“Firms and Industries in Evolutionary Economics: Lessons from Marshall, Young, Steindl and Penrose”

By Harry Bloch and John Finch
Firms and Industries in Evolutionary Economics: Lessons from Marshall, Young, Steindl and Penrose

Harry Bloch
School of Economics and Finance
Curtin University of Technology
GPO Box U1987
Perth, Australia 6845
harry.bloch@cbs.curtin.edu.au

and

John Finch
Department of Marketing
University of Strathclyde Business School
Stenhouse Building
173 Cathedral Street
Glasgow G4 ORQ
UK
john.finch@strath.ac.uk
Abstract

Evolutionary economists have tended to assess firms and industries separately, neglecting the role of their interaction in the process of economic growth and development. We trace the separation of firms and industries to Marshall, whose industrial analysis by means of the representative firm formalizes population thinking as “thin” means of relating firms and industries. Penrose avoids the industry concept by focussing on heterogeneous firms, while Young and Steindl develop mundane explanations of firms’ relations within groups, locating the impetus for growth in a poorly understood environment. We conclude that evolutionary economics should revisit firms’ boundaries, not in the sense of explaining the existence of firms, but in a relating and communicating sense in which boundaries signify selective means of relations with others.

Key words: firms, industries, Marshallian economics, external economies, firms’ boundaries

JEL classifications: B25, L14, 012

*Financial support from the Australian Research Council is gratefully acknowledged. An earlier version of this paper was presented at the 18th HETSA conference held at the University of Western Australia in July 2005. The authors appreciate helpful comments from the audience as well as from Stan Metcalfe, but take full responsibility for any remaining errors or omissions.
1. Introduction

Metcalfe (2007a, p. 2) argues that while Marshall ‘is routinely presented as a founder of modern neoclassical economics’, he should properly be considered ‘as a major figure in a thread of evolutionary reasoning that explores the restless, dynamic nature of modern capitalism, a thread that begins with Adam Smith and leads on through Marx to Schumpeter and Hayek’. In particular, Metcalfe points to the *Principles of Economics* (Marshall, 1920), and *Industry and Trade* (Marshall, 1921) as providing an analysis in which there is clear focus on the self-transforming as well as self-organizing nature of capitalism. We concur with Metcalfe’s assessment and argue that modern evolutionary economics would benefit from incorporating elements of Marshall’s analysis, along with elements from the analysis of some later economists who extended or criticized his method.

The particular aspect of Marshall’s analysis on which we focus here is the relationship between firms and the industry in which they operate. Marshall understands that heterogeneity among firms can lead through internal economies of scale to domination of an industry by a firm that acquires an early advantage. However, he then argues that family-owned firms have built-in life spans and that external economies can explain how the division of labour in the economy continues to expand nonetheless, thereby boosting productivity. Spillovers of knowledge in the environment of the “industrial college” allow an industry to progress while individual firms go through their limited life cycles.

The mix of internal and external economies and, particularly, the linkage between firm and industry created through the analytical device of the representative firm has been problematic for Marshall’s followers. The relationship is simplified
and, arguably, emasculated by Pigou (1920), who changes from the representative firm to the equilibrium firm, and then by Robinson (1931), who replaces the equilibrium firm with the optimum firm. In the process, Marshall’s concept of an industry as a forest composed of growing and declining firms thereby loses its self-ordering and self-transforming nature and is replaced by a mature lifeless plantation of uniform and stationary members.

We seek lessons for an evolutionary explanation of industries and firms by examining the analysis of Marshall and selected later economists who reject the emasculated form of the representative firm. Our discussion compares and contrasts the approach of Marshall to three successors who deal with the theory of the growth of the firm and its implications for the evolution of industries: Young (1928), Penrose (1956, 1959, 1960), and Steindl (1945a, 1945b, 1952). These authors each take issue with some aspects of Marshall’s approach, but in the process provide important pointers that we argue can be used to extend and deepen Marshall’s original insights.

Firms and industries have been central to research in modern evolutionary economics, at least since the modelling by Nelson and Winter (1982). In Nelson and Winter competition is a dynamic process in which firms choose between innovation and imitation strategies with stochastic outcomes. More varied and complex networks for knowledge transfer and co-evolution have been added, involving many more participants, including governments, universities and scientific organizations of various types, (Malerba, 2006). However, there is a certain tendency to deal with industries (as in Malerba 2006) or firms (as in Hodgson and Knudsen 2004, 2006) in isolation or, at least, without allowing full scope for their interaction in the process of economic development. Evolutionary economists can learn important lessons from Marshall, Young, Steindl and Penrose that will aid in the endeavour to bring firms and
industries together into the analysis of the self-ordering and self-transforming nature of capitalism.

2. Marshall’s Vision of Firms and Industries

In this section we focus on three related aspects of Marshall’s analysis of economic change: (1) the role of the firm and the industry in economic development, (2) the tension between internal and external economies, and (3) the dimensions of ongoing corporations and their internal accumulation of capital and knowledge. We then discuss Marshall’s use of the representative firm in his analysis of value and its subsequent re-interpretation in the neoclassical analysis of firms and markets in equilibrium.

