Explaining Vertical Integration: Lessons from the American Automobile Industry

Richard N. Langlois, Paul L. Robertson

Journal of Economic History, Volume 49, Issue 2, The Tasks of Economic History (Jun., 1989), 361-375.

Journal of Economic History
01989 Economic History Association

Explaining Vertical Integration: Lessons from the American Automobile Industry

RICHARD N. LANGLOIS AND PAUL L. ROBERTSON

The early history of the American automobile industry provides fertile hunting grounds for theorists seeking corroboration of various, conflicting theories of vertical integration. An examination of the whole history suggests that no single theory always fits the facts perfectly. A complete explanation must combine specific theories in a way that is attentive to such factors as industry life-cycle, demand, economies of scale, and appropriability. If there is any "general" theory, it lies in the set of "dynamic" transaction-cost approaches rather than in the asset-specificity approach now dominant.

VERTICAL INTEGRATION: THEORY

One starting point for a theory of vertical integration goes back at least to Adam Smith: the division of labor. Writers in this tradition include Allyn Young and George Stigler, who assume that the firms in an industry are initially vertically integrated and that increasing output leads to differentiation as various stages of the production process are spun off into specialized concerns.[1] Whereas small firms in industries with limited output might need to undertake the production of intermediate goods because outside suppliers would not find it profitable to manufacture on such a limited scale, an expansion of the output of final products could permit specialized firms to take over the production of intermediate goods. This differentiation is especially important where there are different levels of economies of scale for the various production stages- where, for example, the most efficient level of output of an intermediate product is greater than that needed by any single manufacturer of the final good. As an example of this kind of differentiation, Young cites the early printing industry, which over time evolved into not only the modern printing industry, but also into firms turning out wood pulp and paper, inks, type, and other inputs into printing.

While there are certainly industries in history that have fit this pattern, we can just as easily point to industries in which the pattern has been quite different. Indeed, as we will suggest shortly, the automobile

The Journal of Economic History, Vol. XLIX, No. 2 (June 1989). ISSN 0022-0507.

Richard Langlois is Associate Professor of Economics, University of Connecticut, Storrs, CT 06269-1063. Paul Robertson is Head of the Department of Economics and Management, University College, University of New South Wales, A.D.F.A., Campbell, ACT 2600, Australia.

The authors would like to thank Lee Alston, Fred Carstensen, Daniel Raff, John D. Robertson, and John R. Robertson for helpful comments.

[1] Allyn A. Young, "Increasing Returns and Economic Progress," *Economic Journal, 38* (Dec. 1928), pp. 523-42; and George J. Stigler, "The Division of Labor Is Limited by the Extent of the Market," *Journal of Politicd Economy,* 59 (June 1951), pp. 185-93.

industry is a potential counterexample: firms were quite differentiated at an early stage in the industry's life but became much more integrated as output expanded.

What makes the approach from the division of labor incomplete is that it considers only one component of cost: the raw technological costs of manufacturing. Since Ronald Coase's famous article in the 1930s, economists have been alert to the importance of transaction costs in explaining the structure of economic organization.[2] This often fuzzy category is intended to comprise such less tangible costs as those involved in search, negotiation, and monitoring. Oliver Williamson is the leader of a group of writers who argue strongly that the pattern of vertical integration in an industry reflects a minimizing of the sum of production and transaction costs, not just production costs alone (as in the Young-Stigler model).[3]

In order to use the concept of transaction cost meaningfully in an explanation of economic organization, one has to specify the nature and sources of the costs one has in mind. One tradition in the literature has pointed to the costs of measuring the output of a stage of production. In the well-known formulation of Armen Alchian and Harold Demsetz, for example, technological indivisibilities in team production make it costly to monitor the marginal products of individual team members, suggesting some form of internal organization as a less costly alternative to market contracts between the members.[4]

Another approach has emphasized the costly possibility that parties to an arm's-length contract-a large buyer; a supplier, monopolistic or otherwise; organized or hard-to-replace laborers-might threaten to withhold the services of the assets they control in order to appropriate a larger share of the quasirents of cooperative production. The potential for "hold-ups" of this sort makes for higher expected costs ex ante than would be the case if the assets involved were jointly owned. The threat of hold-up is least when the assets involved are generalized ones, thickly supplied and therefore easy to replace. Conversely, the threat of hold-up, and the cost of arm's-length contracting, is greatest when assets are highly specific, thinly supplied and hard to replace.

