Introduction

Imagine, perhaps implausibly, that a present-day practitioner of what the economics profession calls ‘Industrial Organization’ had somehow stumbled upon Chapters VIII through XIII of Book IV of Marshall’s *Principles*, the chapters explicitly bearing the term ‘industrial organization’ in their titles. Our accidental economic tourist would find essentially no familiar landmarks. There would be no references to strategic interaction let alone to game theory. No talk of principal–agent problems or misaligned incentives. Instead, the bewildered industrial organization’s economist would discover a disquisition on industrial production understood as a form of evolving social organization and as a manifestation of the process of economic growth, with considerable emphasis on the division of labour and on the evolution of knowledge, technology and organizational form.

Does this mean that Marshall does not, or cannot, speak to the present day? To the contrary, I will argue that Marshall has much to say to practitioners of what is broadly Industrial Economics, a term still somewhat in use – mostly outside North America – to denote a broader interest in how industry functions and is organized. Indeed, it is precisely because of the narrowness – cynics might say sterility – of the mainstream research program in industrial organization that genuinely Marshallian ideas are enjoying resurgence in many areas of present-day thought. I will begin by focusing on one sub-area of Industrial Economics, the Economics of Organization, and then I will turn to what we may call Industrial Geography broadly understood.

Organization and capabilities

As Loasby (1976), Moss (1984), and others have argued, what we think of as mainstream ‘Marshallian’ theory today is in many ways more Pigovian than it is Marshallian. Rather than thinking in population terms as Marshall did and constructing a ‘representative firm’ that reflects the characteristics of the population of firms as a whole (rather than the characteristics of any particular firm), the neoclassical theory of the firm since Pigou begins with identical idealized firms and then builds up to the industry by simple addition. The ‘theory of the
Marshall’s (real) influence  309

firm’ in modern-day price theory builds on the Pigovian foundation. It begins with firms as production functions, each one identical and each one transforming homogeneous inputs into homogeneous outputs according to given technical ‘blueprints’ known to all. One effect of these assumptions has been to reduce the margins on which firms operate to two only: price and quantity. This in turn has led to the notion of ‘perfect’ competition, in which a technically desirable set of assumptions replaces the common-sense notion of competition (Hayek 1948; McNulty 1968).

Now, price theory – whether appreciative2 Marshallian or heavy-metal Pigovian – was never intended to be a theory of the firm as an organization or an institution. As Marshall understood, the firm in price theory is a theoretical link in the explanation of changes in price and quantity (supplied, demanded or traded) in response to changes in exogenous factors (Langlois and Koppl 1991). It was never intended to explain industrial structure, let alone to serve as a guide to industrial policy. More to the point, using this sort of price theory to explain the boundaries of the firm is just plain illogical because the firm’s boundaries in price theory are a matter of assumption.

Marshall’s understanding of the issues of organization was thus not at all ‘Marshallian’ in the sense that most present-day economists understand that term. For Marshall, firms, inter-firm relationships and markets are all modalities of organization that help coordinate differentiated knowledge and further the growth of knowledge. Marshall touches specifically on the firm in Chapter XII of Book IV of the Principles, where he is clear that firms arise in order to provide an organizational or coordination function. Middlemen can sometimes serve as a coordinating intermediary between the consumer and the suppliers of factors inputs. But in most cases, that task of intermediation must be

given into the hands of a specialized body of employers, or to use a more general term, of business men. They ‘adventure’ or ‘undertake’ its risks; they bring together the capital and the labour required for the work; they arrange or ‘engineer’ its general plan, and superintend its minor details.

(Marshall 1920: IV.xii.7)

As an ‘organizer of production’, the manager must have specialized knowledge of his or her field and a keen faculty of judgment, especially in choosing and overseeing workers and subordinates (ibid.: IV.xii.18–19).

The present-day field of the economics of organization traces to Ronald Coase’s famous 1937 paper, ‘The nature of the firm’. Coase’s approach is certainly Marshallian in its emphasis on effects on the margin. He wrote that ‘A firm’

will tend to expand until the costs of organising an extra transaction within the firm become equal to the costs of carrying out the same transaction by means of an exchange on the open market or the costs of organising in another firm.