The overarching argument in the *Principles of Economics* (Marshall, 1920) is that the basis of economic development is found in firms in relation to other firms, rather than in firms considered in isolation. Knowledge diffuses through localized industries, which act as industrial colleges with knowledge of products and productive techniques being ‘in the air’ (Marshall, 1920, p. 217). Knowledge leaks out of its contexts and is communicable through personal contacts among rivals, suppliers, training institutes, trade journals, customers and so on. While the communication of productive knowledge cannot be taken for granted, channels of communication are developed and firms locate in proximity to their rivals to participate and benefit in that industry’s activities. From each firm’s perspective, capital includes both internal organization and external connections (Marshall, 1920, p. 377).
Marshall describes innovation as being gradual, and though a brief period of accumulating knowledge internally might allow one firm to get a start on others, the high likelihood of leakage means that internal accumulation is just enough to provide relatively weak incentives for innovation on a piecemeal basis. Analytically, innovative capability (understood as an indirect or higher-order and cumulative ability) resides to a significant extent in an industry at large (with the broader industrial college among its functions), with limited chance for firms to develop their individual innovative capability. So firms build on others’ contributions with as much likelihood as building on their own. Innovations might as well be undertaken through random draws, as in the simulation of economic development in industries by Nelson and Winter (1982).

Analytically, Marshall’s explanation of economic development needs industries as well as firms, and the two need to be understood as distinct and interacting units of analysis. Though distinct, firms are not easily isolated as they are identified in part through sets of external connections, including gaining access to external economies (or economies specific to that localized industry), which may belong to no-one in particular, or may emanate from publicly or communally-funded research and teaching, such as with universities. Firms are irreducibly heterogeneous, with heterogeneity again including firms’ sets of external connections, but heterogeneity is analytically within the industry’s bounds so presenting the possibility of systemic coherence (Potts, 2001). To re-iterate, firms’ bounds are both relative to one another and are secured in principle through the domination of internal economies by external economies.

Marshall’s firms are heterogeneous at the very least because each has a unique and networked position relative to other firms. Co-incidently, a unique position is
reflected in what we now term a firm’s (internal) resources and capabilities, which
develop as firms cope with their unique settings (Richardson, 1972, 1975; Loasby,
1998). Analytically, heterogeneity can be translated into variation among firms, with
the trajectory of such an argument being towards organizing firms into industry
populations. Hence, variation may be interpreted statistically.

Bounds – additional to those provided by the industry itself – on heterogeneity
are inferred in Marshall’s system from each firm’s presumed likely demise, which
reinforces the domination of internal economies by external economies (Marshall,
1920, p. 287). The demise of firms is connected with an intergenerational model in
which family firms form, flourish and decline. For instance, third-generation
managers are presumed not to have acquired the personal motivation, commitment
and vigour of the founding generation. They do not enthuse their employees with
similar drive, commitment and vigour, especially in adopting and adapting to
innovations that are abroad in their industry. This leads to malaise and failures to
acquire new knowledge and techniques or failure to build upon these if acquired
(Marshall, 1920, p. 299). Marshall’s argument is not solely motivational. He also
implies that firms develop with respect to a particular vintage of technology and then
find it difficult to adapt to the newer technologies upon which new entrants base their
production so there is a corresponding dimension of core rigidities or competence and
technology traps. In sum, bounds on heterogeneity make it easier to perform an
analytical translation of heterogeneity to variation and hence to population thinking of
firms in industries.

Marshall’s framework reflects his expository strategy of “careful ambiguity”,
such that large and small firms, which coexist in industries, can also be reflected in
his theory of firms and industries. Productivity improvements are also acquired easily
as leakages from outside the industry of focus and in turn leak to other firms. Otherwise, if they emanate only from the industry of focus, we face the awkward question of why most firms in this system absorb the external changes easily but no one firm initiates any innovating activities of a more radical nature. The means of transition are expected to be orderly so that many firms can adapt the changes from elsewhere without too much disruption to the industry.

Despite irreducible heterogeneity among firms in industries, competition among firms prevents one firm getting an insurmountable lead and benefiting from faster accumulation or internal economies.\(^1\) The representative firm can stand for the industry because analytically Marshall places limits on – and so constructs a concept of – variation across firms in the industry. Variation’s necessary limits are in part caused by the transmission of knowledge, in part by the lack of opportunities for internal economies, and in part by limits in intergenerational managerial vigour. Also, important is the imperfection of markets, especially capital markets, but including labour and product markets. This imperfection limits the size of the firm in real time, even when there are internal economies of scale.

The limits on variation give Marshall’s representative firm some potential grounding when he comes to consider the determinants of the supply price of an industry. Here he notes,

We shall have to analyse carefully the normal cost of producing the commodity, relatively to a given aggregate volume of production; and for this purpose we shall have to study the expenses of a representative producer for that aggregate volume. On the one hand we will not want to select that new producer just struggling into
business, who works under many disadvantages, and has to be content for a time with little or no profits, but who is satisfied with the fact that he establishing a connection and taking the first steps towards building up a successful business; nor on the other hand, shall we want to take a firm which by exceptionally long-sustained ability and good fortune has got together a vast business, and well-ordered workshops that give it superiority over almost all its rivals…. Thus, a representative firm is in a sense an average firm. (Marshall, 1920, pp. 264-5, italics in original)

The representative firm, which represents or at least stands in for the firms that are interacting in a particular industry, is a wonderful and infuriating instance of Marshall’s careful ambiguity. It hints at population thinking, by which we mean the representation of a set of heterogenous firms in an industry as if they could be isolated and ordered on the basis of some critical characteristics (Hannan and Freeman, 1977). At the same time, we know that this isolation and ordering is – to say the least – difficult to perform empirically because firms exist and are ongoing in significant part through their relationships with one another. In other words, firms continue to exist because they continue to co-exist. Notwithstanding the empirical difficulties, the representative firm attracts our attention because its representative quality derives from its centrality in the notional distribution, and mere mention of a distribution implies a translation of our understanding of the group of firms from heterogeneity to industry and then to population.²

