flow of economic life.'

to moral hazard, output is maximized when that factor becomes the principal, leaving the less-costly factor(s) to be agents.

The source of inefficiency in the asset-specificity approach is the tussle for rents after the fact of a contract (Klein *et al.*, 1978; Williamson, 1985). Such distributive battles can arise only when there are assets that cannot be redeployed costlessly if the contractual arrangement were to end. As a result, the costs to which the asset-specificity view looks are costs of post-contractual 'hold up' rather than post-contractual moral hazard. As in the measurement-cost view, however, the possibility of these postcontractual costs leads to the selection (evolution?) of institutional forms *ex ante* that mitigate the 'opportunism' *ex post*.

The similarity becomes clearer when we look at one set of models that allies itself with the asset-specificity camp, namely the literature on incomplete contracts (Grossman and Hart, 1986; Hart, 1988, 1989). Here contracts are incomplete in that, because the parties are 'boundedly rational,' the contract cannot provide in advance for all the contingencies that might occur, at least not in adequate detail. There are thus two types of rights a contract can allocate: specific rights and residual rights. Specific rights are those spelled out, while residual rights are the 'left over' rights to control in any circumstances not specifically provided for. The possession of residual rights over an asset is what we mean by ownership of that asset. Thus a theory of the allocation of residual rights is a theory of the ownership of assets, which is a theory of the boundaries of the firm.[5]

What is interesting about this literature is the broad similarity of its conclusions to those of Barzel.[6] When contingencies can be adequately specified, or when the decisions of the cooperating parties do not affect one another, contracts are possible and integration is unnecessary. When the decisions are linked, however, that party should get the residual rights whose decisions are more important to the joint enterprise. As in Barzel, there is a kind of Coase-theorem result at work: the residual rights end up in the hands of the party whose possession of them maximizes the joint surplus. There are also a couple of other ways in which these results are more similar to those of Barzel than initially appears. First of all, the concept of an agency cost—which is what really underlies Barzel's notion of measurement cost—

[5] I should note that there is a raging controversy over the definition of a firm. One view, championed by the incomplete-contract theorists (see for example Hart, 1989, esp. 1771–73) is that the firm ought to be defined by the common ownership of tangible assets. The opposing view is that the firm is a 'nexus of contracts' and that ownership of physical assets is irrelevant (Cheung, 1983). The two are not always incompatible, and I certainly agree that one needs to pay attention to the ownership of 'organizational assets' (Klein, 1988) as well as physical assets. (For this reason I will refer to 'ownable' assets rather than physical assets.) But in the end I side with Hart in thinking that the 'property rights' approach, as he correctly calls it, is both the most appealing and the most precise one to take.

[6] A point Barzel notes (1987, p. 105n).

includes not only the technological costs of monitoring but also any residual loss of value that comes from a misalignment of the agent's incentives with those of the principal (Jensen and Meckling, 1976). Thus, to say that the owner of the residual right should be the party with the higher monitoring cost is in fact to say that that owner should be the party whose decisions are more important to the value of the enterprise. Second, the Grossman-and-Hart model relies on hold-up costs: the inefficiencies arise from non-optimal choices of specific investment *ex ante* in the face of a distributive game *ex post*. But the 'property rights' approach, as Hart (1989) aptly calls this incomplete-contracts theory, is in fact more general than this. Indeed, one of the models in Hart (1988) relies on moral hazard *ex post* rather than hold-up.

Transaction costs in the long run

F. A. Hayek (1945, p. 523) once wrote that 'economic problems arise always and only in consequence of change.' My argument is the flip-side: as change diminishes, economic problems recede.[7] Specifically, as learning takes place within a stable environment, transaction costs diminish. As Carl Dahlman (1979) points out, all transaction costs are at base information costs. And, with time and learning, contracting parties gain information about one another's behavior. More importantly, the transacting parties will with time develop or hit upon institutional arrangements that mitigate the sources of transaction costs.

The incomplete-contracts framework makes this argument particularly clear. The reason contracts are incomplete is because of 'bounded rationality,' a somehat misleading expression that better captures the limitedness of the agent's knowledge and decision-making skills than it does imperfection in the agent's rationality. In other words, then, contracts are incomplete because of limitations of knowledge. With time, however, agents engaged in similar transactions will learn the typical outcomes of those transactions and will include increasingly more specific provisions in their contracts. As a result, a progressively greater part of the transactions can be handled through specific rather than residual rights. More concretely, with repeated transactions in a stable environment one can expect (1) contracts to become 'self-enforcing' because of reputation effects and (2) hold-up and moral-hazard problems to be attenuated by the evolution of norms of reciprocity and cooperation (Axelrod, 1984; Sugden, 1986). There is also another aspect to the argument. If the environment is genuinely one in which change is

[7] For another version of this argument, see Langlois (1984). In that article, I tried to make the case that transaction costs are ultimately the product of radical or 'structural' uncertainty. I still think that argument valid, but I will not push that language here.

diminishing, then it is also one in which behavior must be becoming increasingly routine. And routine behavior is necessarily easier to monitor and measure than non-routine behavior. In an environment in which change is absent, the 'plasticity'[8] necessary for moral hazard is also absent. For all these reasons, one would expect transaction costs to play a small role in the long run.

This is not immediately to say that the long run favors vertical disintegration, although there is obviously some reason to follow the likes of Allyn Young (1928) and George Stigler (1951) in this direction. It may well turn out that one of the institutional responses to the moral-hazard or hold-up costs of the short run is in fact vertical integration, that is, common ownership of the ownable assets of adjacent stages of production. If this happens, subsequent organizational learning would take place (at least initially) within the framework of the firm, which may well affect the long-run pattern of integration. To put this another way, the result of a learning process of this sort depends in general not only on the present state of the system but also on the past states through which the system has traveled (Hayek, 1967, p. 75). Such a system may display some of the properties of 'lock-in' that Paul David (1985) and others have discussed. As a result, bursts of economic change may leave their mark in the long run (Langlois, 1984, 1988).

Thus, my point here is not that the effect of learning on transaction costs, let alone on the shape of organization, is obvious. Rather, my point is that one cannot have a complete theory of the boundaries of the firm without considering in detail the process of learning in firms and markets. The reigning transaction-cost theories of vertical integration provide illuminating snapshots of possible institutional responses to a momentary situation. But they do not place those responses in the context of the passage of time. They are short-run theories that, unlike Marshallian price theory, have no long-run correlative.

This paper is an attempt to sketch out a possible approach to connecting the long run and the short. In order to do this, I will reassert the wisdom of the classical (long-run) view of organization, appending to it a theory of organizational learning now gaining attention as the 'capabilities' view.

2. *Organization and capabilities*

Although one can find versions of the idea in Smith, Marshall, and elsewhere, the modern discussion of the capabilities of organizations probably begins with Edith Penrose (1959), who suggested viewing the firm as a 'pool

[8] A term due to Alchian and Woodward (1988).

of resources.' Among the writers who have used and developed this idea are G. B. Richardson (1972), Richard Nelson and Sidney Winter (1982), and David Teece (1980, 1982). To all these authors, the firm is a pool not of tangible but of intangible resources. Capabilities, in the end, are a matter of knowledge. Because of the nature of specialization and the limits to cognition, organizations as well as individuals are limited in what they know how to do effectively. Put the other way, organizations possess a pool of more-or-less embodied 'how to' knowledge useful for particular classes of activities.