This asset-specificity version of transaction-cost theory has been arguably the dominant approach in the study of vertical integration. Its proponents have claimed the support of economic history in general,

[2] Ronald H. Coase, "The Nature of the Firm," *Economica* (new series), 4 (Nov. 1937), pp. 386-*405*.

[3] Oliver E. Williamson, *The Economic Institutions of Capitalism* (New York, 1985), esp. p. 103.

[4] Armen Alchian and Harold Demsetz, "Production, Information Costs, and Economic Organization," *American Economic Review, 62* (Dec. 1972). pp. 777-95. For a more general formulation of this measurement-cost approach, see Yoram Barzel, "Measurement Costs and the Organization of Markets," *Journal of Law and Economics, 25* (Apr. 1982), pp. 27-48; and Steven N. S. Cheung, "The Contractual Nature of the Firm," *Journal of Law and Economics*, 26 (Apr. 1983), pp. 1-21.

with the case of the auto industry figuring prominently in the evidence. Nonetheless, there remains reason to think that this approach may not offer a complete picture of the nature of and motives for vertical integration. For one thing, there may be other sources or types of transaction costs at work, some of which may not be easily subsumed under the rubric of measurement costs. In particular, the transaction costs of economic change may help shape the pattern of vertical integration in an industry. Explanations that center on the costs of economic change, including both the quantitative change of a rapidly growing market and the more qualitative change of technological and organizational innovation, are what we might call "dynamic" transaction-cost theories.

Morris Adelman was one of the first writers to associate vertical integration with economic change.[5] In a rapidly growing industry, he argued, suppliers of intermediate goods may not be able to expand quickly enough to meet the needs of the producer of final goods, thus motivating that producer to integrate backwards.[6] Morris Silver has more recently recast this "bottleneck" explanation in informational terms.[7] Economic change, he argues, often involves not merely quantitative growth but also qualitative change-innovation in the Schumpeterian sense. Existing firms with the expertise to help an innovator could find the strangeness of a new idea so overwhelming that they would dismiss it out of hand. Rather than devoting effort to catching the attention of suppliers and converting them to the idea, the innovating firm could find it cheaper to acquire the capabilities necessary to produce its own inputs or handle its own marketing. After the innovation has become more familiar and outsiders have been able to assimilate the innovator's ideas, later entrants would not face the same bottlenecks and could draw more freely on outsiders. The value of external economies, a concept much appreciated by Alfred Marshall and his British followers, therefore varies according to whether a firm is an innovator or a follower.

In some cases, vertical integration is effectively forced on innovative firms, who would have preferred to economize on managerial resources by buying complementary products and services on the open market. In other cases, however, we might say that integration results from an entrepreneurial decision to supersede an existing network of complementary activities. The innovations of refrigerated meat packing in the nineteenth century or containerized shipping in the twentieth superseded inferior systems that might easily have been judged perfectly

[5] Richard N. Langlois, "Economic Change and the Boundaries of the Firm," *Journal of Institutional and Theoretical Economics, 144* (Sept. 1988), pp. 635-57.

[6] Morris Adelman, "Concept and Statistical Measurement of Vertical Integration," in *Business Concentration and Price Policy* (Princeton, 1955), pp. 318-20.

[7] Morris Silver, *Enterprise and the Scope of the Firm* (London, 1984).

efficient by the standards of neoclassical competition theory. But in the supersession of those systems lay even greater (internal) economies? Disintegration using Marshallian external economies and vertical integration are active strategic alternatives open to the entrepreneur.

A principal argument for the superiority of vertical integration over a network of decentralized producers is the potentially greater ability of the large firm to implement innovation, especially large-scale innovation of a systemic sort. It can be equally well argued, however, that backward integration might retard change by tying firms to a single source of innovation when access to the market would have provided a wider range of choices. As was the case in the British shipbuilding industry before 1914, the presence of a large number of competing suppliers can be an important external economy if they are all committed to improving their products and providing other design services to benefit their customers, especially to the extent that the latter are small and unable to afford research into improving a wide range of inputs.' Indeed, William Abernathy has used the automobile industry as precisely an example in which vertical integration led to stagnation because it isolated firms from innovation, especially product innovation, they might have obtained in a more decentralized system."

