Emergence of the Theory of Industrial Organization, 1890-1990

T.A.B. Corley[1]
University of Reading

Business history, wrote Louis Galambos in 1966, "tries to achieve a synthesis of the characteristics and effects of decentralized economic decision-making on both economic change and economic growth." Like other scholars such as Richard Wohl (1954), he was exercised about what seemed to be business history's chief problem: how to weave the diverse activities of a multitude of independent firms into a coherent account of national, regional, or other developments. While a theoretical structure would clearly strengthen any synthesis, ready-made theories from other disciplines, especially economics, would fail to reflect the richness of business history case studies. Hence business historians would have to make do with eclectic and partial systems of generalization. Yet those hardly would add up to a broad intellectual framework, capable of making case studies more meaningful and of testing economic historians' broad conclusions.

Despite the output of studies since the 1960s, this fundamental dilemma seems to have been infrequently discussed in business history literature. Nearly three decades later, therefore, it is none too early to reopen the issue, but this time from the theoretical side. Is the theory of corporate activity still too rigid to help business historians, and if so, could anything be done to make the theory a more serviceable instrument for them?

To throw light on these questions, it would not be enough to examine the current state of theory. Instead, a broader "history of economic analysis" approach will help to give perspective to the inquiry. By studying the work on the firm by certain mainstream economists of the past, it should be possible to see how their observations evolved into formal

[1] I am very grateful for helpful comments to the participants at the Business History Conference at Baltimore, especially Michael Best, Leslie Hannah and Richard Langlois, and also to Mark Casson and Michael Waterson of the University of Reading. To save space, the very lengthy list of references has had to be omitted. These references can be obtained from myself at the University of Reading, RG6 2AA, UK, but should be traceable through any university library catalogue and the *Index of Economic Journals/Articles*.

BUSINESS AND ECONOMIC HISTORY, Second Series, Volume Nineteen, 1990.
 ISSN 0849-6825.

theories which sometimes erred too much toward over-generality or else over-rigidity.

Included in this discussion will be the writings of Alfred Marshall from 1890 onward, and the posthumous but pervasive influence of Augustin Cournot, which in due time led to Edward Chamberlin's and Joan Robinson's monopolistic or imperfect competition theories of the early 1930s. By confining the firm to an over-restrictive straitjacket, these latter theories deprived scholars of a flexible and realistic body of concepts just when business history studies were beginning to proliferate in the U.S. and Britain after 1945. Economists' subsequent theoretical and empirical studies have helped to illuminate many facets of corporate behavior, but without providing an adequate analytical substitute for the simple Chamberlin-Robinson models. In conclusion, the question is raised as to whether economists, as well as business historians, would benefit from a radical overhaul of industrial organization theory as a means of overcoming the dilemma mentioned earlier.

Marshall and His Era

Alfred Marshall was a pioneer in two respects. He was the first to assemble into some kind of corporate theory the building blocks contributed by many predecessors. Moreover, his unique technique combined mathematical analysis with a practical knowledge of how business and commerce worked.

His basic ideas on the firm centered around competition which he saw in terms of an activity or a process rather than in modern structural terms as portrayed in static sales or average revenue curves. While acknowledging the existence of single or collective monopolies in the private sector, he strongly believed in the pressure of competition through free entry into the industry which helped to curb tendencies toward monopoly. While aware of the power of combative advertising to enhance artificial product differentiation, he thought sales curves could be sloping also as the result of genuine consumer goodwill. Competition bred uncertainty over rivals' responses to a given firm's policies. Whereas the demand for a product would equal supply at the market-clearing price and therefore the industry would be in equilibrium, each firm would not be in equilibrium since it was in a constant process of growth or decline. Marshall therefore chose a representative firm to provide some typical supply curves for his value theory. Economists have never yet properly examined the empirical question of whether firms are actually in equilibrium. However business historians know that firms making for stock use inventories and backlogs to balance supply and demand.