One sort of embodied knowledge is that contained in the firm's physical capital—that is, in machines. By rendering tasks a matter of routine, the division of labor (in the manner of the pinshop) allows for the substitution of skilled machines for skilled labor. But the capabilities embodied in machines are for present purposes the least interesting sorts of capabilities a firm might possess. More important are the sorts of knowledge embodied in the human capital of the firm, especially in those who manage it. Although management is clearly a highly skilled activity, the human capabilities of the firm are nonetheless quite nearly as much a matter of routine as are the skills of machines. 'Routines,' write Nelson and Winter (1982, p. 124), 'are the skills of an organization.' Indeed, as Michael Polanyi (1958) has argued, much of what we think of as skilled human behavior—in sports, the arts, everyday life—is in fact a matter of routine, in the sense that such skill consists in following inarticulate or 'tacit' rules of behavior. Such tacit knowledge is fundamentally empirical: it is gained through imitation and repetition not through conscious analysis or explicit instruction. This certainly does not mean that humans are incapable of innovation; but it does mean that there are limits to what conscious attention can accomplish. It is only because much of life is a matter of tacit knowledge and unconscious rules that conscious attention can produce as much as it does.

In a metaphoric sense, at least, the capabilities or the organization are more than the sum (whatever that means) of the skills of the individuals in the organization. In addition to the 'skill' of the firm's physical capital, there is also the matter of organization. How the firm is organized—how the routines of the humans and machines are linked together—is also part of a firm's capabilities. Indeed, 'skills, organization, and "technology" are intimately intertwined in a functioning routine, and it is difficult to say exactly where one aspect ends and another begins' (Nelson and Winter, 1982, p. 104).

Richardson and Teece have used notions like these to develop a theory of diversification. Just as a technological stage of production may be an 'anti-bottleneck' with excess capacity, so may an organization have excess capacity

in its organizational capabilities. In both cases, the result is the taking on of additional work. But in the case of organizational capabilities, the new activity need not be linked technologically to what the firm had previously been doing; rather, the new activity need only require a similar set of capabilities. In Richardson's terminology, the activities needn't be *complementary*; rather, they must be *similar*.

The flip-side of a theory of diversification, of course, is a theory of non-diversification—a theory of specialization. Such a theory would explain, as Coase once put it, 'why General Motors was not a dominant factor in the coal industry, and why A&P did not manufacture airplanes' (Coase, 1971, p. 67). And the basic answer is that capabilities have their limits. There are diminishing returns to spreading one's capabilities over more activities. This is so not merely for the reasons emanating from traditional span-of-control arguments (e.g. Robinson, 1934), but also because each new activity the firm could consider diversifying into will be increasingly dissimilar to—will require capabilities slightly different from—those the firm started out with. Ultimately, a firm will be restricted to activities that are fundamentally similar along one or another dimension.[9]

What gives this observation its salience, however, is that what is similar need not be what is complementary. That is, the various activities in the chain of production may—or may not—each require skills that are quite distinct. The manufacture of silicon wafers, from which integrated circuits begin, requires capabilities quite different from the fabrication of the semiconductors; as a result, the wafers are supplied by chemical companies, like Wacker Chemie, whose other activities are similar. The manufacture of the optical steppers used in the photolithography of the semiconductors is also unlike the fabrication of chips; but it is quite like the making of other precision optical equipment, which is why Nikon and Canon are among the suppliers of these devices (Langlois *et al.*, 1988).

3. *Capabilities and governance costs*

The capabilities view of the firm is in many ways a modern reformulation of the theory of Smith and Marshall: it is a real-time account of production costs

[9] I am obviously putting aside the phenomenon of conglomerates. But even here, one could argue that the relevant capability in excess capacity is that of financial management. Such conglomerates function largely as internal capital markets, which, as some have argued, may have advantages over the decentralized stock markets during periods in which inflation injects noise into the price system. [For the characterization of the conglomerate as an internal capital market, see Williamson (1985, p. 288). For an argument about the role of inflation, see Boudreaux and Shughart (1989).] Conglomerates now seem out of fashion, however, and many business analysts are arguing that firms are returning—or ought to return—to their 'core competences' (Prahalad and Hamel, 1990). For an interesting discussion of diversification strategies from the point of view of organizational capabilities, see Robertson (1990).

in which knowledge and organization have as important a role as technology. Unlike the Smithian theory narrowly understood, however, the capabilities view of the firm does give us some insights into the boundaries of the firm.

One implication of the boundedness of capabilities is that no firm—even the most integrated—has the capabilities necessary for all activities in the chain of production. The result is that firms must link up with other firms. This often takes place through the simplest of market contracts. One can, for example, buy off-the-shelf parts at spot prices and assemble a finished product out of them. But often—and especially when innovation is involved—the links among firms are of a more complex sort, involving everything from informal swaps of information (von Hippel, 1989) to joint ventures and other formal collaborative arrangements (Mowery, 1989). All firms must rely on the capabilities owned by others, especially to the extent those capabilities are dissimilar to those the firm possesses. A firm could—and many do—acquire dissimilar capabilities complementary to the ones they already own. But there is no particular reason to do so unless there are specific transaction costs impeding contractual arrangements. And there are generally costs to owning dissimilar assets, especially when the acquiring firm cannot use or sell their full capacity.

The existence in the market of complementary (but possibly dissimilar) capabilities is, of course, one kind of external economy Marshall thought important. And the level of relevant external capabilities in an economy will be important to the level of vertical integration we observe in that economy. In developing countries, or in developed economies when innovation renders the market's existing capabilities obsolete, a firm may have to integrate into many dissimilar activities in order to generate all the complementary activities it needs (Silver 1984). Consider the case of the American automobile industry (Langlois and Robertson, 1989). In the early days of the industry, automobile makers were all assemblers, that is, they contracted for almost all the parts that went into the cars, reserving only the assembly stage for themselves. They could do this because the American economy—and the Detroit region in particular—possessed a high level of general purpose machining and metal-working capabilities available in the market. The innovation of the moving assembly line at Ford, however, rendered these capabilities obsolete, in that Ford could mass-produce parts much less expensively than he could buy them on the market.[10] Because Ford could not quickly and cheaply convey to suppliers the (partly tacit) nature of the innovation—which was in any case a slowly unfolding process—he was

[10] Contrary to popular notions, the moving assembly line was significant not primarily as a way to assemble cars but as a way to manufacture parts for cars.

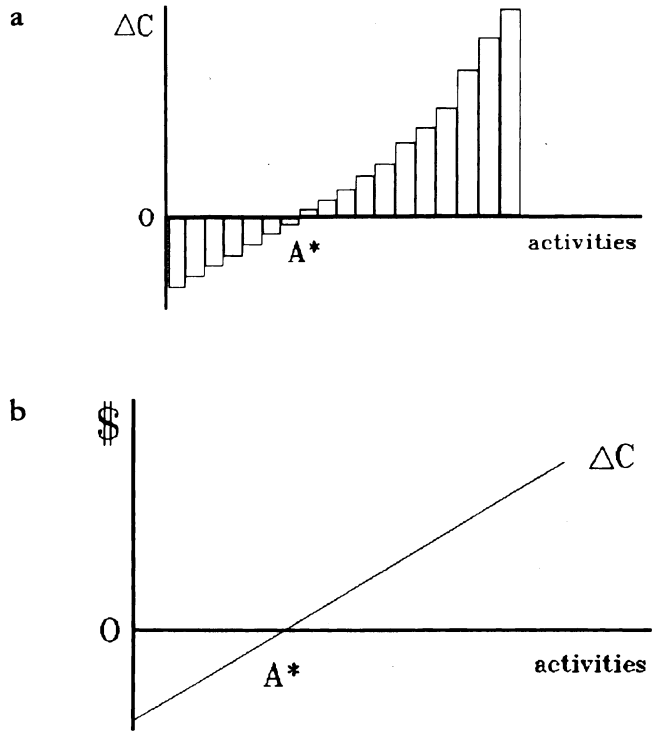

FIGURE 1. (a) Discrete representation; (b) continuous representation.

forced to integrate vertically into parts manufacture. Imitators like General Motors, however, could take advantage of the eventual spread of Ford ideas to the market, and thus needed less integration.