One would in general expect improvements in process technology that depend on economies of scale for their viability to come only after customers have accepted a dominant product technology. When there are several variations of a product competing in a limited market in the early stages of the product life-cycle, producers may not invest in a capital-intensive technology that will pay only at high levels of output. On the other hand, it is possible that the final choice of a product variation will be made on the basis of process innovation that reduces prices to a point at which other considerations become largely irrelevant to customers. Because of the increased uncertainty faced by a producer who adopts a capital-intensive process technology when the nature of the product is still in flux, however, the tendency would be for firms to wait for a resolution of consumer preferences before committing themselves to mass production.

[8] William Lazonick has lately put great emphasis on this point in presenting a theory of the vertically integrated firm that is otherwise much in the spirit of the Adelman-Silver view. See William Lazonick, "The Social Determinants of Technological Change: Innovation and Adaptation in Three Industrial Revolutions" (paper presented at the Second International Conference on the History of Enterprise, Terni, Italy, October 1-4, 1987); and William Lazonick, "The Causes and Consequences of the Modern Corporation: Innovation and Adaptation in the Theory of the Firm" (paper presented at the U.C. Intercampus Group in Economic History Conference, University of California, Santa Cruz, April 29-May 1, 1988).

[9] Sidney Pollard and Paul Robertson, *The British Shipbuilding Industry*, 1870-1914 (Cambridge, MA, 1979), p. 92. For a general theory of innovation in a network of suppliers and producers, see Eric von Hippel, *The Sources of Innovation* (Oxford, 1988).

[10] William Abernathy, *The Productivity Dilemmu: Roadblock to Innovation in the Automobile Industry* (Baltimore, 1978).

Moreover, the value of external economies can differ significantly between innovator and follower. Whereas an innovating concern might make do by purchasing goods or services developed for other purposes by outside suppliers or distributors when capital is limited, adaptive followers would not be faced with the same disability and could plug into existing networks with few if any penalties. As David Teece suggests, the advantages (or disadvantages) of innovative leadership will depend on, among other things, the ease of imitation in the industry and the nature of the assets complementary t o the innovation.[11] If imitation is easy, such as when patents or trade secrets do not effectively protect the innovator, an innovating firm may be at the mercy of competitors (or even suppliers) who can enter quickly and take cheap advantage of the capabilities created at high cost by the innovator. To profit well in such circumstances, the innovator would have to own many of the assets complementary to the innovation. This need not imply vertical integration in any meaningful sense, however, since in principle the innovator need only take financial positions in the complementary assets-long positions in those likely to appreciate and short positions in those likely to depreciate. Only when the complementary assets are highly specific to one another do problems of "hold-up" necessitate joint ownership and control.

THE AUTOMOBILE INDUSTRY: HISTORY

Robert Paul Thomas has divided the early history of the American automobile industry into four periods: the pre-1900 era of invention; the era of product development from 1900 to 1908; the era of rapid expansion from 1908 to 1918; and the era of replacement demand from 1918 to 1929.[12]

In the era of invention, the focus of all manufacturers, if they could be called that, was almost exclusively on product innovation. Here the industry came closest to the pattern suggested by Young and Stigler. Car makers like Winton, Olds, and Ford were craft shops, all highly integrated in the sense that they made, or more accurately, improvised, most of their own parts. One well-known example is Henry Ford making his first cylinder out of a piece of pipe. What distinguishes the automobile industry from such earlier examples as printing or machine tools is that this early vertical integration and use of crafts techniques disappeared very quickly once genuine commercial production began in

[11] David J. Teece, "Profiting from Technological Innovation: Implications for Integration, Collaboration, Licensing, and Public Policy," *Research Policy*, 15 (Dec. 1986), pp. 285-305.

[12] Robert Paul Thomas, *An Analysis of the Pattern of Growth of the Automobile Industry, 1895-1929* (New York, 1977), pp. 6-8. For a more detailed account of the history of vertical integration in the auto industry, see Paul L. Robertson and Richard N. Langlois, "Innovation and Vertical Integration in the American Automobile Industry, 1900-1940" (working paper, economics and management dept., University College, University of New South Wales, Jan. 1989).

earnest. By the turn of the twentieth century, the capabilities to manufacture parts adaptable to the automobile already existed in the American economy, and these external economies obviated the early stage of vertical integration that had been thrust on many nineteenth-century industries. At the same time, the novel and experimental nature of the product made financing difficult and left firms starved for capital, further working against large-scale integration. Most importantly, perhaps, the novelty of the product meant that a dominant design paradigm had not yet coalesced in the minds of customers, making a commitment to large-scale integrated production extremely risky.