If internal economies are substantial and managerial vigour is sustained, the process of accumulation together with down-sloping cost curves over normal zones of

²
production, even for one firm, leads to the break down of Marshall’s theoretical and systemic account of industrial activity and development.\(^3\) Internal economies come to dominate external economies. That is, firms with lower costs can devote resources to Schumpeterian process innovations, such as establishing professional research and development functions, as well as to innovations in products and physical capital (note that in the citation above, Marshall refers to ‘well-ordered workshops’). Firms with lower costs no longer receive random innovation draws. Rather, the draws are now biased in their favour (as in Steindl’s analysis of absolute concentration discussed below). Even if knowledge is still ‘in the air’, other members of the community now find it a little harder to make sense of, as the many small steps of connection start showing gaps, requiring bigger leaps among potential imitator-innovators, or leaps that require the devotion of more and more accumulated resources (Cantner and Pyka, 1998). An implication is that firms are less embedded in one another, though as a consequence we can anticipate that heterogeneity is more pronounced, as discussed by Penrose (1959). Recognizing firms as irreducibly heterogeneous presents a procedural way of isolating firms, rather than dealing with them through the analytical means of population.\(^4\)

Marshall’s view of firms, which provides essential grounding to his analysis, has strong empirical bases, though it may have become less valid as conditions changed during the late 19\(^{th}\) century and beyond. Arguably, Marshall’s empirical understanding of trends in scientific management in business is accurate but not fully reflected in his theoretical framing of firms and their interdependence on one another. Hence, ‘The head of a large business can reserve all his strength for the broadest and most fundamental problems of his trade’, while ‘the small employer has not the time if he has the ability’ (Marshall, 1920, p. 284). He sees firms in an industry as moving
towards being roughly similar in size and accumulative capacity, based on the domination of external over internal economies and supported by a perceived lack of sustained vigour in managerial capability. This leads to Marshall’s (1920, pp. 315-316) ‘trees in the forest’ metaphor, with the industry (forest) remaining seemingly unchanged as individual firms (trees) are born, grow, mature and die. The industry comes to the fore and establishes dynamic balance between progressive and declining firms, with today’s progressives inevitably becoming tomorrow’s decliners (Marshall, 1920, p. 287). So Marshall can reasonably impose his model of the representative firm as a caricature of the industrial structure and processes, which can be described as in static equilibrium for the whole (Sutton, 2000). Of course, as Sraffa (1926) notes, Marshall also requires an assumption of rising costs for the representative firm to ensure competitive equilibrium for his theoretical system.

Marshall (1920, pp. 302-304) devotes some pages to joint-stock companies and to the specialist task of management. In *Industry and Trade* (Marshall, 1921) he refers to firms with ongoing accumulation of capital and knowledge, supported by scientific managerial techniques (Whitaker, 1999, provides a stimulating discussion). Scientific management is one part of the story of the large firm and also of the growing firm, which, if admitted by Marshall into his theoretical system, would undermine his explanation of economic development based on industries. Marshall’s (1921, p. 315) most telling remark is where he considers that it might be feasible for a firm to take over large amounts of economic activity, but that no one firm had demonstrated sufficient vigour and longevity, or had access to sufficient capital, for such an undertaking. Once again, Marshall is forced into having to rely on a phenomenon associated with the family-owned firms of the 19th Century to deal with
the modern corporation that was already taking hold at the beginning of the 20th Century (Penrose, 1952, p. 805).

However, a second volume of Principles was not written and so the descriptive, informal and evolutionary or developmental insights in Industry and Trade have had precious little impact on economists’ later readings of the more formal analysis of Principles (Comin, 2000; Raffaelli, 2004). While ‘it all could be in Marshall’, we have arrived at no conclusions on what the ‘it’ is. Part of the problem is in applying Marshall’s insights to the subsequent changes in corporate form; changes that are supported by the co-evolution of scientific managerial techniques, stock markets and legal regulation of corporate governance; trends which he noticed.

Modern neoclassical analysis has restricted the notion of the representative firm to provide theoretical precision, although the precision and formalism is not in the sense of evolution. The subtlety (and ambiguity) of Marshall’s arguments concerning the balance of internal and external economies and the declining intergenerational vigour of management have been progressively usurped by increasingly precise concepts that dismiss the importance of heterogeneity among firms for the purpose of analysing industry behaviour. Pigou (1920, p. 790) replaces the representative firm with the analytical construct of an ‘equilibrium firm’, which is defined as a firm that is in equilibrium whenever the industry is in equilibrium.6

Robinson (1931, p.11) goes further and adopts the concept of an ‘optimum firm’, which is defined as ‘a firm operating at that scale at which in existing conditions of technique and organising ability it has the lowest average cost of production, when all those costs of production which must be covered in the long run are included.’ While Robinson clearly recognizes that his optimum firm may differ from Marshall’s concept of a representative firm, he argues ‘The optimum firm is
likely to result from the ordinary play of economic forces where the market is perfect and sufficient to maintain a large number of firms of optimum size.’ Robinson (1931, p.12) Technology then becomes the sole determinant of profit-maximizing firm size, at least as long as the market is sufficiently large to accommodate a reasonable number of firms and thereby prevent the emergence of market power. It is this conception of the representative firm that dominates modern textbook treatments of the theory of the firm and competition.\(^7\) It provides no basis for evolutionary analysis because there is no meaningful population (actualizations) and no meaningful set of connections among firms (knowledge structure).\(^8\)