In short, then, the capabilities view of the firm suggests that the boundaries of the firm are determined (at least in part) by the relative strength of internal and external capabilities, that is, capabilities internal to the firm and those available through contract with other firms. Consider Figure 1.[11] On the X-axis we can array activities or stages of production in order of increasing cost of internal production. Specifically, ΔC graphs the normalized per-unit cost premium the firm must pay for the output of a particular activity if it integrates into that activity, measured relative to the per-unit cost it would incur by obtaining the output on contract from a distinct firm. Whenever this premium is negative, there is a cost advantage to internal organization. And the firm will acquire increasingly dissimilar activities

[11] This figure is inspired by, but modified from, Silver (1984, p. 44).

until the premium is zero, in this case at A*. Activities in the range OA* are within the boundaries of the firm; the rest are left to the market.[12]

The cost premium, and therefore the location of A*, will depend on a number of factors. As transaction-cost economics suggests, it will depend on the bureaucratic costs of internal organization and the transaction costs of market relations. But in this story, the location of the ΔC curve also depends on the internal capabilities of the firm and the external capabilities available in the market. That is to say, the price premium includes both governance-cost and production-cost differences.

If we hold capabilities constant, then we get the familiar account: whatever lowers bureaucratic costs on the margin will increase the extent of integration; whatever lowers transaction costs will reduce the extent of integration. If capabilities were unbounded, then governance costs alone would determine the boundaries of the firm. In such a case, the activities would be ordered according to the normalized per-unit governance-cost premium for internal over market procurement. By contrast, if governance costs were zero, capabilities alone would determine the boundaries of the firm. In this case, the activities would be ordered according to decreasing similarity, measured from the activity in which the firm has the greatest cost advantage over the market. In both polar cases, the firm consists of all the activities in the range OA*; but those are not necessarily the same set of activities in each case (see Figure 2).

In the long run, I have argued, transaction costs might be expected to approach zero. One might also argue this for governance costs generally. In the long run, activities have become increasingly routine. This reduces the cost of contracting, not in the sense that contracts have become cheaper to write but in the sense that contracts are increasingly unnecessary: everything is done tomorrow the way it was done today. In this sense, then, the long run also arguably reduces the cost of internal management by reducing decision-making costs.[13] Thus, one might argue that, in the long run as I have defined it, the boundaries of the firm are determined entirely by the capabilities of the firm relative to the capabilities of the market.

If, however, we follow Marshall in seeing the long run as the asymptotic end-state of a process of learning, then we also have to consider the ways in

[12] Such a diagram captures what Coase meant when he wrote that 'a firm will tend to expand until the costs of organizing an extra transaction within the firm become equal to the costs of carrying out the same transactions by means of an exchange in the open market or the costs of organizing in another firm' (Coase, 1937, p. 395).

[13] This is not to say, of course, that the long run favors internal organization. Since, as we will see, the benefits of internal management lie largely in the superior flexibility (of a specific kind) such management offers, we might well expect the *benefits* of internal management to decline faster than the costs in the long run, since flexibility becomes less important in a world of routine.

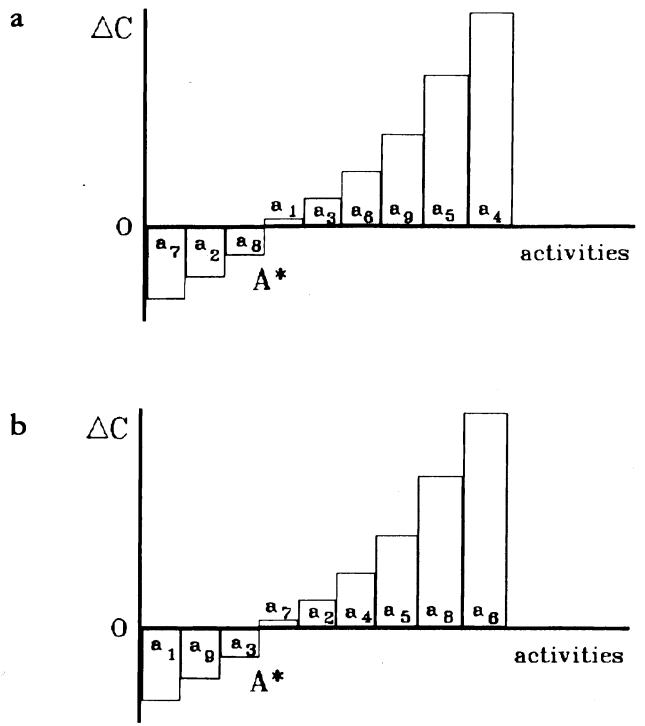

FIGURE 2. (a) Unbounded capabilities; positive governance costs. (b) Bounded capabilities; no governance costs.

which capabilities change over time. And here, it seems to me, there are also two opposing effects. On the one hand, the firm is likely to become more 'capable' over time. As more and more of the firm's activities take on the nature of routines, and as the firm's routines become more finely tuned, both the firm's total managerial capacity and its free managerial capacity will increase. Other things equal, this will shift ΔC down and increase the extent of OA*. On the other hand, however, the market will also become more 'capable' as time passes. Other firms will also be increasing their capabilities. And techniques pioneered by one firm may diffuse to and be imitated by other firms. All other things equal, this will have the effect of shifting ΔC up and lowering the extent of OA*.

The classical presumption was that this latter effect predominates: in the longest of runs capability diffuses completely into the market, leading to full specialization and vertical disintegration (see Figure 3). In general, the relative strengths of these effects will depend on the relative learning abilities of the firm and the market. The firm's learning ability will depend on its

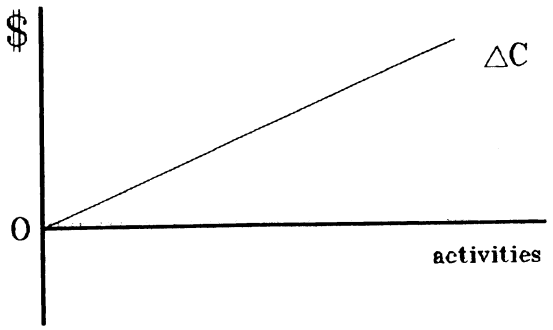

FIGURE 3. The classical presupposition.

internal organization. And the learning ability of the market will depend on technical and institutional factors, as well as on the learning abilities of the firms it comprises, considered both individually and as a system. The remainder of this paper is devoted to considering these two learning systems in slightly more detail. More specifically, it will set out some preliminary generalizations about how the level of capabilities in the firm and the market—and the nature of *change* in those capabilities—effects the boundaries of the firm.

Before turning to that task, however, let me raise a conceptual issue. I have so far tried to portray capabilities as in the nature of production costs, something distinct from transaction costs. But the line is actually far more blurry. Assume no transaction costs of the measurement or hold-up kind. And suppose that a firm chooses to undertake a particular activity internally rather than relying on the market. This must mean that the firm has a cost advantage over the market. But is that advantage in the nature of a production cost or a transaction cost? In a recent critique of the transaction-cost approach to the multinational, Paul Hallwood (1991) argues that much foreign direct investment can be explained simply by production-cost advantages of the foreign firm over indigenous firms. For example, a foreign soft-drink company may choose to set up its own bottling plants in a particular country instead of licensing indigenous bottlers simply because its bottling capabilities are superior to—and thus its costs lower than—those of local plants. Viewed from another perspective, however, might we not say that the foreign licensor integrates into local production because the transaction costs of using the local market capabilities were prohibitive? This is not merely tautological if we specify that the transaction costs involved are those of somehow transmitting to local market participants the knowledge that

would provide them with the necessary capabilities. As Hallwood suggests, this is not necessarily a helpful way to look at the issue, especially when, as we have argued, capabilities involve tacit knowledge that can be gained only by a long process of apprenticeship.