(Coase 1937: 395)
The Post-Coasean research program focused on identifying the sources of (what came to be called) the transaction costs of market exchange. Curiously, little effort attached to the other blade of Coase’s Marshallian scissors: the costs of organizing within firms. This oversight is related, I believe, to the surge to prominence of Pigovian price theory at the time Coase was writing. Because the production function, with its attendant assumptions of perfect knowledge, was understood to describe the inner workings of firms as productive units, the neglected margin of transactions between firms was the only margin with which Coase had to work (Langlois and Foss 1999).

Nonetheless, Coase’s own account of why firms arise – of why explicit organization frequently interposes itself between consumers and factor owners – is not far removed from that of Marshall. In the end, a Coasean firm reduces transaction costs because it allows for managerial coordination in the face of uncertainty. Frank Knight’s story was along similar lines, echoing Marshall’s portrayal of management as a Smithian specialization in the (non-contractible) faculty of judgment (ibid.).

A re-quickening of interest in the economics of organization since the 1970s has brought a rediscovery of Coase’s 1937 paper and the questions it asked. For the most part, however, the present-day literature has chosen to answer these questions rather differently. Despite some occasional interest in coordination, especially early on, writers nowadays tend to focus almost exclusively on issues of incentive misalignment and ‘opportunism’ as sources of the costs of market transaction. Sometimes, this problem takes the form of moral hazard arising out of the difficulties of monitoring and measuring output; more often, it takes the form of a threat of rent expropriation when assets are highly specialized to a transaction (Williamson 1985). Ironically, perhaps, the asset-specificity literature is often given to announce Marshall as an inspiration and precursor (ibid.: 52). This is not, of course, because of Marshall’s own treatment of industrial organization, but rather because of his notion of the composite quasi-rent (Marshall 1920: VI.viii.35) – a quasi-rent created by the cooperation of two (or more) parties whose distribution among those parties cannot be determined on some margin but is necessarily a matter of bargaining. It is a composite quasi-rent over which the owners of highly specific assets tussle when they threaten to hold one another up. (Of course, Marshall did not view composite quasi-rents as a terrible problem, in part, because he saw ‘custom and . . . notions of fairness’ as helping to smooth the bargaining.)

Although the mainstream literature (especially that emanating from economics departments) continues to be absorbed with problems of incentives, there has arisen a dissident movement for whom Marshall’s own views on industrial organization are a genuine inspiration. This ‘capabilities’ approach – closely allied to what in business schools is often called the Resource-Based View of the firm – takes seriously the Marshallian concern with coordination and with managerial knowledge and judgment (Langlois and Foss 1999). Indeed, Nicolai Foss goes so far as to refer the principal figures of this movement as Post-Marshallians (Foss 1998).
A central notion here is what George Richardson (1972) called *economic capabilities* or what Edith Penrose (1959) referred as *resources*. Richardson (1972: 888) describes capabilities as ‘the knowledge, experience, and skills’ of the firm. Taking issue with the representation of knowledge in the production function approach, he writes:

Of course I realise that production functions presume a certain level of managerial and material technology. The point is not that production is thus dependent on the state of the arts but that it has to be undertaken (as Mrs. Penrose has so very well explained) by organisations embodying specifically appropriate experience and skill. It is this circumstance that formal production theory tends to put out of focus, and justifiably, no doubt, given the character of the optimisation problem that it is designed to handle; nevertheless, it seems to me that we cannot hope to construct an adequate theory of industrial organization and in particular to answer our question about the division of labour between firm and market, unless the elements of organisation, knowledge, experience and skills are brought back to the foreground of our vision.

(Ibid.)

We might say that, in an important sense, to restore to consciousness, these issues of knowledge, experience, and skill are to bring Marshall back in.

How does the notion of capabilities lead to a theory of the boundaries of the firm? In Richardson’s system, production can be broken down into various stages or *activities*. Some activities are *similar*, in that they draw on the same general capabilities. Activities can also be *complementary*, in that they are connected in the chain of production and therefore need to be coordinated with one another. The central problem of economic coordination lies in the fact that what is complementary need not be similar:

Where activities are both similar and complementary they could be co-ordinated by direction within an individual business. Generally, however, this would not be the case and the activities to be co-ordinated, being dissimilar, would be the responsibility of different firms. Co-ordination would then have to be brought about either through cooperation, firms agreeing to match their plans *ex ante*, or through the processes of adjustment set in train by the market mechanism.