3. Allyn Young’s Reinterpretation of External Economies

The domination of internal economies by external economies is the main means of coherence empirically among firms and between firms and the industry as Marshall wrestles with the possibility of the differential and cumulative growth of some firms. Young (1928) develops the arguments of Smith and Marshall further in explaining how the industry itself could grow, thereby subtly altering the analytical focus from firms that are connected to one another, to the group of firms (Richardson, 1975, p. 352). For Young, external economies (beyond a particular industry) provide the direct impetus for economic growth, whereas for Marshall the connection was indirect, in preserving competition between firms in a particular industry as the means of growth. Following Young (1928, p. 528),
Although the internal economies of some firms producing, let us say, materials or appliances may figure as the external economies of other firms, not all the economies which are properly to be called external can be accounted for by adding up the internal economies of all the separate firms.

In Young’s sense, some economies “belong” initially and distinctly to firms, but seep out to others in an industrial setting by deliberate and knowledgeable imitation (White 2002).

Whereas with Marshall the processes of internal and external economies are intimately related as firms come to imitate one another as a normal practice, Young introduces something of schism between types of external economies by placing greater emphasis on the additional possibility of novelty and productivity gains from outside the industry of focus. Firms can gain internal economies through adapting to an increasing output, but are within circumscribed bounds. Internal economies can arise, for instance, where managers prioritize coordination while planning to adapt to higher levels of output. External economies are in a firm’s environment, implying that firms have little control or influence over these. In Young’s argument we know only a little of why there is an increase in output, even though plans are made to adjust to it. Hence, ‘out beyond, in that obscurer field from which derives its external economies, changes of another order are occurring. New products are appearing, firms are assuming new tasks, and new industries are coming into being’ (Young 1928, p. 528). Young is much clearer than Marshall in referring to boundaries that distinguish the industry and its environment.
Young describes the capture by firms of external economies in prosaic terms, implying similar rather than different and uneven understandings among firms. The most clearly understood part of economic development for Young, drawing from Marshall and also Smith, is also the part that features incremental changes, as firms adapt to production at a larger scale through adapting their internal managerial procedures to new means of production. The adaptive processes can be uneven and disruptive but work themselves out in ways that can be compared among all surviving firms. For example, adaptation could be through firms in a different or counterpart industry producing higher-order industrial goods (fixed capital), with the capital goods being easily applicable and absorbable for many firms working at similar levels of productive and absorptive capacity. In short, and from each firm’s perspective, changes in scale are established somewhere else and most firms adapt quite easily. A Marshallian twist is that opportunities arise for new entrants to replace older firms that lack the motivation and vigour to adapt.

Young’s reinterpretation of external economies, now external to the industry, provides a “cleaner” means of identifying firms individually because an important source of external economies is from ‘that obscurer field’; the environment, which is necessarily poorly understood. Further, firms establish their identities within a competitive group rather than an industry. “Competitive group” implies clear-cut connections between firms through strategic interaction, rather than the complex co-creation of external economies. If the now mainly competitive relationships between firms in a competitive group can be interpreted as orthogonal to the development of the firms or the group, the translation of heterogeneous firms into a population of firms rather than into the more complex concept of the industry may be feasible.
4. Steindl’s External Limits to the Growth of Firms

In his discussion of the determinants of firm size, Steindl (1945b, p. 3) starts by criticizing Marshall’s use of the representative firm. He notes that according to Marshall, ‘The limitation to the size of the representative firm, to sum up, is due to the limits of the market and, is obviously assumed, to the large scale economies becoming less important from a certain size on.’ Steindl (1945b, p.10) rejects this admittedly overly simplistic interpretation of Marshall’s view of the determinants of firm size, instead arguing that small and large firms coexist with a ‘general advantage of the bigger firm.’ (italics in the original).

Steindl then addresses the co-existence of near-monopoly capital alongside entrepreneurial and small businesses. In effect, Steindl is dealing with the consequences of Marshall’s inability to develop a theoretical framework that can accommodate large firms and growing firms. Steindl’s explanation is clearly connected to that of Young (1928). Steindl inherits Young’s reinterpretation of internal and external economies, but includes capital market imperfection, greater variety among firms in an industry codified as variation in asset size, and indivisibilities in innovations embodied in capital goods.

Steindl (1945a, 1945b, pp. 13-18) recognizes that entrepreneurs demand a risk premium on investments with uncertain returns to compensate for exposure to bankruptcy and for the risk associated with variance in returns. The risk premium
increases with the amount of finance required relative to the firm’s own capital.\textsuperscript{14} This means that the ability of firms to expand their productive capacity through the acquisition of additional capital equipment is realistically limited at any point in time. Further, small and large firms face different opportunities for undertaking risk-bearing activities that offer higher rates of return, because economies of scale tend to raise the return to large units of capital above that of small units of capital.

Recognition of differential opportunities for small and large business underpins Steindl’s subsequent examination of trends in concentration in \textit{Maturity and Stagnation in American Capitalism} (Steindl, 1952). Here, the higher returns associated with economies of scale are combined with technical progress that brings improvements in productivity. However, improvements that occur at an uneven pace across firms in the same industry yield differences in the level of production cost, even among firms in the same size class.\textsuperscript{15}

Firms with differing levels of production cost can coexist in the same industry due to imperfect competition, which leads to a general tendency towards price rigidity (Steindl, 1952, pp. 14-17).\textsuperscript{16} When prices are rigid, cost-reducing innovations lead in the first instance to an increase in the gross profit margins of the innovating firms. If the excess capacity of the firms with the lowest unit costs is within acceptable limits, these “progressive firms” have no incentive to cut prices. This allows high-cost firms, which are small or technologically backward, to survive, even when these “marginal firms” do not gain access to the cost-reducing technology.