The entrepreneur was central not merely to short-term operations but also to the firm's long-term survival and prosperity. Marshall was the first economist specifically to integrate the entrepreneur into his analysis of value, adding organization to the existing factors of land, labor, and capital. The factor reward, pure profit, was not contractual but residual. While the more successful entrepreneurs might enjoy a rent of ability, competitive pressures would tend to reduce profits to a normal level, sometimes allowing for the

difficulties of work done and the degree of risk undergone. Since in the real world there was not a perfect market in information, entrepreneurs' ability to win was helped by superior knowledge and by experience acquired over time.

To achieve success, entrepreneurs constantly had to adapt their firms to changing circumstances. Improved machinery would make production cheaper and more accurate, while a progressive subdivision of labor would also help to keep down costs. Marshall sought to study the firm's activities as a whole, not merely the pricing and output decisions but also the growth and technical progress. While many of his assertions were subsequently eroded by the analytical trends to be discussed below, some insights and hints did anticipate later work.

Cournot's Legacy, 1890-1933

After 1870 the essentially literary subject of political economy transformed itself into the mathematically-based social science called economics. This process rapidly accelerated after 1890. Hence Marshall's rich verbal insights, based on detailed personal knowledge, were progressively set aside in favor of analytically rigorous models concentrating on limited aspects of the firm's activities. Those models owed much to Cournot's work on value theory, published in 1838 but largely neglected until the 1870s. Cournot analyzed firms as creators of value. He began with the monopolistic case and progressively extended the number of producers in the market until he reached the opposite pole of unlimited competition. At this pole, each firm contributed too small a proportion of the whole to affect the going industry price. In diagrammatic terms not drawn by him, monopoly was associated with a sloping sales curve and unlimited competition was associated with a horizontal curve. Cournot discussed duopoly, suggesting that self-interest would induce the two rivals concerned to reach a determinate and mutually advantageous solution. However, he failed to analyze the commonest market form in advanced economies, namely oligopoly.

Marshall had discussed monopoly purely in terms of a gas undertaking, with sales and cost curves yielding the unique point of profit maximization. He may have derived his diagram from Dionysius Lardner, who in *Railway Economy* (1850) had produced a comparable model based on total rather than average values. The practice of using marginal values came much later. Marshall did not regard private-sector monopolies as worth analysis.

In the U.S., despite the passing of the Sherman Anti-trust Act of 1890, economists were either lukewarm or actively hostile to the act, and consequently felt little interest in pursuing the analysis of monopolies. John B. Clark in 1899 merely acknowledged the conceptual difficulties in integrating monopoly into static theory. Competition, on the other hand, was the subject of a lively debate. In 1906 Henry L. Moore criticized Marshall for treating any deviation from the competitive mode as a friction to be disregarded, and stressed the urgent need to explore the intermediate area between the two poles. Among British economists, Arthur C. Pigou

(1912) implicitly agreed with Moore's view that private monopolies or trusts were common enough to merit serious analysis. However, unlike Cournot he believed that oligopolistic conduct could not easily be analyzed because any solutions were indeterminate.

Soon after Marshall's death in 1924, the anatomy of the firm began to appear in a shape recognizable to modern scholars, the skeleton being the revenue and cost schedules, and the market form being indicated by the slope of the sales curve. Economists portrayed average cost curves as U-shaped, to reflect scale economies in the downward leg and diminishing returns to scale thereafter. Concerned as they were to refine theory along the newer mathematical lines, they chose to disregard Marshall's scruples and assumed that firms in general--and not merely statutory undertakings--would aim to be in equilibrium at the level of output where profit was maximized. They were plainly influenced by Marshall's and Lardner's diagrams showing equilibrium in gas and railway enterprises respectively. Whereas public monopolies could, within certain limits, calculate a sales curve in the absence of close substitutes, commercial firms were unable to do so because of uncertainty about future sales and about rivals' reactions.

This refining process for theory steadily increased after 1925. Economists assumed that most firms were in perfectly competitive markets, and hence would seek to produce at the level where the horizontal sales curve was at a tangent to the average cost curve. Yet observation showed that real life firms operated on the increasing returns leg of the cost curve, able to reduce average costs if they raised output. Piero Sraffa, author of a lengthy 1925 article in Italian on peripheral matters, was asked by the editors of the *Economic Journal* to produce an English version. Sraffa included a postscript designed to solve the increasing returns problem by suggesting that firms in general were monopolists and not perfectly competitive, and thus free to alter their prices at will without wholly losing market shares.