At the same time, of course, 'technology transfer' is in fact an option that presents itself to firms. Instead of producing oneself, one can teach others how to produce and persuade them to do so. In principle, Henry Ford *could* have taught outside parts suppliers the techniques of the moving assembly line and persuaded them to use them for high-volume production. The costs of doing so would have been prohibitive; but they would have been costs, in a straightforward foregone-opportunity sense. Similarly, a firm that opts to procure an input on the market *might* have chosen to acquire the necessary capabilities to produce internally. The cost of acquiring those capabilities is the diminution in profit from not having taken advantage of the market's capabilities. We might not want to call these transaction costs, but they are certainly information or knowledge costs. And it is these costs that permit the notion of 'capabilities,' unlike the classical or neoclassical notion of production costs, to help explain the boundaries of the firm. Consider the following case. A firm's capabilities grow, allowing it to apply those capabilities to a new activity similar to the ones in which it is already active. If there are truly no transaction costs, and the issue were solely the firm's advantage over other firms in production costs, then there would be no reason for the firm in fact to own the new stage of production. It could costlessly license its knowledge to others and earn a specific contractual return. To say that the firm must extend its boundaries to encompass the new activity implies that it is costly to license its capabilities, which means that there must be transition costs of *some* kind. [14]

For want of a better term, I propose to call these *dynamic* transaction costs. They are a kind of cost that has been largely neglected in the explanation of the boundaries of the firm. As I will explain more fully below, I will mean by dynamic transaction costs the costs of persuading, negotiating, coordinating and teaching outside suppliers. Another—if perhaps fast and loose—way to look at these transaction costs is as the costs of not having the capabilities you need when you need them. Indeed, if one follows Demsetz (1988) in using the term transaction cost to refer only to costs of using the market and never to costs of internal governance, then one ought to call these dynamic *governance* costs, since, as we'll see, it can also be a cost of internal organization not to have the capabilities one needs at the right time. When the market cannot provide the right capabilities at the right time, vertical integration may

[14] For a discussion of the cost of licensing that is somewhat in the spirit of the present essay, see Caves *et al.* (1983)

result; and when the firm lacks the right capabilities at the right time, vertical disintegration may occur. Let us examine these two possibilities in order.

4. *Capabilities, learning, and vertical integration*

Internal capabilities

As we have seen, the distinction between capabilities and transaction costs suggest two (non-exclusive) possibilities. On the one hand, a firm may need to internalize a stage of production because the complementary capabilities that stage represents do not exist or are more expensive in the market. This would be a pure capabilities explanation for internalization. Perhaps the case of Henry Ford and the moving assembly line might fit this possibility. His process innovation gave him a cost advantage over outside suppliers, motivating a high degree of vertical integration. On the other hand, a firm may wish to internalize a stage of production even when the market possesses the requisite capabilities to at least the same degree as does the firm itself. If the firm does internalize, it must be because there are other costs to using the markets. An explanation along these lines would be a pure transaction-cost explanation.

What might the sources of these transactions costs be? Asset specificity is a possibility with which the literature is enthralled. In a sense, this fits with the theme I have suggested: when there is a threat of hold-up, one might be afraid of not having the right capabilities available at the right time. For example, Alfred Chandler (1977) sees the backward integration of large manufacturing firms as largely if not entirely a defensive stratagem to avoid supply disruptions and ensure high-volume throughput. As I have already hinted, however, the problem with the hold-up view is that it is neither sufficient nor necessary as an explanation for integration. It is not sufficient because, in the absence of uncertainty and a divergence of expectations about the future, long-term contracts, reputation effects, and other devices can remove the costs of arm's-length arrangements. It is not necessary because, in the presence of uncertainty and a divergence of expectations about the future, arm's-length arrangements can be costly even without highly specific assets.

Consider the case of the diesel locomotive (General Motors Corporation, 1975). In Charles F. Kettering's research laboratories, General Motors developed in the 1920s a generic capability to produce powerful but lightweight diesel engines. Their earliest use was in submarines. But GM

chairman Alfred P. Sloan saw the possibility of applying the technology to a diesel-electric railroad locomotive. The existing steam-locomotive firms possessed many necessary capabilities GM lacked, as did General Electric and Westinghouse in all-electric technology. The three sets of capabilities might have combined by some kind of contract or joint venture. But the steam manufacturers—Alco, Baldwin, and Lima—failed to co-operate. This was not, however, because they feared hold-up in the face of highly specific assets. Rather, it was because they actively denied the desirability of the diesel and fought its introduction at every step. General Motors was forced to create its own capabilities in locomotive manufacture, although it did initially buy electrical systems from GE and Westinghouse. As Morris Silver (1984) has argued, the costs that impair market arrangements in such situations are in the nature of persuasion costs—the costs of getting the participants on the same wavelength. The necessity of investing in expensive transaction-specific assets may make such persuasion more difficult, but a divergence of entrepreneurial expectations is enough to do the trick.

Perhaps a more general way to look at these costs of persuasion is as costs of co-ordinating separate stages of production. David Teece encapsulates the argument nicely.

> If there is a high degree of interdependence among successive stages of production, and if occasions for adaptation are unpredictable yet common, coordinated responses may be difficult to secure if the separated stages are operated independently. Interdependence by itself does not cause difficulty if the pattern of interdependence is stable and fixed. Difficulties arise only if program execution rests on contingencies that cannot be predicted perfectly in advance. In this case, coordinated activity is required to secure agreement about the estimates that will be used as a basis for action. Vertical integration facilitates such coordination.
>
> This argument also reduces, at least in some respects, to a contractual-incompleteness argument. Were it feasible to stipulate exhaustively the appropriate conditional responses, coordination could proceed by long-term contract. However, long-term contracts are unsatisfactory when most of the relevant contingencies cannot be delineated. Given these limitations, short-term contracts are likely to be considered instead. . . . Even if short-term contracts are defective neither on account of investment disincentives nor first-mover advantages, the costs of negotiations and the time required to bring the system into adjustment by exclusive reliance on market signals are apt to be greater than the costs of administrative processes under vertical integration. (Teece, 1976, p. 13.)

Another way to say this is that unpredictable change makes it costly to

specify contractual provisions, implying the need for expanded residual rights of control.[15]

Teece mentions this possibility as one of a string of possible explanations for vertical integration. My contention is that this is in fact the general explanation, and that all other transaction-cost explanations are either derivative of this argument or apply only on an *ad hoc* basis to special situations. Ultimately, the costs that lead to vertical integration are the (dynamic) transaction costs of persuading, negotiating with, coordinating among, and teaching outside suppliers in the face of economic change or innovation.

When would such costs be likely? That is to say, when would we expect vertical integration? As Teece suggests, the costs of coordinating among stages would be greatest when there is a high degree of interdependence among the relevant stages of production. But more than mere interdependence is necessary: the interdependence must be of a sort such that a change in one stage of production requires a corresponding change in one or more distinct stages.

For simplicity, picture the chain of production as literally that: a linear progression from one stage to the next. We can say that an innovation is *autonomous* if it affects only one stage in the chain. By contrast, an innovation is *systemic* if it requires the coordination of change across more than one stage (Teece, 1986).

It was autonomous innovation that Adam Smith had in mind when he argued that the division of labor enhanced innovation: each operative, by seeking ways to make his or her lot easier, would discover improved methods of performing the particular operation (Smith, 1976, I.i.8, p. 20).[16] The

[15] For an argument that this was ultimately Frank Knight's theory of the firm, see Langlois and Cosgel (1990). One could also make an argument that this was in the end Coase's theory as well. Consider the following passage.