Steindl argues that investment by firms is dominated by internal accumulation (the financing of investment by retained earnings).\textsuperscript{17} Higher profits earned by progressive firms therefore lead to the expansion of their productive capacity relative to marginal firms. Eventually, the progressive firms become the largest firms in the
industry. If the number of marginal producers is constant, the industry is subject to relative concentration through the faster rate of growth and growing market share for the limited number of largest firms. However, a sufficiently high rate of growth of industry demand may attract new entrants, as small and relatively high-cost firms, thereby postponing the onset of relative concentration (Steindl, 1952, pp. 40-42).

Steindl’s introduction of technological progress into the analysis loosens the finance constraint on firm growth. When technical progress raises the profits of progressive firms, the rates of internal accumulation and growth for these firms also increase. Given a predetermined rate of growth for industry demand, unplanned excess capacity eventually emerges as a result of the enlargement of capacity.  

Initially, progressive firms react to this unplanned excess capacity by engaging in aggressive price or selling competition. Marginal firms cannot match the aggressive competition due to their smaller gross profit margins, so they are forced to cede market share to the progressive firms. Some of them will go bankrupt and exit the industry. The reduced gross profit margins also dissuade entry of new firms into the industry. Concentration of the industry rises in absolute terms in the sense that, with the decline in the number and size of the marginal firms, there is a decline in the total sales of small firms and a rise in the total sales of large firms (Steindl, 1952, pp. 42-43).

Steindl’s progressive firms are able to overcome the external constraint on growth posed by market demand growth below their rate of internal accumulation through the use of aggressive competition. However, this aggressive competition reduces the profit margins for themselves as well as for competing firms, hence reducing the rate of internal accumulation and the rate of firm growth throughout the industry. Thus, the environment influences firm growth, but only in a way that is intermediated through the development process of firms and of the competition
between firms. This is most clearly indicated when Steindl suggests that the increasing industry concentration eventually leads to an abatement of capacity expansion by the progressive firms, even though profits are available for investment. Managers of the progressive firms come to recognize that there is little possibility of capturing enough market share from the remaining competing firms to maintain growth in sales (Steindl, 1952, pp. 53-55).

Steindl uses the division of firms into the two generic types of marginal and progressive to analyse competition as a process, as “an ideal pattern of competition”. In this pattern, the internal constraint on progressive firm growth of limited finance is sequentially replaced by the constraint of the presence of marginal firms in the market and then by the limited growth of the overall market. Only at this last step of the process, which Steindl terms “maturity”, is a type of equilibrium obtained (Steindl, 1952, p.60). Thus, heterogeneity of firms drives the analysis in a manner that would be impossible with the representative firm as an ideal type, or even with the broader interpretation of Marshall’s notion of the representative firm as an average type.

5. Penrose on Internal Limits to the Growth of Firms

Penrose (1959, p. 2) is clear in her rejection of the idea that there is a firm size that is somehow best: ‘It is often presumed that there is a “most profitable” size of firm and that no further explanation than the search for profit is needed of how and why firms reach that size. Such an explanation of the size of firms will be rejected in this study.’ She goes on to state that, ‘it will be argued that size is but a by-product of the process of growth, that there is no “optimum”, or even most profitable, size of firm. As we
shall see, traditional theory has always had trouble with the limits to the size of firms, and I think we shall find the source of the trouble.'

Penrose shares with Steindl the notion that it is the growth of firms, rather than their size, which is limited under capitalism. However, where Steindl argues that firm growth is constrained by an external factor of limited access to capital, Penrose (1959) locates the constraint on growth within the firm. She argues that investment exhibits path dependence, with the increments to firm’s resources being specific to its unique growth trajectory. In other words, additional resources are of value as productive services only in a complementary and intangible connection with the firm’s established productive services drawn from its prior resources.

Initially, the freed-up resources are most useful in expanding its current production. However, as the firm grows and develops, attractive opportunities for the employment of these productive services may be in areas outside the firm’s current area of specialization, defined in terms of either technologies or markets (Penrose, 1959, pp. 109-111). Thus, in Penrose’s analysis even the size of the market for the firm’s original products poses no limit to firm growth as diversification is a likely outcome to the continued internal development of firms.

The potential for diversification complicates the definition of the industry. Penrose makes no clear distinction between internal and external economies, probably because to her “external” is everything that is beyond the firm and its immediate network of suppliers and buyers. Indeed, because the firm has fuzzy boundaries, and a significant, though still fuzzy, boundary internally between the senior management team and the firm’s operations, it is difficult to establish meanings of “internal and external”. Instead, Penrose (1959, pp, 99-102) acknowledges as ‘economies of growth’ the possibility of freely available productivity improvements, which can
emerge organically as by-products of the firm’s growth. Penrose’s economies of
growth, which are not divided into internal and external components, fulfil a
comparable function and also share some of the characteristics of Marshall’s external
economies because they have to be acquired and integrated into an overall managerial
or entrepreneurial vision. Indeed, Loasby (1999) argues that Penrose is thoroughly
Marshallian without really referring to Marshall.