In many respects, the expansion of the industry in the first two decades of the century resembled a textbook example of the benefits of external economies. Early in its development, the industry became concentrated in Michigan, Ohio, and Indiana. Localization allowed assemblers to take advantage of external economies that flowed from concentrations of suppliers. Although the degree of vertical integration varied from firm to firm, virtually all automobile companies began as assemblers rather than manufacturers. Early cars could be easily put together from components developed for other purposes, such as bicycle wheels, or from variations on known themes, such as wooden bodies. Particularly in the United States, the term "horseless carriage" was an accurate description of the design concept that lay behind many cars. They were seen initially as improvements on existing forms of transport rather than as a new product that needed to be thought out from scratch. As a result, finding willing components manufacturers and dealers does not seem to have been a problem, although raising capital was.

In the era of rapid expansion, the design of automobiles began to coalesce into a more or less unified paradigm. One important move toward standardization came with the replacement of basic American designs with more advanced and complex French ideas. Although Americans originally conceived of cars as buggies without horses (like the curved dash Oldsmobile runabout), the French thought in terms of a road locomotive. French autos had heavier, multicylinder engines mounted at the front, multigear transmissions, and differentials, all attached to a steel frame. The preference of wealthy customers for French imports induced American firms to begin manufacturing cars on the French pattern in 1901. After 1905, the preference for cars on the French model was reflected in a shift of demand patterns toward more expensive cars. This left an enormous unfilled demand for a model that delivered the benefits of the French concept at a price comparable to the Oldsmobile. Ford's success with the Model T was based on precisely that combination of modern features and low price.

As is typical in the product life-cycle, the radical changes in product characteristics that occur in an early era gave way to incremental

product change. The focus of innovation began to shift to process innovation, including the development of mass automobile production using the moving assembly line, which was the ultimate key to the success of the Model T.

Mass production involved Ford in the large-scale design of machine tools, at least some of which also seem to have been manufactured by the firm. Although Thomas contends that the automobile industry was well served by outside machine-tool firms in its early days, the greater extent to which Ford took the division of labor when he developed mass production and the assembly line called for a far wider use of special-purpose machinery than had been necessary for other firms.[13] To meet their new needs, Ford and his engineers were forced to develop many of these devices themselves. As Silver hypothesizes, this was not because the technical advances in the tools were beyond the understanding of independent tool makers; rather, it was because only the men at Ford understood the uses to which the new machines were to be put. It would have cost more for Ford Motor Company to explain the process of automobile -manufacture to the tool makers (a process that was in any case still evolving in the minds of Ford and his assistants) than it did to diversify into tool design on an ad hoc basis. The Ford techniques could have diffused to outside suppliers and indeed quickly did so, but the massive demand for the Model T generated by Ford's low-price policy made even small delays costly. In this sense, integration was a strategy forced upon Ford as much as it was a conscious entrepreneurial decision.

There is also more to the story than the radical newness of the Ford machines. Central to the innovation of mass production is its systemic character-its organization and timing of production activities-which had the effect of making vertical integration self-reinforcing in certain ways. Because much of the assembly activity was collected in one place, Ford engineers were able to perceive opportunities for grasping economies of scale that would not have been apparent to decentralized parts suppliers. An example of this is the development of metal-stamping techniques. Pressed-steel frames, which were much stronger than their structural-steel predecessors, were generally adopted in the industry by 1909. A couple of years before, however, the managers of the John R. Keim Mills, a producer of pressed steel in Buffalo, had approached Ford with a proposal to supply axle housings. Ford agreed, and by 1908 Keim was supplying a variety of parts. Engineers from the two firms cooperated in perfecting the pressing process; Ford invested heavily in machinery for the Buffalo plant, and eventually purchased Keim in 1911. When the workers at the Keim plant struck the following

[13] Thomas, *Pattern* of *Growth,* p. 269. Compare David A. Hounshell, *From the American System to Mass Production,* 1800-1932 (Baltimore, 1984), p. 233.

year, Ford ordered all of the machinery removed from Buffalo and shipped to Highland Park. The technology sparked the interest of Ford's engineers, and metal-stamping was soon adapted to producing crankcases, axles, housings, and even bodies. Thus integration initially motivated by a desire to avoid hold-up problems and supply disruptions led to further integration because of the possibilities for innovation it opened up in the Ford environment.