> It may be desired to make a long-term contract for the supply of some article or service. . . . Now, owing to the difficulty of forecasting, the longer the period of the contract is for the supply of the commodity or service, the less possible, and indeed, the less desirable it is for the person purchasing to specify what the other contracting party is expected to do. It may well be a matter of indifference to the person supplying the service or commodity which of several courses of action is taken, but not to the purchaser of that commodity or service. But the purchaser will not know which of these several courses he will want the supplier to take. Therefore, the service which is being provided is expressed in general terms, the exact details being left until a later date. . . . The details of what the supplier is expected to do is not stated in the contract but is decided later by the purchaser. When the direction of resources (within the limits of the contract) becomes dependent on the buyer in this way, that relationship which I term a 'firm' may be obtained. (Coase, 1937, pp. 391–392.)

[16] In fact, however, Smith also saw the division of labor as leading to systemic innovations: 'All the improvements in machinery, however, have by no means been the inventions of those who had occasion to use the machines. Many improvements have been made by the ingenuity of the makers of the machines, when to make them became the business of a peculiar trade; and some by that of those who are called philosophers or men of speculation, whose trade it is, not to do any thing, but to observe every thing; and who, upon that account, are often capable of combining together the powers of the most distant and dissimilar objects.' (I.i.9, p. 21) More on this below.

improvements he had in mind were such that they improved the efficiency of a particular stage without any implication for the operation of other stages. Autonomous innovation of this sort may even further the division of labor to the extent that it involves the cutting up of a task into two or more separate operations. Instead of being *differentiating* in this way, however, an innovation may be *integrating*, in the sense that the new way of doing things—a new machine, say—performs in one step what had previously needed two or more steps (Robertson and Alston, 1992). More generally, a systemic innovation may require small modifications of the way work is performed at each of a number of stages, and would thus require coordination among those stages.

This possibility of interconnectedness has been the basis for an argument that vertical disintegration may retard innovation.[17] Innovation may mean replacing assets at more than one stage in the chain of production. If decision-making is decentralized, the costs of coordinating the innovation may be high, and the innovation may never take place. This is particularly significant if some of the existing asset-holders, or the suppliers of factors complementary to the existing assets, have the power to block innovation (through trade unionism, for example) to protect their rent streams. If innovation does occur, it may take place elsewhere in the economy (and perhaps elsewhere in the world) under the direction of a unified asset-holder and decision-maker who can ignore existing task boundaries.

The empirical significance of this argument, especially as applied to the case of Britain at the turn of the twentieth century, is a subject of intense dispute. As a theoretical matter, however, this argument would seem most applicable to particular kinds of innovations, namely those that integrate operations. And what kinds of innovation do this? Surely one class of important systemic innovations comprises major organizational shifts. Examples would include the factory mode of production (Leijonhufvud, 1986), the moving assembly line (Hounshell, 1984), refrigerated meat-packing (Silver, 1984, pp. 28–29), and containerized shipping (Teece, 1986). All of these examples are ultimately *process* innovations. And one might argue that, although process innovation may also proceed in an autonomous way,[18] there

[17] An early version of the argument is by Marvin Frankel (1955). (See also Gordon, 1956, and Frankel, 1956.) The idea has more recently found a champion in William Lazonick (1981; Elbaum and Lazonick 1986). Such notables as Sir Arthur Lewis (1957, pp. 583–584) and Charles Kindleberger (1969), pp. 146–147) have also pointed to vertical fragmentation as a cause of British industrial decline.

[18] There are in fact many examples of this in industry. An example at the forefront of modern technology is semiconductor fabrication. A large group of (mostly American) firms is pushing for a modular equipment architecture standard (MESA) that will allow a fabricator of semiconductors to mix and match process equipment from many different manufacturers (Winkler, 1990). Each unit will be able to 'bolt on' to the system so long as it obeys the standard. This will allow a greater degree of autonomous innovation in process technology. The main reasons for this development seem to be the highly

are typically advantages to systemic process innovation. For one thing, process innovation is often integrating, requiring the consolidation of several stages of production in a single (usually mechanized) stage. More generally, process innovation is frequently a matter of fine-tuning the production process in the face of steady and predictable growth in demand: learning how to shave time off operations, to eliminate steps, to substitute stamping technology for casting, etc. This class learning- or experience-curve effect arguably proceeds faster in an integrated environment in which systemic change is relatively inexpensive.

How would learning proceed once the integrated organization was established? As change slows, the boundaries between the stages will begin to stabilize, and the costs of coordinating among stages will decrease. One would thus expect a greater decentralization of operations in general,[19] including a spinning off of the activities most dissimilar to the firm's 'core competence.' The extent to which this happens will depend on the relative learning abilities of the market and the firm. If capabilities diffuse easily to the market, we would expect more spinning off and thus less integration with time. If capabilities do not diffuse easily, disintegration will be slowed. Moreover, the firm may be more or less able to learn over time. For example, it may have an R&D lab, or possess a structure and culture conducive to learning.

Cohen and Levinthal (1989, 1990a, b) argue that a firm's ability to learn is governed by its 'absorptive capacity.' This capacity is typically, though not exclusively, a byproduct of R&D. A firm engages in R&D not only to create new knowledge but also, and often more importantly, to increase its ability to perceive and utilize knowledge generated outside the company. That is to say, the ability to learn is itself a capability the firm possesses and in which it can invest. Indeed, since one's ability to assimilate new knowledge is arguably a function of the related knowledge one already possesses, absorptive capacity is cumulative: one's capacity to acquire new capabilities depends on one's existing level of related capabilities. This reinforces the observation that organizations may be good at certain kinds of learning by doing, so long as the economically relevant knowledge the organization needs to learn does not stray far from what it already knows.

---

decentralized nature of American capabilities in semiconductor fabrication equipment and the inadequate capabilities of even the largest producers to create all the elements necessary for the increasingly integrated process technology of modern semiconductors.

[19] This was in fact true in the case of the moving assembly line. Once the innovation had taken its basic form, Ford found it desirable, in the face of a growing extent of the market, increasingly to set up geographically dispersed plants specializing in the fabrication of particular parts (Langlois and Robertson, 1989, p. 368). In this way, he replicated something like the network of decentralized producers existing before the innovation—except that the new network used the moving assembly line and was all owned by Ford.

External capabilities

The possibility that a firm may need to integrate vertically because the market somehow can't deliver has been a theme in the literature at least since Coase. What has received considerably less attention is the way in which failures of internal capabilities can force a firm to disintegrate, that is, to turn to the market for the capabilities it needs. Clearly, the need to coordinate innovation across stages of production can present a decentralized network of firms with transaction costs. But there are also benefits to decentralization. And, in some cases and for some types of innovation, these benefits can greatly outweigh any transaction costs.

In general, vertical disintegration would prove superior to vertical integration when complementary capabilities either don't exist within the firm or are inferior to those available in the market. Marshall talked about external economies as an explanation of economic progress. By analogy, we can talk about *external capabilities* available to the firm through contract. A firm may choose to rely on external capabilities if the (dynamic) governance costs of generating those capabilities internally are high.