(Ibid.: 895)

Note, with Brian Loasby (1991), that Richardson here stands on its head a principal presumption of the mainstream economics of organization, namely that contractual relationships among firms must be fraught with hazards and thus that integration must be widely desirable. For Richardson, the dissimilarity among capabilities makes integration costly, to such an extent that, even when there are transaction costs, market transactions may typically prove a cheaper alternative.
Another significant Marshallian feature of Richardson’s approach is that it returns attention to the blurriness of the idea of a firm’s ‘boundaries’ and to the importance of the wide array of contractual and ownership arrangements that coordinate production in a modern economy.

The evolution of industrial structure

The question of the boundaries of the firm, even including the issues of inter-firm relations and ‘hybrid’ forms to which Richardson pointed, is really only one aspect of Marshall’s version of industrial organization. In a conceptual way in the Principles, and in a more historical and empirical way in Industry and Trade, Marshall was concerned with the dynamics of organizational form, which he understood in strong analogy to the evolution of biological structure.

One manifestation of Marshall’s concern with knowledge and the evolution of industrial form is his famous discussion of industrial districts, often cited as inspiration for (or at least as ancestry to) the so-called New Economic Geography (NEG). But Paul Krugman, perhaps the central figure in this literature, is quick to concede that the NEG does not capture Marshall’s full vision. ‘In modern terminology’, Krugman says that Marshall argued that industrial districts arise because of (i) knowledge spillovers (‘the mysteries of the trade become no mysteries; but are as it were, in the air’), (ii) the advantages of thick markets for specialised skills, and (iii) the backward and forward linkages associated with large local markets. The new economic geography in effect considers only the third, one that is arguably less important in practice – but easier to formalise – than the others.

(Fujita and Krugman 2004: 153)

Masahisa Fujita, another founder of the NEG, goes further to suggest that only an (as-yet-nonexistent) dynamic theory of knowledge spillovers will begin to capture what Marshall was getting at (ibid.: 160).

Again, however, there are some less-mainstream writers who have already tried to follow Marshall in dynamic, and sometimes even evolutionary, directions. In the case of the firm, as we saw, Edith Penrose’s (1959) dynamic account of organizational growth has been highly influential on the Resource-Based View of the firm in the management and strategy literature. In her theory, firms consist of acquired pools of resources – including, importantly, managerial resources – that come in lumpy bundles. In order to take advantage of excess capacity in some of the lumps, the firm may expand or diversify into areas in which that capacity is useful. This in turn may lead the firm to acquire other complementary capabilities, which will lead to further excess capacity, and so on. Penrose’s ideas found their empirical correlative in Alfred Chandler’s (1977) account of the rise to dominance of the large managerial corporation in the late nineteenth and early twentieth centuries, a rise that seemed to follow the pattern of diversification and internal growth in capabilities Penrose had outlined.
But, as Marshall understood, the firm – even the large firm – is imbedded in a wider system of capabilities residing in the industry, the economy and the society. A number of authors have championed what (especially in Scandinavia) is called industrial dynamics (Carlsson 1989). Grounded in Marshall but leavened with Schumpeter, this approach tries to look at the growth of capabilities not only in firms but also in the wider industry. Relatedly, some business historians – enjoying the benefit of several additional decades of history – have called into question, both theoretically and empirically, the triumph of the large managerial corporation and have sought to imbed the Penrose–Chandler story in a roomier framework that admits of the decline and the growth of the large multi-unit enterprise (Langlois 2004). At the end of the twentieth century, Chandler’s powerful Redwoods were supplanted by a stand of younger – and less vertically integrated – trees of the forest.

Another related strand of thought has sought to widen the horizon beyond even industry and market to include social institutions more broadly. The literature on national systems of innovation (Nelson 1993) looks at the extent to which capabilities for industrial development reside in the institutions and policies of government and in non-profit institutions of research and education. A central issue of debate is whether individual countries can be said to possess distinctive ‘national’ capabilities for innovation and industrial growth. Because many of the issues turn on questions of industrial clustering and inter-firm linkages, this literature also intersects with Marshallian notions of industrial districts and does so in a way that sometimes connects with, but often broadens beyond, the NEG.

