

The growth of firms, competition and the maturity of industries¹

Harry Bloch, School of Economics and Finance, Curtin University of Technology, GPO

Box U1987, Perth, Western Australia 6845

harry.bloch@cbs.curtin.edu.au

John Finch (corresponding author), University of Aberdeen Business School, Edward

Wright Building, Dunbar Street, Aberdeen AB24 3QY, Scotland, UK

j.h.finch@abdn.ac.uk

Abstract

We develop a three-way categorization of capacity and its utilization covering micro, meso and macro framings, beginning with the approaches of Marshall, Steindl and Penrose. Within firms, we combine Steindl's analysis on fixed capital with Penrose's consideration of resources, including a higher-order form of capacity. We examine the implications of capacity expansion for competition and for the maturity of the industry, which reflects back on the value of the firm. We then argue that our higher-order concept of capacity can be made operational for empirical study by identifying a critical bottleneck that restrains firm growth at each point in its development.

Key words: capacity, utilization, firms, industries, economic development

JEL classifications: L13, L16, M11, O31

1. Introduction

Capitalism is dynamic. There is in-built provision for accumulation and growth through channelling an income stream to those with ownership interests in the production process. Yet, economic theory generally treats the capitalist firm as a static participant in the economic process, with no explicit role for its growth or development over time. An exception is Edith Penrose's (1995) *The Theory of the Growth of the Firm*.

Penrose's analysis is micro in orientation. There is no consideration of the impact of firm on the larger economy. The missing element in Penrose's analysis is the industry, which provides a meso level of analysis linking micro analysis with the macroeconomy. Marshall (1920) quite properly identifies the industry as the appropriate level of aggregation for analyzing the role of competition under capitalism. Competition is essential to understanding the evolution of capitalism, as it provides the mechanism whereby selection operates to distinguish the dynamics of macroeconomy from the dynamics of the micro units of which it is comprised. With competition and the selection process it generates, behaviour of the whole is not simply the sum of the parts.

An analysis of firm growth that links firm growth to the macroeconomy and provides for consideration of the role of competition is Steindl's (1976) *Maturity and Stagnation in American Capitalism*. Here, firms reinvest retained earnings to expand their capacity to produce current products, but an obstacle exists to unlimited expansion, namely the existence of other firms supplying similar products – the industry. Leading firms are able to overcome this obstacle temporarily through aggressive competition, which slows capacity growth through the reduction profits for reinvestment and the elimination of fringe producers. The latter impact is a type of Schumpeterian “creative

destruction”, with the creation of capacity by leading firms destroying the capacity of fringe firms (Bloch, 2000). Eventually though, the industry matures with the elimination of fringe producers leaving a concentrated core of producers. This concentrated core demurs from further reinvestment because they realize the adverse consequences for their profitability. Hence, maturity is associated with stagnation.

Penrosian firms change as they grow. Managerial resources are the key constraint on growth of enterprising firms in Penrose’s analysis, but managers learn from their experience, adding to their capability for expansion. The expanded managerial capability can be used for overseeing the expansion of productive capacity for existing products through reinvestment of profits as in Steindl. However, it can also be used for overcoming obstacles, like limited demand for a particular product or aggressive competition within a particular product market, by seeking out opportunities for growth through diversification or new product development. Thus, managerial capability represents a higher-order concept of capacity than Steindl’s productive capacity. It also represents a mechanism for rejuvenating a mature industry or shifting a firm’s expansion to a less mature industry, thereby overcoming the tendency toward stagnation.

In this paper we explore the interaction between expansion of capacity at the level of the individual firm and competition at the level of the industry. We are particularly concerned with the implications for the macroeconomy, especially the prospects for avoiding a tendency towards stagnation as capitalism evolves. In the following section, we review literature that deals with capacity at the level of the firm and at the level of the industry, including the work of Marshall, Steindl and Penrose. Section 3 argues for a broadened understanding of capacity, both for the firm and for the industry. In Section 4,

we examine the implications of capacity expansion for competition and for the maturity of the industry, which reflects back on the value of the firm. Section 5 suggests our higher-order concept of capacity can be made operational for empirical study by identifying a critical bottleneck that restrains firm growth at each point in its development. Section 6 concludes.