Penrose’s analysis of the development of productive knowledge is carried out
primarily within firms, while acknowledging some role for resources that are close at
hand to firms either in supply, in competitive relations or among product users. This is
the closest that Penrose gets to industries. Even this is more focused on connections
of exchange rather than on interactions between firms that share productive
techniques or other criteria for identifying industries. In fact, Penrose goes much
further than Chamberlin (1933) in freeing her analysis, and so her heterogeneous
firms, from the conflation of firms and industry in Marshall’s representative firm. Her
heterogeneous firms are not even anchored to particular markets or industries.

Penrose has a generic firm type, replacing the “representative firm” of
neoclassical analysis with a “growing firm” in order to explain processes that are
presumed to be common to all firms that are pursuing growth, but not determining in
any particular case. Importantly, it is a distinct subset of special firms that pursue
growth, so that Penrose’s generic firm type diverges sharply from Marshall’s notion
of the representative firm as an average firm.

All firms that seek to grow have the same general processes, but emerge from
and are shaped by idiosyncratic historical paths. Firms develop different resources
(which become so in connection with an entrepreneurial plan developed within the
senior managerial team) and managers have different capabilities and outlooks; these
differences affect how the firms carry out their productive activities. Hence, free productive resources, which Marshall termed external economies, are still ‘in the air’, but there is a very limited group of individuals, including managers of the particular focal firm and perhaps other firms located in close proximity, that are in a position to devote other resources to interpreting and assimilating these, or which have opportunities of interpreting and assimilating these. Most of Marshall’s firms could absorb external, and more or less free, improvements to productivity quite easily. Penrose’s growing firms need managerial competence (which cannot be hired at short notice and with immediate effectiveness) and other resources to capture and direct the economies of growth. Further, the necessary incumbent managerial competence is available episodically as ongoing activities (selected in previous episodes) become routine, freeing up resources, particularly managerial resources.

For Penrose (1959) the factor that limits the internal development and growth of the firm ultimately is the capacity of its management. She argues that, ‘Since the services from “inherited” managerial resources control the amount of new managerial resources that can be absorbed, they create a fundamental and inescapable limit to the amount of expansion that a firm can undertake at any time.’ (Penrose, 1959, p.48) Thus, managerial resources too can develop over time. Penrose views firms as repositories of resources, which include fixed capital alongside intangible resources, such as knowledge and routines, from among which managers can select and configure plans in the form of combinations of productive services. This neatly captures Penrose’s (1952) objections to biological analogies in economics, a cautionary point for modern evolutionary economics.

Penrose (1959, pp. 107-108) pays attention to the relationship between firm, market and industry, and her discussion recognizes concepts such as diversification.
and product to be inherently relative or ambiguous. All growing firms have unique bundles of resources from which managers configure sets of productive services (ibid., p. 77). Firms sell in markets by undertaking spending efforts that seek to build customer loyalty, but Penrose emphasizes that this does not mean a firm’s resources are tied to a particular market (ibid., pp. 116-118). Penrose is so successful in isolating firms from one another, and making the growth of firms dependent on accumulated capacities aligned with the entrepreneurial vision of its managerial team, that it becomes very difficult to translate heterogeneity into an industry, let alone a population.20

6. Lessons for Today’s Evolutionary Economics

What are the lessons for modern evolutionary economics from the efforts of Marshall, Young, Steindl and Penrose? We start with a general observation about the nature of work of all four and then point to lessons regarding the treatment of firms and industries from each of the authors in turn.

The general observation that links the theorizing of each of our authors is that their analyses are conducted in a specific empirical context. This distinguishes their work from that of most mainstream theorists after Marshall. Compare the rich mixture of empirical observation and theory in Marshall’s Principles to the high-level abstraction of subsequent theoretical tracts in mainstream economics, for example Hick’s (1946) Value and Capital, Samuelson’s (1947) Foundations of Economic Analysis or Varian’s (1992) Microeconomic Analysis. Young, Steindl and Penrose follow the Marshallian approach to theorizing, rather than participating in the
mainstream pursuit of a universal economic theory devoid of any historical or institutional specificity. Given that our authors all give due heed to historical and institutional context that is central to evolutionary economics, they belong to a group of what might be termed, empirical evolutionary economists, that goes back to Adam Smith and before.

Lesson one comes from Marshall (1920 and 1921) and tells us to treat firms and industries as interdependent rather than focus on one and either ignore the other or treat it as derivative (as Nelson and Winter, 1982, do in simulating changes in industry concentration from stochastic innovation outcomes of individual firms). Marshall has each firm identified in part through its unique set of relations with other entities, including other firms as well as organizations such as universities and industrial journalism. Relationships, or external connections, form part of each firm’s organizational capital and are means to new knowledge and understanding of industrial arts. Further, some external economies seem to belong to no one in particular, and rather are as if ‘in the air’. While firms can of course be identified separately, they can owe their continuation to informal networks of relations with other entities, which also grants each firm a unique perspective on the network and so an irreducible basis of heterogeneity (Loasby, 2001, p. 408).

This is a version of Granovetter’s (1985) ‘problem of embeddedness’, and of Uzzi’s (1997) empirically grounded ‘paradox of embeddedness’, in which firms cannot escape their networks of social and economic connections. Analytically, one way out, which might be feasible, is to recast embeddedness as nested selection, such that firms make selections of connections (that is choose or at least modify their own network position), and are then selected in part on the basis of network selections. In other words, industries are understood as entities that emerge from the network
properties and interactions among a set of firms. However, Marshall’s firms are both partly constitutive of one another and also of the industry so are redefined rather than eliminated through emergence. The challenge for today is to further develop the analysis of the co-evolution of industries and their constituent firms.