It is important to emphasize that integration at Ford during that period was a phenomenon of novelty and innovation. As soon as the outlines of the assembly-line innovation were clear, it diffused rapidly to independent parts suppliers, a process aided by Ford's openness to trade journalists. Moreover, Ford engineers soon learned that the production process could be efficiently decentralized.[14] Once the innovation of mass production of parts became assimilated, and as the extent of the market grew, centralization became more costly and less beneficial. At Ford, the ensuing decentralization took place within a vertically integrated structure for largely historical reasons.

Significantly, the one other firm to attempt a high level of integration before 1920 was also run by a strong-minded visionary. In his two forays into General Motors, William Crapo Durant brought together not only a large number of automobile assemblers but also a collection of components suppliers.

Like Ford, Durant believed strongly in the latent demand for automobiles, especially at the lower end of the price range. Unlike Ford, however, Durant's innovations lay not in production of the vehicles but in the creation of a marketing and distribution network. From the point of view of process technology, General Motors (GM) was very much a follower. For both of these reasons, integration at General Motors was of a very different sort than that at Ford. Under Durant, there was no systematic policy behind the expansion of General Motors, and no attempt to integrate the various divisions into a coordinated manufacturing group on the Ford model. It was only after a decade of wrestling with their legacy that Durant's successors were able to rationalize the firm and move toward genuine integration.

Vertical integration at GM also arose from motives that were quite different from those at Ford. Alfred Chandler and others have cited as a principal reason for vertical integration Durant's desire to ensure adequate and timely supplies.[15] This is no doubt part of the story. But an arguably more important motive was Durant's desire to appropriate the benefits of his vision and his marketing innovations by taking positions in as many assets complementary to auto manufacture as possible. Durant was by no means a passive stockholder in these

[14]Henry Ford with Samuel Crowther, *My Life and Work* (Garden City, 1923), pp. 83-84.
[15] Alfred D. Chandler, *Strategy and Structure* (Cambridge, MA, 1962), p. 116.

enterprises. Capital markets for automotive securities were not yet developed enough to allow appropriation by passive investment: the first appearance of automotive stock on the New York exchange was the sale of General Motors voting trust certificates in 1911.

By the 1920s, the market for cars had changed radically. As both product and process innovation became less important, cost relationships between automobile manufacturers and potential suppliers became increasingly prominent. By the end of that decade, a modest reduction in the degree of vertical integration had begun, based on a new appreciation of the flexibility that could be derived from outside suppliers. Between 1922 and 1926, the importance of components purchased from outside suppliers had declined from 55 percent to 26 percent of the wholesale value of American motor vehicles? Even in dollar terms, the value of components had declined despite a near doubling in the total value of finished vehicles. By 1927, however, purchases of some components were already beginning to increase, and vertical integration became progressively less important in the early years of the Depression."

This is surprising in view of earlier trends. It is especially interesting for the early years of the Depression, when one would expect a capital-intensive firm to protect its own output of components at the expense of independent firms. The explanation for this pattern involves several key factors: the growth of the market for spare parts; organizational innovation; and the introduction of the annual model change.

Despite the increasing degree of vertical integration in the early 1920s, the aging of the national fleet of automobiles led to a proliferation of parts firms to supply the replacement market. While employment in automobile plants proper decreased by 43,628 over the period from 1923 to 1925, employment in parts firms grew by 64,628.[18] Because of the low investment required to produce a limited range of items, entry was easy and most of the new parts firms were small or medium sized. Competition among them reduced costs to levels that frequently could not be reached by the automobile firms themselves, particularly as the need for flexibility grew near the end of the decade.[19] As novelty became an important selling factor, firms were forced to change models more frequently. Moreover, the number of models produced by each firm

[16] Lawrence E. Seltzer, *A Financial History of the American Automobile Industry* (Boston, 1928), p. 59.