**Industrial districts and communities of practice**

The main feature that distinguishes industrial districts, sectoral systems of innovation (Malerba 2004) and similar groupings from ordinary market competition as understood in price theory is their high levels of ‘social embeddedness’ (Granovetter 1985). Because of their structure, industrial districts offer important benefits in innovation processes. For one thing, the high levels of differentiation and specialization allow firms, in Smithian fashion, to focus on aspects of the supply chain in which they are especially competent. Second, commentators since at least the time of Marshall have recognized the importance of close social relationship among entrepreneurs and workers in industrial districts (Bellandi 2003). The tight geographical proximity of competing firms within a district works to increase social ties within districts, and both the leaders of firms and their workers are ‘embedded’ (Granovetter 1985) in networks outside their work places. This embeddedness operates along three dimensions: territorial, social, and network. The resulting meetings may be purely extramural (sharing drinks at a bar, schmoozing at the kids’s soccer game) but still promote discussion of common problems – and of new initiatives. Strong ties (Granovetter 1973) among workers, including managers, can increase the amount of information available to firms...
and the readiness of people to share what they know when relationships gain a dimension of friendship to counterbalance the competitiveness among firms.

Labour mobility further enhances the spread of knowledge within industrial districts. When there are many employers, workers can change jobs and roles, moving to other firms to become foremen as Marshall (1975) suggests, or setting up in business themselves if capital requirements are low or financing is easily available. New firms may fail, but talented people who have gone out on their own can then be reabsorbed as employees in other firms, especially where, as in Silicon Valley, entrepreneurship is rewarded but failure is not severely stigmatized (Saxenian 1994).

When embeddedness is strong, the creation of communities of practice (Wenger 1998; Brown and Duguid 2000) generates competences that, although possessed by individuals, are collective in that they are based on a set of practices that is common to all members of a community. These competences (both tacit and codified) can transcend firm boundaries and become characteristics of an entire industrial district. Even when a community of practice is not as all-embracing as Marshall suggests, novices become socialized to a community’s mores and procedures as a result of continual association with colleagues. Communities of practice are also important as arenas of learning in which tacit knowledge is transmitted especially well (Lave and Wenger 1991; Wenger 1998), even though the range of ideas transmitted can be narrowed artificially by the stress placed on the local practices followed within the community. Although, in some cases, the knowledge held by a community can be classed as shared routines, it often has dynamic aspects that help to direct attention to solving problems that are widespread within the community.

Relationships within industrial districts therefore lead to diffusion but also to the creation of new knowledge through shared preoccupations. Because many people or firms can work on a problem simultaneously, a number of different solutions may be found. The result is a larger and stronger ‘gene pool’ within the sector (Loasby 1990: 117), with the further advantage that solutions that are originally regarded as competing may turn out to be complementary and well suited to different niches within the district.

### Differentiation and modularity

In addition to these casual relationships, close proximity within industrial districts can enhance the deliberate exchange of information. Managers who meet cheaply and frequently with suppliers, customers, and competitors can gain a better appreciation of problems in a sector than when forced to communicate at a distance and through writing. The resulting changes to the system can then be integrated by lead firms that collect information along several segments of a supply chain. Lead firms can provide coordination not only of ideas and inputs but also of people and entire firms, who might otherwise not be aware that they have complementary needs and knowledge. This integrating function can be performed by merchants who, as in the early modern putting out system, are in touch with distant markets and able to communicate to small localized firms, information
about what is popular with consumers; but it may also be a function of lead manufacturers that coordinate changes in the physical configuration of technology as well as in design. More recently, as in Silicon Valley, the integrating role has on occasion been undertaken by venture capitalists or lawyers who have a broad generalist knowledge of what is happening in a district and arrange packages of services and make other connections among small highly specialized firms (Kenney and Florida 2000).

Some of these integrating activities can take place without spatial proximity (Jacobson et al. 2001; Heanue and Jacobson 2001–2002). For example, networks of professionals such as those in law or medicine are communities of practice that arguably constitute a geographically dispersed ‘virtual’ industrial district (Savage 1994). In this case, the virtual character of the network has to do in part with the dispersion of customers and the need to produce the product (provide the service) near the consumer rather than in a central location. Shared routines and capabilities in professions are thus a partial substitute for the economies of scale that might otherwise have been gained through centralized production in the mode of Chandler (1977).