Lesson two comes from Young (1928) and tells us to look at the industry (in its relation to individual firms) and beyond to its connections with the rest of the economy for the driving mechanism of economic progress. This may be taken as a corollary to Marshall’s lesson on interdependence of firms and their industry. Young gives the industry a life independent of that of its constituent firms. This is an approach that has resonated with many evolutionary economists. For example, Malerba (2006, p. 18) refers to Young as providing a first discussion at a theoretical level of the co-evolution of vertically related industries due to the interdependence of the extent of the division of labour.

Lesson three is from Steindl’s analysis and tells us to incorporate external constraints on firms into the analysis of dynamics of firms and industries. One constraint, the limited access that firms have to debt or equity finance for the expansion of productive capacity, comes from the broader economy but becomes industry specific through the mechanism of retained earnings as a source of internal finance. The second external constraint of limited demand for the industry’s products interacts with the finance constraint through Steindl’s ideal pattern of competition to provide for the co-evolution of firms and the industry, leading up a mature industry in which firms behave differently towards investment and pricing than during the ideal pattern of competition.

Steindl’s account provides an example that falls into the evolutionary economics category of subset selection (Hodgson and Knudsen, 2006). Firms are
selected on the basis of differential cost through firms with lower costs having faster
growth rates and being able to survive aggressive price competition during the ideal
pattern of competition. The reason for the differences in costs is not clearly explained
by Steindl. If the differences are due to a trait that is distributed over the group of
firms, such as willingness (as discussed by Marshall and Young) to adopt and adapt to
new vintages of capital produced elsewhere, generative selection can replace subset
selection as variation in the trait occurs due the interaction of firms with their
environment (the industry and beyond). Thus, Steindl’s analysis offers a base on
which to fruitfully apply evolutionary theorizing in a context that includes the
interaction of firms and industries.

Lesson four comes from Penrose and tells us to consider the constraints on
firm growth arising from within the firm, particularly limits to managerial capability.
Importantly the internal constraints are associated with the imperfect knowledge
within the firm as an organization. Means of organizing improve within the firm over
time, allowing the firm to extend the scale and scope of its enterprise without adding
to managerial resources. Knowledge flows are at the core of much evolutionary
theorizing (Loasby, 1999 and 2002). Furthermore, the conception of the firm as
having an internal structure that evolves through time has been taken up in formal
evolutionary theorizing (Hodgson and Knudsen 2006). Penrose only deals vaguely
with the industry concept, as she allows firms to transform their product and
marketing orientation to escape markets that might otherwise constrain their growth.
However, there is an implicit lesson that industries need to be construed broadly and,
indeed, allowed to co-evolve with constituent firms.
7. Conclusions

We proceed on the basis that evolutionary economists are not fully aware of the rich vein of economic analysis from Marshall, Young, Steindl and Penrose that deals with firms and industries empirically in an evolutionary spirit. We aim to draw attention to this work and to tease out some of its salient features that are particularly relevant to current developments in evolutionary theorizing. We hope that this encourages more researchers to treat firms and industries as interdependent and co-evolving.

Our concern in writing this paper is to plot the means by which evolutionary researchers have kept firms and industries as distinct foci, especially where pursuing mainly empirical research projects. While industries comprise firms, they are not aggregations of firms because firms are a combination, even compromise, of internalizing and externalizing tendencies. In short, industries comprise firms not because firms share common technology, managerial techniques or potential customers, but because these shared dimensions emerge through significant interactions and communications between firms, whether collaborative or competitive.

Our paper shows that researchers, even those who undertake research close to firms and in the spirit of Marshall’s evolutionary framing, deploy considerable discretion in stabilizing the relationship between firms and industries, thereby isolating one from the other. Firms are irreducibly heterogeneous because they have idiosyncratic histories and experiences, extending to their personnel. Heterogeneity establishes notable and stimulating differences among firms, which is a basis for there
being an industry. A dominant analytical strategy for coping with difference is to convert heterogeneity to variety and then variation, which is found in Marshall’s construct of the representative firm, and in population thinking more generally. Marshall also employs the auxiliary or reinforcing assumptions of ‘the death of firms’ and the ‘localizing of firms’ in order to place meaningful bounds on variation. The theorists discussed in this paper have nascent systems theories and place the analytical distinction of (firm or industrial) system and its environment differently. For instance, Penrose focuses on heterogeneous firms and so considers relations among firms, which may form loose groupings or associations temporarily, informally. Young and Steindl group firms into industries and characterize their relationships in mundane terms, compared with activities that are of greater significance to the group which are presumed to be occurring elsewhere, in the poorly understood environment.

The above are largely negative conclusions in that they account for how researchers have used analytical devices to loosen the connection of firms and industries. In a positive sense, our paper suggests that evolutionary research should be redirected fundamentally at the question of firms’ boundaries. Given evolutionary interests in development and growth, firms’ boundaries lose their deterministic and foundational quality, which signify the existence of firms. Rather, corporate boundaries are recast as means of communication and interaction. The question facing firms is not so much where our boundaries are, but rather with whom do we share boundaries and how are the boundaries shared? This leads to the question of how does sharing of boundaries change in the co-evolution of firms and industries.
Notes

1 Firms focus their competitive efforts on selling or marketing their products to end users. Heterogeneity in firms’ productive capabilities can be reflected further in heterogeneous products, although of course there are many ways of making comparable products. Consumers also cope with heterogeneous products in seeking to compare products.