[17] Norman G. Shidle, "Trend Toward More Car Models Helping Outside Suppliers," *Automotive Industries* (July 30, 1927), p. 146; and Harold Katz, *The Decline of Competition in the Automobile Industry, 1920–1940* (New York, 1977), p. 260.

[18] Seltzer, *Financial History*, p. 50.

[19] Eva Flugge, "Possibilities and Problems of Integration in the Automobile Industry," *Journal of Political Economy, 37* (Apr. 1929). p. 166.

multiplied. This reduced the economies of scale open to automobile manufacturers, giving further advantage to the small parts firms.

Organizational innovation also helped shape the pattern of vertical integration in this era. In the twenties, the principal innovation had the effect of shifting the margin between firm and market (to use Coase's metaphor) in the direction of the market. Following the scares of the early decade, firms sought to keep their inventories small by purchasing only amounts that were immediately necessary. This led to the adoption of "hand-to-mouth' ' purchasing, a more colorfully named predecessor of the now-fashionable "just-in-time" practices. And, if modern proponents of this technique are correct, the adoption of hand-to-mouth purchasing had efficiency advantages over and above its ability to economize on inventories.

But the most important factor influencing the pattern of vertical integration in the twenties and thirties was the annual model change and the product innovation that accompanied it. The used-car market opened up competition that new-car makers had not faced before. In order to give their new cars an advantage over used models, car makers turned to increased product innovation. The decade of the twenties amounted to, in effect, what we might think of as a discontinuity in the automobile's product cycle. Instead of a continued evolutionary pace in product and process innovation, relatively radical product innovation had again become necessary. And here, as in the early days of the industry, those firms able to draw on the product innovations of decentralized suppliers gained at least a temporary advantage. It was thus the small, less integrated firms who were best able initially to succeed in the era of replacement demand. Abernathy believes, for example, that Chrysler's "strategy of design flexibility and shallow vertical integration proved very successful in the prewar period, when the rate of technological change in the product was rapid.''[20]

By contrast, Ford had the greatest difficulty in adapting to the changed conditions. Despite notable evolutionary changes in the Model T in the mid-1920s, Ford proved incapable of making a smooth transition to a new model. The firm was obliged to suspend production for nine months and faced severe teething problems thereafter. The Model A contained a number of important components that either were not manufactured by Ford or were so different from those of the Model T that the existing plant and machinery could not be converted.

The new regime thus had the effect not only of promoting the fortunes of firms employing relatively less vertical integration but also of prodding the highly integrated to become less so. "Major product innovations," as Abernathy argues, "destroy old paths of backward vertical integration and create opportunities for new ones. Product

[20] Abernathy, *Productivity Dilemma,* p. 37.

innovations thus generally reduce the degree of backward integration."*' This certainly seems to have been the case at Ford. Moreover, despite the legend that Ford did not keep records, the firm knew as early as the mid-1920s that some components could be purchased more cheaply than Ford manufactured them. This message was reinforced by the transitional problems that the firm experienced. By the early 1930s, Ford executives were convinced that a smooth introduction of a new model was more likely if the burden of change were spread over a large number of firms.

Of all the firms in the industry, however, General Motors was in the best position to take advantage of the new economics of replacement demand. Unlike Ford, GM had not optimized its production process for the single-minded manufacture of one unchanging model. Under Ford defector William Knudsen, GM installed a system of "flexible" mass production that allowed for model changes without trauma.** GM was also less integrated than Ford, and in a much looser way, which made it easier to take advantage of outside suppliers' ideas. Moreover, the company had always had something of a strategy of product innovation and model change. Dating back to Durant's early days at Buick, GM chose not to compete by reducing price on a standardized model, and instead improved the performance and amenity characteristics of a car with a more or less constant price. At the same time, General Motors was also in a better position than its smaller rivals to appropriate the benefits of product innovation and model change. GM possessed a richer package of complementary assets, notably its large distribution, dealer, and consumer-finance network. This gave the firm an ability to benefit, in a classic way, from both its own innovations and those developed by smaller competitors and independents.