The flip side of this observation is that the geographical proximity may encourage implicit integration of firms. When common practices within an industrial district lead to high degrees of consistency of products and processes, the introduction of formal and informal modularity is easier. Formal modularity occurs when there are ‘design rules’ and specified interfaces between components that allow firms to change the components they produce although knowing that this will not require adjustments to other parts of an assembly (Baldwin and Clark 2000). One of the benefits of formal modularity is that it obviates common ownership across stages of production. Because the use of design rules reduces transaction costs, it allows firms to communicate cheaply with little, if any, hierarchical coordination. Codified design rules may be unnecessary in industrial districts, however, as informal modularity can arise when firms within a district have a common vision of what their business is and how they are expected to go about it. The self-image of such firms, as well as their public image, may involve distinctive designs or particular market niches (e.g. expensive or cheap products), in this way, providing guidance to firms along a supply chain about the kinds of innovations that are likely to succeed in the marketplace. On a technical level, familiarity with production processes within a district gives firms, including suppliers of capital goods, a good working knowledge of how their products relate to existing configurations of components. Thus, differentiation and specialization within an industrial district can lead to implicit integration that is highly effective despite its informality because, as long as particular design and production paradigms do not change dramatically, they offer inexpensive guidance on the types of innovation that firms in a district can expect to succeed.

Disadvantages of industrial districts

Despite many advantages of industrial districts for innovation, the factors underlying successful innovation in some industrial districts may turn out to be weaknesses
depending on the broader innovation environment within a trade or industry. Firms in an industrial district may simply be slow to notice changes arising outside their district because they do not have good external channels of communication. As Marshall (Loasby 1990) recognized, close relationships among firms and their workers could reduce their access to knowledge developed outside the district and their willingness to consider ideas from unfamiliar or distant sources.

For instance, decentralized systems of innovation (including industrial districts) may be at a disadvantage in generating genuinely systemic innovations (Teece 1986), that is, innovations that require the development of new components and new ways of integrating components. In such a case, the location of much of the relevant knowledge within a tightly coupled system is likely to facilitate innovation. This need not mean a single vertically integrated firm, but it does mean that the lead or coordinating firms – in modern terminology, systems integrators – must possess a wide range of knowledge or capabilities and must indeed ‘know more than they do’ (Brusoni et al. 2001). They also need to be powerful enough to force other firms to follow their lead.

In general, the ability of firms in an industrial district to jump from one technological trajectory to another (Robertson and Langlois 1994) is often limited by the cumbersome decentralized organization of many districts. Because of high degrees of specialization and the large number of firms that participate in the production process, re-education procedures are likely to be lengthy. Attributes that once were strengths, such as the presence of implicit standards, can turn into weaknesses that retard a transition from one technology to another. Thus, during periods of major change, the role of integrator firms with strong connections to the external environment is especially important because it is unlikely that smaller suppliers of inputs would have the resources to gather information from diverse sources quickly. The upshot could be major centralization of power and, perhaps, the destruction of many smaller firms as they consolidate or disappear. In some cases, exogenous technological shifts can render obsolete virtually the entire set of competences of an industrial district. One such example is the venerable Swiss watch industry that saw its advantage in mechanical watch movements destroyed by the development of the electronic movement in Japan (Langlois 1998). In such a case, no incremental or endogenous processes of innovation could have been expected to respond adequately to the challenge. In the event, the Swiss industry adapted with a centralized response that incorporated some existing competences (such as design and marketing) but left the industry far more vertically integrated – far less an industrial district – than it had been.

Conclusion

Marshall is alive and well in some of the most vibrant reaches of what I have called Industrial Economics. In the economics of organization, especially the dynamic capabilities and resource-based approaches, and in Industrial Dynamics and Industrial Geography, genuinely, Marshallian ideas are alive and well. Like Marshall, these literatures are concerned with economic growth and the evolution
of industrial structure through processes of specialization, differentiation and reintegration.

Notes

1 This chapter draws on Langlois (2006) and Robertson et al. (2008). My co-authors are probably responsible for all the good ideas in what follows, but I remain liable for any errors.

2 On the distinction between formal and appreciative theory, see Nelson and Winter (1982: 46).

3 What G.L.S. Shackle (1967) called ‘the years of high theory’.

4 ‘To use a mode of speaking which workmen themselves use, the skill required for their work “is in the air, and children breathe it as they grow up”’ (Marshall 1975: 197).

References