Consider the case of the personal computer (Langlois, 1991). In entering the PC market in the early 1980s, IBM understood both (1) that the market possessed a high level of capabilities and (2) that IBM's own capabilities were severely lacking. This latter was the case partly because the company had focused on larger computers and did not possess all the capabilities necessary for smaller machines. But it was also and more importantly because the company's hierarchical structure, internal sourcing procedures, and elaborate system of controls made it too inflexible to respond well to a rapidly changing market. As a result, IBM chose in effect to disintegrate vertically into the production of PCs. They spun off a small group of executives and engineers, exempted them from IBM internal sourcing and other procedures, and treated them as, in effect, a venture-capital investment. The original IBM PC was in fact almost completely assembled from parts available in the market, very few of which were produced in IBM plants. IBM's motives for *disintegration* were in this regard strikingly similar to Henry Ford's motives for *integration*: the need to access quickly capabilities that would not otherwise have been available in time.

With a slower pace of change—and/or the resources to subsidize a short-run cost disadvantage—firms may of course choose to invest in internal capabilities as part of a longer-term strategy. But internal capabilities are not always good substitutes for external capabilities. To see why this may be so, we need to turn back to one of the Marshallian themes with which we started: the market as an evolutionary system. Perhaps the central difference

between Darwinian and Linnaean biology is that Darwin highlighted not what was common to the organisms in a species but what was *different*: the natural variation among organisms is what made evolutionary selection possible. Marshall saw the economic system in the same way. 'The tendency to variation,' he believed, 'is a chief cause of progress' (Marshall, 1961, V.iv.3, p. 355).[20] Thus, a high degree of variation—a high rate of technological and organizational experiment—is crucial to economic progress (Nelson and Winter, 1977). And the ability of a vertically disintegrated industry to generate, transmit and assimilate new ideas is thus a potentially powerful external capability, external in the same sense that Marshall understood external economies: it is a property of the system as a whole and cannot be reduced to the internal capabilities of firms taken individually. A market form of organization is capable of learning and creating new capabilities, often in a self-reinforcing and synergistic way. Marshall describes just such a system when he talks about the benefits of localized industry.

> The mysteries of the trade become no mysteries; but are as it were in the air, and children learn many of them unconsciously. Good work is rightly appreciated, inventions and improvements in machinery, in processes and the general organization of the business have their merits promptly discussed: if one man starts a new idea, it is taken up by others and combined with suggestions of their own; and thus it becomes the source of further new ideas. And presently subsidiary trades grow up in the neighbourhood, supplying it with implements and materials, organizing its traffic, and in many ways conducing to the economy of its materials. (Marshall, 1961, IV.x.3, p. 271.)

In this sense, the ability of a large organization to coordinate the implementation of an innovation, which is clearly an advantage in some situations, may be a *disadvantage* in other ways. Coordination means getting everyone on the same wavelength. But the variation that drives an evolutionary learning system depends on people being on different wavelengths—it depends, in effect, on outbreeding. This is something much more difficult to achieve in a large organization than in a disintegrated system. Indeed, as Cohen and Levinthal (1990a, p. 132) point out, an organization experiencing rapid change ought in effect to emulate a market in its ability to expose to the environment a broad range of knowledge-gathering 'receptors.'

Vertical integration, I argued, might be most conducive to systemic, integrative innovations, especially those involving process improvements

[20] On this see also Loasby (1990).

when demand is high and predictable. By contrast, vertical integration may be less desirable—and may be undesirable—in the case of differentiating or autonomous innovations. Such innovations require less coordination, and vertical integration in such cases may serve only to cut off alternative approaches. Moreover, disintegration might be most beneficial in situations of high uncertainty: situations in which the product is changing rapidly, the characteristics of demand are still unknown, and production is either unproblematical or production costs play a minor role in competition. In such cases the coordinating benefits of vertical integration are far outweighed by the evolutionary benefits of disintegration.

In part, this is a matter of the so-called product life-cycle (Utterback, 1979). At the early stages of the life-cycle, uncertainty is high and the product is fluid. Here a diversity of approaches can blanket the product space and seek out desirable forms quickly. As the product develops a dominant paradigm, however, change becomes more incremental, and fine-tuning of production for low-cost, high-volume output comes to the fore—and vertical integration gains the advantage. Nevertheless, it may also be the case that disintegration retains its benefits for some time. In some cases product innovation may not slack off very fast. In other cases, low-cost, high-volume production may be already available as an external economy (in a way that it was not available to Henry Ford).[21]

How would learning proceed in a system of decentralized capabilities? As I've already suggested, progress would take place autonomously within the decentralized stages. There would be no need for integration unless a systemic innovation offering superior performance arrives on the scene. Indeed, as we have seen, fixed task boundaries and standardized connections between stages might make innovation difficult within the existing structure, requiring a kind of creative destruction[22] (Schumpeter, 1950). Within the individual (autonomous) stages, however, learning can proceed in systemic fashion. The production process for or internal structure of a modem or stereo amplifier are irrelevant so long as those items can connect in properly with other components. Thus, it may not be surprising to find that vertical integration may be more important in the production of the components of a larger system than in the system itself. To put it another way, one's appraisal of the benefits of vertical integration may depend on how one defines the 'product' under study.

[21] More interestingly, perhaps, the product may develop into a modular system (Langlois and Robertson, 1990). In such a system, the connections among the parts making up the product become standardized, rendering innovation much more autonomous than in cases in which the connections among parts are fluid.

[22] Tushman and Anderson (1986) talk of technological change that is competence enhancing or, as in this case, competence destroying.

Markets and hierarchies

As G. B. Richardson (1972) pointed out some time ago, the easy partition of alternatives into markets (relying on price signals) and firms (relying on authority relations or hierarchies) is not a good description of how the world works. In fact, what we see out there is a mixture of modes of ownership and contract. As Imai and Itami (1984) put it, there is typically an 'interpenetration' of organization and market, leading to organization-like markets and market-like organizations as well as the ideal types of pure market and pure hierarchy.

The system of product development typical in Japan illustrates the mixed nature of ownership and contract in the modern economy. It also amplifies many of the points about organizational and market learning touched on above. As described by Imai *et al.* (1985), the Japanese process of product development has these seven characteristics: (1) top management as a catalyst; (2) a self-organizing project team; (3) overlapping development phases; (4) multi-learning; (5) subtle control; (6) organizational transfer of learning; and (7) organizational network building. Some of these characteristics have to do with the way Japanese firms organize to learn. But one important characteristic of the system—number 7—speaks to the way in which Japanese firms rely on decentralized capabilities outside their boundaries.

Although the authors do not put it this way, several of these characteristics are motivated by the tradeoff inherent in innovation between diversity and synthesis. As Adam Smith understood, innovation often involves 'combining together the powers of the most distant and dissimilar objects' (1976, I.i.9, p. 21). This requires both having knowledge of the distant and dissimilar and being able to combine them together. A decentralized market is good at generating diversity, whereas a more unified institution—Smith's 'philosopher' or his organizational counterpart—would have the advantage in combining the diverse elements together. Japanese firms effectively compromise. They encourage diversity in the project team, which is often staffed by generalists who must learn what they need to know and therefore bring a fresh perspective. This is part of what Imai *et al.* mean by 'multi-learning.' Also, the overlapping of development phases helps prevent too-early specialization. At the same time, however, the development team is tight-knit, permitting effective communication; and the galvanizing and 'subtle control' functions of management help synthesize what is learned into a useful form.

Most interesting—and perhaps surprising in light of popular Western images of the monolithic Japanese enterprise—is, however, the remarkable extent to which the innovating firm relies on a large and interactive network

of outside contractors. This is true not merely at the manufacturing stage but even during product development. In the case studies Imai *et al.* examine, between 65 and 90 per cent of the parts at the mass-production stage came from outside, as did as much as 90 per cent at the product-development stage.[23] Some contractors even specialize in product-development work, aggregating the demands of many teams at many companies. In short, the case of the IBM PC arguably fits the Japanese model of product development more closely than does, say, the case of the Model T.