And when GM lacked some of the relevant assets, it moved to acquire them. The most important example of this was the acquisition of Fisher Body, a pioneer in the development of the closed body. Although this was still a relatively unpopular option at the time, GM decided to tap into the technology by acquiring 60 percent of Fisher in 1919. At the same time, GM signed a ten-year agreement to buy all of its closed bodies from Fisher, and in return insisted on stringent contract terms in order to avoid opportunistic behavior. As overall demand grew and preferences shifted toward closed bodies, General Motors became discontented with its arrangement with Fisher, which had become an extremely important supplier. GM believed the price was too high in view of the lower unit capital costs that Fisher Body incurred, and resented Fisher's refusal to locate its body plants adjacent to GM assembly plants, a move that would have benefited General Motors but

[21] Ibid., p. 64.
[22] Hounshell, *American System,* pp. 265-66.

would also have limited Fisher's flexibility in dealing with other customers. Ultimately, General Motors took over the remainder of Fisher Body in 1926 in order to protect its interests.[23]

Benjamin Klein, Robert Crawford, and Alchian explain this episode narrowly in terms of the highly specific assets Fisher and GM had to commit to the venture and the attendant threats of hold-up and expropriation of rents. But it was not the specificity of the assets that made integration look better to GM than contractual alternatives. As we saw, the later years of the twenties and the early thirties saw a trend toward increased dependence on suppliers for parts, and many of these suppliers certainly employed assets highly specific to the task and produced components critical to the assemblers' flow of production. What led to high transaction costs in the case of Fisher was the very rapid shift in demand toward the closed body. In contrast to Ford's experience with the Model T, General Motors was not in the forefront of innovation in this instance. Closed bodies had been available for years and had also been adopted by competitors. In the long run, GM might well have been able to rely on outside body manufacturers as all other major auto firms, including Ford and Chrysler, were able to do to some extent in the 1930s. In the mid-1920s, however, GM's larger volume and emphasis on flexibility made it especially vulnerable to hold-up problems. Because it was not an innovator, GM had to compete with other auto makers for scarce closed bodies as demand accelerated. If it could not keep pace with rising consumer demand for closed bodies, GM stood to lose the benefits that its policy of up-to-date styling was supposed to provide. This might in fact have led to a permanent loss in market share if the public bought other brands by default. Rather than pay a premium to guarantee a reliable supply in a seller's market, GM decided to invest a relatively small sum to gain full control of a supplier in which it already held a majority interest.[24]

Note that although sunk cost may be sunk, GM's prior investment made vertical integration into body building cheaper than for other firms that needed to begin from scratch. As the additional investments required to overcome the problem of hold-up were greater for other auto makers, those firms were more willing to incur whatever transaction costs derived from association with independent body manufacturers,

[23] Benjamin Klein, Robert Crawford, and Armen Alchian, "Vertical Integration, Appropriable Rents, and the Competitive Contracting Process," ***Journal of Law and Economics, 21*** (Oct. 1978), pp. 308–10.

[24] In a recent re-examination of the Fisher Body case (which came to our attention after this paragraph was written), Klein presents an interpretation substantially in accord with our own. Benjamin Klein, "Vertical Integration as Organizational Ownership: The Fisher Body-General Motors Relationship Revisited," ***Journal of Law, Economics, and Organization,*** 4 (Spring 1988), pp. 199-213.

especially since the car makers could also offset these costs to a degree with the external economies of a competitive market.

Detailed estimates of vertical integration are not available for the years of the 1930s, but Harold Katz believes that the use of outside suppliers was reduced as firms learned to cope with complexity and change.[25] The effect of reduced innovation on the degree of vertical integration is an area in which theory is not well developed. In the case of automobiles during this period, however, there were exogenous factors that tended to militate in favor of reintegration. One of these was the market for spare parts, an area over which car makers had an incentive to retain control in a period when replacement demand was growing. As original equipment and replacement parts were jointly produced-indeed, were identical-they could be turned out more cheaply if producers could operate on the larger scale allowed by sales to both markets. Moreover, replacement parts have traditionally carried a higher profit margin than original equipment. Because automobile manufacturers forced their dealers to use only authorized replacement parts, they were in a position to pad their profits by producing parts internally.

Another factor favoring integration was organizational innovation. The corporate reorganization that Alfred Sloan put in place at GM in the 1920s is now hailed as a pioneering application of the multidivisional structure (the M-form). It was a structure well adapted to the institutionalized change of the annual-model strategy: it decentralized day-to-day decision-making to an extent that approached that of the market, while retaining a centralized strategic and coordination ability.[26] The M-form innovation thus arguably reduced the costs (increased the benefits) of internal organization, and helped move the firm-market margin closer to the firm.