Sraffa's English article, published in 1926, seemed to provide an approach that covered all market forms short of perfect (or unlimited) competition. Then in 1928, while offering the proposition that Marshall's representative firm would be in equilibrium even if all other firms in the industry were not, Pigou brought back into mainstream analysis the geometrical relationship between marginal and average cost. This relationship, first put forward by Francis Y. Edgeworth in 1913, showed the one curve intersecting the other at the minimum average cost point. By 1932 Richard Kahn had pursued the theoretical implications of Sraffa's rather densely packed logic to the point where Joan Robinson was able that year to combine his analysis on the demand side with the Edgeworth-Pigou cost curves into a self-contained model, in which-for the first time-*every* firm was held to be in equilibrium. Scholars' unease at her many heroic assumptions were hardly assuaged when her extended treatment appeared in 1933. For instance, she assumed, without empirical justification, that each firm was its own monopoly, so that the all-important oligopolistic reactions of rival firms could be neglected. She admitted at the outset to be offering no more than a box of tools, and was thus content to draw conclusions *from*

geometrical diagrams rather than arguing *to* those diagrams: a practice deplored by Marshall and others.

In contrast with Joan Robinson's ultimately dead-end analysis of 1932-1933, Edward H. Chamberlin's *Theory of Monopolistic Competition* (1933) has been hailed as truly revolutionary. Similar to Lardner, he arrived at his theme through studying practical questions about railway rates, but his methodology owed much to Cournot. That did not deter him from incorporating into his analysis such highly relevant topics as oligopoly, selling costs, product differentiation, and quality variation. Robinson and Kahn made bold attempts to eliminate entrepreneurship as a productive factor by claiming that its marginal value product was negative. Chamberlin however was convinced that entrepreneurs had a genuine function in having to judge the appropriate degree of interaction with rivals in the group, or industry, with adjustment taking place over time. Unfortunately, his pioneering attempt to analyze the intermediate area of oligopoly, in this presentation of a small-group model, was scotched by Robert Triffin (1940). Triffin set out to restate the core of Chamberlin's analysis in terms of Leon Walras's general equilibrium theory, and in the process denied that the concept of the group or industry had any usefulness because of product differentiation. Chamberlin later accepted these criticisms, and thereby robbed his analysis of one of its most realistic features.

Regrettably, these high-profile treatments by Robinson and Chamberlin diverted economists' attention from some important contributions to the understanding of the enterprise. Frank H. Knight in 1921 and Joseph A. Schumpeter in 1934 had assigned crucial roles to the entrepreneur, such as coordination, decision-making, bearing uninsurable risk, and promoting innovations. Knight maintained that the entrepreneur was essential because of uncertainty. In 1934 Adolph A. Berle and Gardiner Means explored the consequences of ownership being separated from control in joint-stock companies, while three years later Ronald H. Coase made an equally path-breaking distinction between the market and the firm, with the price mechanism being superseded in the latter. All these ideas stimulated a number of significant developments in corporate studies, as will be seen below.

Era of Controversy, 1933-1951

An unexpected consequence of the developments up to 1933 was the gradual realization of the existence of an entirely new academic subject, the theory of the firm. Earlier economists' work on corporate topics had been undertaken merely to clarify aspects of value theory. Only by 1942 was the new subject well enough established for Kenneth E. Boulding to maintain that the theory of the firm had existed since 1932. Boulding himself was not unduly impressed with the marginalist analysis, on the grounds that it simplified the exposition of theory but at the cost of neglecting some highly important aspects such as oligopoly and uncertainty. These criticisms foreshadowed a very bitter controversy that soon erupted between the supporters of Robinson's and Chamberlin's marginal approach and those who rejected it as essentially rigid and unrealistic.