2 In attempting to translate heterogeneous firms in an industry empirically into a population of firms, we know that the ordering should be in the basis of average costs. In which case, the representative firm is something of a lagging indicator, suitable also for static or equilibrium analysis. If Marshall’s emphasis were on growing firms ahead of the developing industry, surely more attention would be focussed on both the new entering firms and also on the exceptional long-lived firms that are enjoying internal economies, as indicated in the passage from Marshall, which we cite above. Metcalfe (2007b) provides an enlightening analysis of how the representative firm may be modelled to evolve in a manner capturing both the lagging and leading aspects of the concept.

3 Marshall recognises that large firms can come to dominate industries: ‘The advantages which a large business has over a small one are conspicuous in manufacturing …. But there is a strong tendency for large establishments to drive out small ones in many other industries’ (Marshall, 1920, p. 297).

4 Once the firm is treated in isolation it makes no sense to speak of the competitiveness of the firm, for competitiveness is a relational concept. This point
that is often lost in business school discussions developed from Penrose’s analysis of the growing firm. The recent exchange between Rugman and Verbeke (2002) and Lockett and Thompson (2004) is instructive in this respect. Rugman and Verbeke argue that Penrose provides little basis for ‘isolating mechanisms’, which they argue is integral to the resource-based view. Locket and Thompson counter that Penrose’s approach is compatible with managers seeking to create and protect rents through isolating mechanisms.

5 Raffaelli (2004, p. 210) argues that Marshall had a coherent and general theory of development or evolution, encompassing deliberative innovation and systemic selection and reproduction among innovations, embodied in repeatable or autonomous routines that exhibit tendencies of inertia.

6 Pigou (1920, p. 790) acknowledges in a footnote that Marshall had in mind something more, such as a typical firm or firm of average size. However he argues that such complications are irrelevant to his purpose of analysing the industry supply curve using the equilibrium firm to stand for all firms.

7 There are alternative neoclassical theories of the firm, particularly the theory due to Coase (1937), Williamson (1975) and Simon (1976) that relaxes the assumption of perfect knowledge and relates the boundaries of firm activity to transaction costs. However, this simply adds further external conditions, the degree of imperfect information and the costs of the technology for overcoming it, to the exogenous determinants of firm size.

8 The modern representative firm or agent is a travesty since it relates to uniform agency, which is to say it is representative of itself – most odd.

9 “Higher-order” is used in the sense of Menger (1976, p. 56): ‘a large number of other things in our economy that cannot be put in any direct causal connection with
the satisfaction of our needs, but which posses goods-character no less certainly than goods of first order.’

10 Metcalfe et al. (2006) draw on Young’s understanding of interdependence among industries in their emergent modelling of economic growth.

11 The competitive group re-appears as a unit of analysis in Caves and Porter (1979) and influences strategy research thereafter.

12 Marshall does this too, but in a descriptive rather than systemic sense.

13 If innovations are embodied in capital goods, adoption may result in indivisibilities either in the form of minimum output scale for the equipment or the requirement that a whole production facility be redesigned to accommodate the new machines.

14 Steindl’s argument concerning the relationship between the variance of return and the risk premium closely follows Kalecki’s principle of increasing risk (Kalecki, 1937).

15 Steindl follows the standard practice of statistical agencies, particularly the US Census Bureau, and defines industries in terms of common or overlapping production technology. This definition is appropriate in terms of Steindl’s focus on scale and adoption of best-practice technology as sources of cost advantage for firms in an industry, but ignores potential competition between products with competing uses that are produced using different production technologies.

16 This view of price rigidity as endemic with imperfect competition follows the arguments of Berle and Means (1932) on the prevalence of administered prices in big business.

17 Internal accumulation also features prominently in the post-Keynesian analyses of Eichner (1976) and Harcourt and Kenyon (1976), in which dominant firms set their profit margins to generate sufficient retained profits to carry out their desired
investment in the expansion of productive capacity. However, causality differs from that in Steindl, where investment increases to the level set by the interaction of a rigid price with falling unit cost. In the post-Keynesian analyses, it is price that increases to the level required to finance desired investment expenditures, given the firm’s unit cost and demand for its product.

18 This restrictive aspect of Steindl’s analysis is relaxed by Levine (1981) and Shapiro (1986), who consider innovation in terms of new product development as a means to overcoming a given growth rate of market demand for established products. Bloch (2006) considers the implications of new product development for Steindl’s analysis of the pattern of competition between progressive and marginal firms.

19 See Sraffa (1926) for a discussion of alternative criteria for defining an industry.

20 By contrast, Richardson (1975) in his essay on Adam Smith examines activities and capabilities in the context of industries. He argues that economies of scale affect industries but possibly with differential effects on clusters of activities and capabilities within firm. Hence, growth will encourage firms to specialize. There may be tendencies towards concentration, but Richardson’s adopts an evolutionary perspective such that firm’s experiments in specializing continually interrupt these tendencies. Langlois (1992) develops this perspective in the context of the firm and its boundaries.
References


Loasby, B.J. (1999) The significance of Penrose’s theory for the development of
economics, *Contributions to Political Economy*, 18, 31-45.

Loasby, B.J. (2001) Time, knowledge and evolutionary dynamics: Why connections
matter’, *Journal of Evolutionary Economics*, 11, 393-412.

Loasby, B.J. (2002) The evolution of knowledge: Beyond the biological model,

Lockett, A. and Thompson, S. (2004) Edith Penrose’s contributions to the resource-
based view: An alternative perspective, *Journal of Management Studies*, 41,
193-203.


Evolutionary Economics*, 16, 3-23

Bert F. Hoselitz with an introduction by F.A. Hayek (New York and London,
New York University Press).