5. *Summary and conclusions*

I have attempted in this paper to place the theory of the boundaries of the firm within the context of the passage of time. More precisely, I have tried to resurrect and place in a modern frame some of the insights of the classical theory of organization. In the Marshallian long run (correctly understood), transaction costs approach zero, and the boundaries of the firm become irrelevant. Governance costs—the transaction costs of markets and the bureaucratic costs of organizations—are thus short-run phenomena.

The modern reinterpretation of the classical theory of organization centers around the 'capabilities' view of the firm. Rather than thinking in terms of production technology as a set of blueprints, consider the firm instead to possess a set of partly tacit knowledge, routines, and skills applicable to certain activities. Taken together with governance costs, these capabilities determine the boundaries of the firm in the short run. Those activities for which the firm has a cost advantage over the market—either because of superior capabilities or because of favorable governance costs or both—will be within the boundaries of the firm. The remaining complementary activities will be accessed through arm's-length arrangements.

One might think that, as governance costs diminish in the long run, the boundaries of the firm would be determined solely by capabilities. But capabilities also change over time as firms—and markets—learn. The classical presumption was that the firm's capabilities would diffuse completely to the market in the long run, leading to complete vertical disintegration. This reinforces the point that capabilities are more than a matter of production costs in the neoclassical sense and, more importantly, suggests that the notion of a firm's capabilities implies a kind of information or knowledge cost—the cost of transferring the firm's capability to the market (other firms) or vice versa. These costs are a neglected kind of governance cost, which I

[23] It has been observed elsewhere that Japanese engineers 'reach instinctively for the parts catalogue,' and use off-the-shelf components to a much greater extent than do their American counterparts. (Anonymous, 1991, p. 61).

call 'dynamic' governance costs. These are the costs of transferring capabilities: the costs of persuading, negotiating and coordinating with, and teaching others. These costs arise in the face of change, notably technological and organizational innovation. They are in effect the costs of not having the capabilities you need when you need them.

These costs provide an explanation for vertical integration that is arguably more general than those dominant in the literature of transaction costs. In the face of uncertainty and divergent views of the future, common ownership of multiple stages of production is a superior institutional arrangement for coordinating systemic change. This observation is by no means entirely inconsistent with the existing literature. For example, Williamson (1985) stresses the firm's superior capacity for adaptive, sequential decision-making in the face of both uncertainty and highly specific assets. I assert, however, that asset-specificity is neither necessary nor sufficient for these dynamic transaction costs to lead to integration.

What has received considerably less attention in the literature is the possibility that these dynamic governance costs may also afflict internal organization. It may sometimes be costly—in terms of persuasion, negotiation and teaching—to create internally capabilities readily available on the market. The firm may nonetheless choose to bear these costs. And if it is desirable to coordinate the relevant activities systemically (e.g. fine-tuning process technology), the firm's decision to create internal capabilities may put it on a learning path that eventually gives it a cost superiority over the market. But in cases in which systemic coordination is not the issue, the market may turn out to be the superior learning engine because of its ability to generate rapid trial-and-error learning.

Acknowledgements

The author would like to thank Fred Carstensen, Wes Cohen, Joshua Haimson, William Lazonick, Scott Masten, Lanse Minkler, Richard Nelson and Paul Robertson for helpful comments and discussions.

References

Alchian, A. and H. Demsetz (1972), 'Production, Information Costs, and Economic Organization,' *American Economic Review*, 62, 772–795.

Alchian, A. and S. Woodward (1988), 'The Firm Is Dead; Long Live the Firm: A Review of Oliver E. Williamson's *The Economic Institutions of Capitalism*', *Journal of Economic Literature*, 26, 65–79.

Ames, E. and N. Rosenberg (1865), 'The Progressive Division and Specialization of Industries,' *The Journal of Development Studies*, 1, 363–383.

Anonymous (1991), 'What Makes Yoshio Invent,' *The Economist*, January 12.

Axelrod, R. (1984), *The Evolution of Cooperation*. New York: Basic Books.

Barzel, Y. (1982), 'Measurement Costs and the Organization of Markets,' *Journal of Law and Economics*, 25, 27–48.

Barzel, Y. (1987), 'The Entrepreneur's Reward for Self-Policing,' *Economic Inquiry*, 25, 103–116.

Blaug, M. (1987), *Economic Theory in Retrospect*. Cambridge: Cambridge University Press, 4th edition.

Boudreaux, D. J. and W. F. Shugart, II (1989), 'The Effects of Monetary Instability on the Extent of Vertical Integration,' *Atlantic Economic Journal*, 17, 1–10.

Caves, R., H. Crookell and P. J. Killing (1983), 'The Imperfect Market for Technology Licenses,' *Oxford Bulletin of Economic Statistics*, 45, 249–267.

Chandler, A. D. (1977), *The Visible Hand: The Managerial Revolution in American Business*. Cambridge: The Belknap Press.

Cheung, S. N. S. (1983), 'The Contractual Nature of the Firm,' *Journal of Law and Economics* 26, 1–22.

Coase, R. H. (1937), 'The Nature of the Firm,' *Economica (N.S.)* 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*. New York: NBER.

Cohen, W. M. and D. A. Levinthal (1989), 'Innovation and Learning: the Two faces of R&D,' *The Economic Journal*, 99, 569–596.

Cohen, W. M. and D. A. Levinthal (1990a), 'Absorptive Capacity: A New Perspective on Learning and Innovation,' *Administrative Science Quarterly*, 35, 128–152.

Cohen, W. M. and D. A. Levinthal (1990b), 'Fortune Favors the Prepared Firm,' Mimeo, Carnegie Mellon University.

Currie, M. and I. Steedman (1990), *Wrestling with Time: Problems in Economic Theory*. Ann Arbor: University of Michigan Press.

Dahlman, C. (1979), 'The Problem of Externality,' *Journal of Law and Economics*, 22, 141–162.

David, P. (1985), 'Clio and the Economics of QWERTY,' *American Economic Review*, 75, 332–337.

Demsetz, H. (1988), 'The Theory of the Firm Revisited,' *Journal of Law, Economics, and Organization*, 4, 141–161.

Elbaum, B. and W. Lazonick (1986), 'An Institutional Perspective on British Decline,' in B. Elbaum and W. Lazonick, (eds), *The Decline of the British Economy*. Oxford: Clarendon Press, pp. 1–17.

Frankel, M. (1955), 'Obsolescence and Technological Change in a Maturing Economy,' *American Economic Review*, 45, 296–319.

Frankel, M. (1956), 'Reply,' *American Economic Review*, 46, 652–656.

General Motors Corporation (1975), 'The Locomotive Industry and General Motors,' in Y. Brozen, (ed.), *The Competitive Economy*. Morristown, NJ: General Learning Press.

Gordon, D. F. (1956), 'Obsolescence and Technological Change: Comment,' *American Economic Review*, 46, 646–652.

Grossman, S. J. and O. D. Hart (1986), 'The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration,' *Journal of Political Economy*, 94, 691–719.

Hallwood, C. P. (1991), 'Production Costs, Measurement Costs, and the Theory of the Multinational Corporation,' Working Paper 91–1001, Department of Economics, University of Connecticut.

Hart, O. D. (1988), 'Incomplete Contracts and the Theory of the Firm,' *Journal of Law, Economics, and Organization*, 4, 119–140.

Hart, O. D. (1989), 'An Economist's Perspective on the Theory of the Firm,' *Columbia Law Review*, 89, 1757–1774.

Hayek, F. A. (1945), 'The Use of Knowledge in Society,' *American Economic Review*, 35, 519–530.

Hayek, F. A. (1967), *Studies in Philosophy, Politics, and Economics*. Chicago: University of Chicago Press.

Hounshell, D. A. (1984), *From the American System to Mass Production, 1800–1932*. Baltimore: Johns Hopkins University Press.