In 1939 some Oxford economists reported on a research project that had involved questioning businessmen on how price and output decisions were reached. They found that two paramount concerns for these businessmen were uncertainty and oligopolistic relations, both--as Boulding had pointed out--neglected by the 1933 theorists. Price therefore had to be fixed according to full costs, and as much as possible would be sold at that price. A younger economist, Philip W. S. Andrews, later slightly modified this full-cost principle to a normal-cost one, which would reflect also the current buoyancy of the market. In the U.S., Richard A. Lester (1946) and Robert A. Gordon (1948) also challenged marginalist theories on both logical and empirical grounds, only to come under attack from such upholders of marginalism as Fritz Machlup.

Such controversies took up all too much time of economists working in the field, so that this period can be seen as an interlude before the pace of constructive work accelerated in the 1950s. Some of that work will be discussed in the following section.

The 1950s, Emergence of Industrial Organization Studies

The modern theory of industrial organization was born out of a number of academic projects in the U.S., already yielding significant results by the early 1950s. In Britain the term "industrial economics" was preferred, and important research findings on the subject were disseminated in the *Journal of Industrial Economics*, founded by Andrews in 1952. Of the American contributions, three are examined here.

The leading project was the one at Harvard, where Chamberlin and Edward S. Mason promoted industry-wide studies, helping to test the hypothesis that market or industrial structures determined member firms' conduct and performance. While this structure-conduct-performance relationship featured prominently in industrial organization studies until the 1970s, its importance declined once empirical research yielded only weak practical results and scholars began to accept that the widespread market form of oligopoly made business behavior very difficult to forecast.

One breakthrough was achieved by a pupil of Mason's, Joe S. Bain, who in 1949 suggested that pricing theory should take account of two relatively neglected factors, namely time and potential entry. Today's price might be governed by tomorrow's profit targets, while the *threat* of competition could well be as effective in determining business conduct as the current market structure. In his 1956 book he both categorized barriers to entry and showed how firms facing various heights of barrier could hold prices above minimum unit costs without encouraging entry. Despite attracting some criticism, for instance regarding definitions of barrier heights and the smallness of his sample in the empirical sections, his work demonstrated how research of this kind could test and refine theories.

The second project reflected the emergence after 1945 of a new field of study, the economics of development. A panel of labor economists from Harvard, MIT, Princeton, Berkeley, and Chicago, investigating labor problems in developing countries, identified the lack of an entrepreneurial (as opposed to a mercantile) cadre as a factor impeding such countries'

economic advance. That approach helped to revive academic interest in the entrepreneur, which had been all but struck out of the scenario during the 1930s.

In 1956 one of the Princeton team, Frederick Harbison, drew on the lately neglected views about the entrepreneur put forward by Marshall, Knight, and Schumpeter. Since the quality of entrepreneurship could clearly affect the degree of organizational efficiency, Harbison suggested that so-called inefficiency could be due to entrepreneurs behaving rationally in pursuing other goals than profit maximization, such as social advancement. Efficiency could also be reduced by inadequate knowledge and by inappropriate organizational structures which could lead to loss of effective control over subordinates. These important ideas later were developed further by one of the labor economists' panel, Harvey Leibenstein, in his discussions of organizational or X-inefficiency, which he was to show as being of far greater magnitude than the allocative inefficiencies resulting from, for example, monopoly or tariffs.

In 1957 while analyzing development questions in depth, Leibenstein elaborated some of his views on entrepreneurship and its origins. He stressed the role of knowledge, which could become an economic resource through the entrepreneur's activities, although in a poor economy the cost of acquiring and utilizing knowledge could be too high for an innovation to be exploited there. On the supply of entrepreneurship, Leibenstein believed that entrepreneurs would emerge in the economic system at a rate depending on their anticipation of growth in income per head. Thus in a stagnant economy, there would be few entrepreneurs, but they would increase as the economy embarked on its development. His "growth agent expansion curve" postulated a mutual interaction between expected and actual income growth, set off by steadily increasing numbers of entrepreneurs. He did not at this time draw a demand schedule for entrepreneurs, although in subsequent works he attempted to do so.