TOWARD A THEORY OF APPLICABILITY

The early history of the American automobile industry suggests that no single theory adequately explains vertical integration in all periods and'in all its guises. If there is any candidate for a "general" theory, it is among the "dynamic" transaction-cost theories. These theories explain the early integration at Ford and GM, and, in an important sense, they also explain the Fisher Body episode of 1926, which has been a keystone example for the asset-specificity view: it was the "disequilibrium" of a rapid rise in demand for closed bodies, not highly specific assets as such, that, as Klein now agrees, "moved the contrac-

[25] Katz, *Decline of Competition,* p. 258.

[26] Williamson, *Economic Institutions,* pp. 279-85.

tual arrangement outside of the self-enforcing range and made it profitable for Fisher to hold up General Motors."*'

The search for a general theory of vertical integration is perhaps less interesting than seeking a schema that determines when various particular explanations are applicable. What this history of the automobile industry suggests is that certain levels of integration are more likely to prevail at specific stages of the industry life-cycle and under specific conditions of demand, economies of scale, and appropriability . We can summarize these relationships in a tentative way as follows.

In the early stages of an industry's development, when product technology is still in flux and markets are small, innovating firms can be expected to eschew vertical integration if possible. In part this is because these firms will not need enough of many of their inputs to produce them efficiently; moreover, small innovating firms rarely have excess financial or managerial resources to devote to multiple stages of production. If outside suppliers cannot be found, however, because the needs of the innovating firm are either too specialized or too difficult to communicate, then vertical integration must occur if the innovation is to succeed. This is particularly likely when the innovation is in the nature of a systemic rearrangement of production technology or organization.

As product innovation slows down, new entrants will not face the same need to integrate vertically because there will now be sufficient economies of scale and knowledge of the product to permit the existence of independent suppliers. Integrated innovators may spin off intermediate stages at this point and rely on outside suppliers shared with later entrants, or they may decide they cannot recoup their investments through selling off suppliers and therefore remain more heavily integrated than new competitors.

A reduction in the rate of product innovation as an industry matures increases the scope for process innovations leading to increased economies of scale. Firms that embark on strategies of mass production may find it desirable to increase either forward or backward vertical integration at this stage in order to protect their sources of supply and distribution networks so they can achieve economies of throughput and rapidly amortize their heavy fixed investments. The extent of vertical integration should nevertheless be tempered when the firm can take advantage of markets in which transaction costs are moderate.

Further innovation may be undertaken either through internal development or by relying on outside suppliers. Economies of scale, both in

[27] Klein, "Vertical Integration," p. 202. In this formulation, it appears that economic change, "uncertainty" in Klein's language, plays as important a role as the degree of asset specificity in explaining vertical integration. This has been noticed before. See Richard N. Langlois, "Internal Organization in a Dynamic Context: Some Theoretical Considerations," in M. Jussawalla and H. Ebenfield, eds., *Communication and Information Economics: New Perspectives* (Amsterdam, 1984), pp. 31-33.

the production process and in the R & D function, may be important in determining the choice. In particular, changes in the relative economies of scale of intermediate and final products may make it desirable for a vertically integrated firm to revert to independent suppliers. When a change in process technology creates economies of scale for intermediate goods that exceed the requirements of any manufacturer of a final good, then a shift to outside suppliers can be profitable. A move in the opposite direction can also bring about disintegration because a reduction in the relative economies of scale of producing an intermediate good may eliminate the advantage gained from backward integration and open up attractive possibilities for playing competing suppliers off against one another. This may increase the range of intermediate-product innovations available to manufacturers of final goods, as well as yield cost advantages, if the supplying firms are capable of undertaking research and development. Otherwise, the final-goods producer may continue to bear the responsibility for innovation in intermediate goods even if it buys from independent suppliers. History may also play an important role here. A firm that starts out highly integrated may develop a bias toward certain kinds of innovations (for example, process innovations of a systemic sort) that further reinforce its integrated structure. And a firm that is less integrated may rely on innovation from decentralized suppliers in a way that reinforces a less integrated structure, at least in the short run.