Imai, K. and H. Itami, (1984), 'Interpenetration of Organization and Market: Japan's Firm and Market in Comparison with the U.S.,' *International Journal of Industrial Organization*, 2, 285–310.

Imai, K., I. Nonaka and H. Takeuchi (1985), 'Managing the New Product Development Process: How Japanese Companies Learn and Unlearn,' in K. B. Clark, R. H. Hayes and C. Lorenz (eds), *The Uneasy Alliance: Managing the Productivity-Technology Dilemma*. Boston: Harvard Business School Press.

Jensen, M. C. and W. H. Meckling (1976), 'Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure,' *Journal of Financial Economics*, 3, 305–360.

Kindleberger, C. (1969), *Economic Growth in France and Britain, 1851–1950*. New York: Simon and Schuster.

Klein, B. (1988), 'Vertical Integration as Organizational Ownership: The Fisher Body-General Motors Relationship Revisited,' *Journal of Law, Economics, and Organization*, 4, 199–213.

Klein, B., R. G. Crawford and A. Alchian (1978), 'Vertical Integration, Appropriable Rents, and the Competitive Contracting Process,' *Journal of Law and Economics*, 21, 297–326.

Langlois, R. N. (1984), 'Internal Organization in a Dynamic Context: Some Theoretical Considerations,' in M. Jussawalla and H. Ebenfield (eds), *Communication and Information Economics: New Perspectives*. Amsterdam: North-Holland.

Langlois, R. N. (1986), 'Rationality, Institutions, and Explanation,' in R. N. Langlois (ed.), *Economics as a Process: Essays in the New Institutional Economics*. New York: Cambridge University Press.

Langlois, R. N. (1988), 'Economic Change and the Boundaries of the Firm,' *Journal of Institutional and Theoretical Economics*, 144, 635–657, reprinted in Bo Carlsson (ed.), *Industrial Dynamics: Technological, Organizational, and Structural Changes in Industries and Firms*. Dordrecht: Kluwer Academic Publishers, 1989, pp. 85–107.

Langlois, R. N. (1991), 'External Economies and Economic Progress: The Case of the Microcomputer Industry,' Working Paper 91–1502, Department of Economics, the University of Connecticut.

Langlois, R. N. and M. M. Cosgel (1990), 'Knight on Risk, Uncertainty, and the Firm: A New Interpretation,' manuscript, University of Connecticut.

Langlois, R. N., T. A. Pugel, C. S. Haklisch, R. R. Nelson and W. G. Egelhoff (1988), *Microelectronics: An Industry in Transition*, London: Unwin Hyman.

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

Langlois, R. N. and P. L. Robertson (1990), 'Networks and Innovation in a Modular System: Lessons from the Microcomputer and Stereo Component Industries,' Discussion Paper, Department of Economics and Management, University College, University of New South Wales (forthcoming in *Research Policy*).

Lazonick, W. (1981), 'Competition, Specialization, and Industrial Decline,' *Journal of Economic History*, 41, 31–38.

Leijonhufvud, A. (1986), 'Capitalism and the Factory System', in R. N. Langlois (ed.), *Economics as a Process: Essays in the New Institutional Economics*. New York: Cambridge University Press, pp. 203–223.

Lewis, W. (1957), 'International Competition in Manufacturers,' *American Economic Review*, 47, 583–584 (supplement).

Loasby, B. J. (1989), 'Knowledge and Organization: Marshall's Theory of Economic Progress and Coordination,' in B. J. Loasby, *The Mind and Method of the Economist*. Edward Elgar.

Loasby, B. J. (1990), 'Firms, Markets, and the Principle of Continuity,' in J. K. Whitaker (ed.), *Centenary Essays on Alfred Marshall*. Cambridge: Cambridge University Press.

Marshall, A. (1961), *Principles of Economics*. London: Macmillan, 9th (variorum) edition, volume I.

Mowery, D. (1989), 'Collaborative Ventures Between U.S. and Foreign Manufacturing Firms,' *Research Policy*, 18, 19–32.

Nelson, R. R. and S. G. Winter (1977), 'In Search of More Useful Theory of Innovation,' *Research Policy*, 5, 36–76.

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

O'Driscoll, G. P., Jr and M. J. Rizzo (1985), *The Economics of Time and Ignorance*. Oxford: Basil Blackwell.

Penrose, E. T. (1959), *The Theory of the Growth of the Firm*. Oxford: Basil Blackwell.

Polanyi, M. (1958), *Personal Knowledge*. Chicago: University of Chicago Press.

Prahalad, C. K. and G. Hamel (1990), 'The Core Competence of the Corporation,' *Harvard Business Review*, pp. 79–91 (May–June).

Richardson, G. B. (1972), 'The Organisation of Industry,' *Economic Journal*, **82**, 883–896.

Robertson, P. L. (1990), 'Economies of Scope, Organizational Culture, and the Choice of Diversification Strategies,' Economics and Management Working Paper #2, University College, University of New South Wales.

Robertson, P. L. and L. J. Alston (1992), 'Technological Choice and the Organization of Work in Capitalist Firms,' *Economic History Review*, 45, in press.

Robinson, E. A. G. (1934), 'The Problem of Management and the Size of Firms,' *Economic Journal*, 44, 242–257.

Rosenberg, N. (1963), 'Technological Change in the Machine Tool Industry, 1840–1910,' *Journal of Economic History*, **23**, 414–443.

Schumpeter, J. A. (1934), *The Theory of Economic Development*. Cambridge: Harvard University Press (New York: Oxford University Press, 1961).

Schumpeter, J. A. (1950), *Capitalism, Socialism, and Democracy*. New York: Harper and Brothers, 2nd edition.

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

Smith, A. (1976), *An Inquiry into the Nature and Causes of the Wealth of Nations*. Glasgow edition. Oxford: Clarendon Press. (First published in 1776.)

Stigler, G. (1951), 'The Division of Labour Is Limited by the Extent of the Market,' *Journal of Political Economy*, **59**, 185–193.

Sugden, R. (1986), *The Economics of Rights, Cooperation, and Welfare*. Oxford: Basil Blackwell.

Teece, D. J. (1976), *Vertical Integration and Vertical Divestiture in the US Oil Industry: Analysis and Policy Implications*. Stanford: Stanford University Institute for Energy Studies.

Teece, D. J. (1980), 'Economies of Scope and the Scope of the Enterprise,' *Journal of Economic Behavior and Organization*, 1, 223–247.

Teece, D. J. (1982), 'Towards an Economic Theory of the Multiproduct Firm,' *Journal of Economic Behavior and Organization*, 3, 39–63.

Teece, D. J. (1986), 'Profiting from Technological Innovation: Implications for Integration, Collaboration, Licensing, and Public Policy,' *Research Policy*, 15, 285–305.

Tushman, M. and P. Anderson, (1986), 'Technological Discontinuities and Organizational Environments,' *Administrative Science Quarterly*, 31, 439–465.

Utterback, J. M. (1979), 'The Dynamics of Product and Process Innovation,' in C. T. Hill and J. M. Utterback (eds), *Technological Innovation for a Dynamic Economy*. New York: Pergamon Press, pp. 40–65.

von Hippel, E. (1989), 'Cooperation Between Rivals: Informal Know-how Trading,' in B. Carlsson (ed.), *Industrial Dynamics: Technological, Organizational, and Structural Changes in Industries and Firms*. Dordrecht: Kluwer Academic Publishers.

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

Winkler, E. (1990), 'MESA, Applied Fued Stirring Over Pact,' *Electronic News*, April 16, p. 26.

Young, A. A. (1928), 'Increasing Returns and Economic Progress,' *Economic Journal*, **28**, 523–542.