The third research project, at Johns Hopkins under Machlup, was on the growth of the firm. In 1952 one of that team, Edith Penrose, stressed the biological analogies in the theory of the firm, favored by Marshall but since discarded in the drive toward rigor under the influence of Cournot's ideas. In her book of 1959, Penrose proved to be a pioneer in two respects. First, in setting out a general theory of corporate growth, she moved discussion away from the scale of the firm to its growth. Second, she focused attention on the firm as an organization in its own right, rather than merely as a unit in a given market structure. For her the engine of growth, whether in one product or by diversification, was the existence of unused managerial services. Penrose recognized the value of good business histories and of evidence from businessmen for testing hypotheses on the process of corporate growth. Since it turned its back on the marginalist theories, her book was attacked by mainstream economists for insufficient analytical rigor. However, it has proved a greater spur to subsequent research than the equilibrium growth models offered by Baumol and by Robin Marris in the early 1960s.

Individual economists were busy in devising more realistic alternative business motives to profit maximization. In 1935 John Hicks pointed to the

"quiet life" as the best of all monopoly profits, and in 1943 Tibor Scitovsky suggested for entrepreneurs the same kind of work-leisure choice that theorists assumed for workers able to vary the hours or intensity of their labor. Business historians are familiar with the tendency of entrepreneurs initially to work all hours, but then to ease off once their firms had been established. In 1958 Baumol put forward the sales revenue maximization hypothesis, since the typical U.S. corporation sought that goal rather than profit maximizing at all costs, but with a minimum profit constraint.

The motive of securing no more than satisfactory profits had been suggested by Gordon in 1948, and followed up after 1955 by Henry A. Simon with his psychological concepts of drives and aspiration levels. In 1963 Oliver E. Williamson's managerial discretion model took account of the separation of ownership and control. Managers, as those effectively in charge of limited companies, would seek to grant themselves rewards well beyond their actual productivity.

By the end of the decade, the ever elusive topic of oligopoly was once again the subject of attention. Two important forerunners had been John von Neumann and Oscar Morgenstern's theory of games (1944), related more specifically to firms' operations by Martin Shubik in 1959, and K. W. Rothschild's article of 1947, which likened oligopolistic rivalry to war, especially in the alternate periods of intense and often costly hostilities, and lengthy spells of inactivity. Paolo Sylos Labini, in an Italian work cited by Franco Modigliani but not translated into English until 1962, analyzed oligopolistic strategies from the viewpoint of aspiring entrants and established firms. The latter's forestalling tactics could pay off even if new firms believed themselves able to produce at costs comparable with those of existing firms. Modigliani helpfully discussed Sylos Labini's work alongside that of Bain. While both contributions were essentially static, they did provide a joint framework for future developments capable of yielding operationally testable propositions.

Recent Developments, 1960-1990

Putting recent events into perspective is always difficult, and in the present case the problem is compounded by the substantial quantity of industrial organization studies in this period. When in the early 1970s, the research impetus of the previous two decades appeared to slacken the National Bureau of Economic Research commissioned four leading scholars to identify current problems and possible future directions of study. Two areas which they specified were the internal organization of the firm and technical progress. Despite being a key determinant of U.S. industrial growth, the latter had been inadequately analyzed, probably being regarded as falling outside the scope of traditional corporate theory. These topics helped to influence thinking in the 1970s, and in turn created an empirical renaissance in industrial organization studies during the following decade.

The work of Alfred D. Chandler greatly boosted the progress of studies on internal organization. During the 1950s Chandler had been researching into the topic of how large-scale business evolved in the U.S. Chandler made a unique contribution to business history by a combination

of three factors. He had a very broad and relevant theme; he took account of certain large economic forces such as demographic and technological changes; and he had a thorough research grounding in the history of the fifty largest U.S. corporations which he studied.

In 1961 Chandler and Fritz Redlich, as former colleagues in the Harvard Research Center in Entrepreneurial History, jointly set out to update the theory of the entrepreneur, hitherto assumed to be one individual or a small team. Their study took account of organizational changes necessitated by the emergence of diversified giants in the U.S. The various categories of enterprise which they described, from the small single-product one to large ones with many products and functions, provided a useful paradigm for scholars. After the article appeared, the journal's editor canvassed eight eminent business historians for their views. That initiative yielded no serious discussion of the authors' main thesis that "business historical material can be used in developing economic theory."

Chandler's *Strategy and Structure* (1962) did offer two concepts that were to be influential well beyond business history: that structure follows strategy, and that large firms needed to evolve an organizational form, such as a multidivisional structure, to permit effective control. It was an economist, Oliver Williamson, who in 1970 drew on *Strategy and Structure* for a distinction between U-form or unitary firms and M-form or multidivisional firms. In *Markets and Hierarchies* (1975), Williamson grafted his exposition of Chandler's concepts on to the distinction by Coase (1937) between the firm and the market, and the concept of transaction costs which had been largely neglected by the post-Cournot economists. Williamson's article of 1981 on "The Modern Corporation" was impressive in covering many facets of the subject and identifying, along Galambos's lines, some of the differences in viewpoint between business historians and economists. Yet the article did not prove as seminal on the business history side, no doubt because Williamson sought to explain the firm's workings chiefly by the overriding need to economize in transaction costs, an approach not very amenable to empirical testing.

Although the second topic of study mentioned by the NBER scholars, namely technical progress, subsequently attracted much academic research, it was considerably extended after the early 1970s. The traditional concept of technological-often mechanical-innovations, arising mainly from research and development, proved to be only one source of corporate advance, to be supplemented by innovations in the fields of marketing, finance, and managerial expertise. The most progressive firms therefore possessed a substantial store of knowledge or information. These broader ideas flourished in the context of yet another new economic subject, the theory of the multinational enterprise (MNE). Normally arising from a home firm's foreign direct investment (FDI), the MNE and its origins received virtually no theoretical attention before 1960, when Stephen Hymer drew a definitive distinction between portfolio and direct investment, the latter being undertaken to earn not interest but profits. Hence, he concluded, MNEs should properly be studied through the theory of the firm.

However, the marginalist models of the 1930s were not robust enough to permit Hymer to build up a satisfactory theory of the MNE. Although

he and successors such as Charles Kindleberger and Richard Caves attributed FDI mainly to monopolistic motives, emphasizing price discrimination through separation of markets, that explanation soon lost ground to more flexible ones. In the 1970s John Dunning at Reading proposed a more eclectic approach that stressed the synergistic motives for producing overseas, with firms perceiving advantages which might be organizational or locational. Other economists were fruitfully applying Coase's ideas to the MNE. They saw a progression, from arms-length marketing of exports to setting up overseas branches, as the internalization of activities formerly undertaken through the market. The new MNEs would enjoy two benefits: profits that had previously gone to others, and reliable quality control. The head office's stock of information could best be safeguarded by being transferred across boundaries while remaining under home control.

Although by the 1980s virtually all large firms were MNEs, industrial organization studies were by no means exclusively MNE-oriented. As Richard Langlois has shown, William Lazonick among others was working towards a theory of the giant entrepreneurial firm along Schumpeterian and Chandlerian lines. Rather than being suspicious of such giants for enjoying market power, Schumpeter had stressed their ability to generate significant innovations. Chandler in turn had emphasized the substantial economies of high-volume production and marketing that the giants were able to achieve. Yet toward the end of the decade, specifically MNE theory was beginning to converge with the old theory of the firm implicitly assumed to be a uninational and single-plant unit. Both of these theories could now be seen as special cases of a general theory of the enterprise in geographical space, with many (but not all) of the considerations affecting MNE behavior being applicable also to multi-regional and multi-plant firms.

Conclusion

Despite the inevitable selectiveness of the above account, the substantial progress made in many aspects of corporate studies will be clear. Yet the dilemma, stated at the beginning of this paper, remains. While undoubtedly superior to partial generalizations for bringing order to the myriad individual case studies, theory so far is scarcely in a state to be really useful to business historians. The problem about economic theories in this area is not so much that they are over--rigid--although parts such as the 1933 models undoubtedly are-but that the various branches of these theories are poorly coordinated and lack an inner logical core.

At the same time, a holistic theory is likely to cover every possible category, is likely to be too general and as suggested at the outset could do violence to the richness of business history case studies. Perhaps the infinitely varied corporate experience would be better reflected in the careful building up of analysis that highlighted the essential similarities and differences of various types of firm, such as manufacturing, distribution, and non-profit making. It is to be hoped that the second century of research into industrial organization will be as fruitful as the first century has been.