2. Introducing meso-phenomena to the capacity debate

The manner in which we consider the industry as a distinct unit of analysis makes a significant difference to how we conceive of capacity and its utilization for both microeconomic and macroeconomic analyses. Given the question of capacity and its utilization, the industry is a meso and emergent unit of analysis, which is irreducible to the number of firms (as micro units of analysis) and makes a difference to how we capture the effects of decisions made in firms for macroeconomic performance (Dopfer et al., 2004). We set out a dynamic framework of firms interacting with one another in industrial settings characterized by imperfect competition, and explain how this can be extended for the purposes of evolutionary analyses.

The received literature on capacity and utilization tends to focus on macroeconomic performance and is a development of the Phillips curve debate on inflation and the natural rate (Shepherd and Driver, 2003; Driver and Shepherd, 2005). Following Eliasson (1991), it can be described generally as a micro-macro approach, distinct from our introduction of an additional meso level of analysis. Further, following Dopfer et al. (2004), we envisage a chain of connections between levels of analysis in

which each level has a degree of autonomy, but in which there are “vertical spillovers” between levels, the magnitude and direction of which depend on interactions within a particular level.

Received micro analyses associated with macro-capacity and utilization studies tend either to be predicated upon surveys among firms as categorized within broad productive sectors of economies or based implicitly on theoretical visions of a representative or average firm. We suggest an alternative micro basis, with firms considered as active constituents of industries, themselves seen as coherent functioning meso units of analysis. The basis of function and coherence among firms in the industry is the norm of imperfect competition. By changing the focus of capacity and utilization analysis to the firms and industries, considered as units of analysis (micro and meso levels), we seek to uncover how macro outcomes can be perceived and captured differently to the established micro-macro analyses of capacity.

Our analysis has a distinctly Marshallian flavour to it, drawing from Marshall’s industry analysis (Marshall, 1919, 1920; Loasby, 1978; Raffaelli, 2004). Marshall’s formal analysis of firms and industries develops from his identification of time periods, and especially the short and long runs. The evolutionary aspect of Marshall’s analysis derives from there being heterogeneous firms in industries, even though he does not pursue the full implications of this. Analytically, his invocation of the representative firm prevents a more thorough evolutionary analysis. Empirically or historically, Marshall develops a discursive three-generational model of firms, believing them to lack sufficient longevity and vigour as on-going organizational forms, which again truncates any evolutionary analysis by imposing an additional and exogenous form of selection.

Marshall's time periods provide one dimension to understanding the interaction of capacity and utilization, and one that is additional to the established micro-macro approach. In the short run, capacity is fixed and we can isolate utilization. Many macroeconomic studies interpret their data in this way, despite sometimes using terminology that conflates the time periods and their implicit assumptions with real time (such as using "term" for "run" or discussing "longer or shorter runs") (Langlois 1992). Conceptually, capacity can be varied alongside utilization in the long run, and it may also be the case that capacity is extended through utilization in real time, for instance in learning-by-doing.

Steindl (1976) presents a significant development of the Marshallian theme of heterogeneous firms within industry systems, in which surplus capacity is a cause and consequence of the pattern of competition.² This provides us with interactions in the form of imperfect competition among firms in an industry setting as a second micro-meso dimension of analysis additional to the more established micro-macro approaches. The question we ask of Steindl's analysis is whether his explanation of a competitive process is self-contained or whether there are spillovers with macro consequences. In systems terms, we assess the extent to which interaction among a defined set of heterogeneous firms can be represented as a 'nearly-decomposable' process of competition (Simon 1962).

A third dimension to our analysis – again additional to the established micro-macro approach – draws upon Penrose (1995), the subsequent resource-based view of the firm, and the contemporary focus on dynamic capabilities (Teece et al., 1997; Eisenhardt and Martin, 2000; Dosi et al., 2000; Winter, 2003). These generally resource-based or

behavioural analyses focus on firms to the exclusion of industries, at least in terms of industries being considered as coherent and so defined systems rather than unbounded environments or components of environments.³ But they provide a two-(or more)-layered understanding of capacity, categorizing firms' capabilities as direct and indirect (Loasby 1999A), where the capabilities extend beyond fixed capital. Similarly, Pindyck (1988: 970) argues that a firm generates options to invest in further capacity by virtue that: 'managerial resources, reputation, market position, and possibly scale, all of which have been built up over time, enable it to productively undertake investments that individuals or other firms cannot undertake.'⁴

While understanding a proportion or dimension of capacity as capabilities raises considerable measurement problems and also problems of aggregation across direct and indirect types of capability, especially as the different layers of types may interact, it helps in generating novel research questions. Again, we are interested in the likely consequences of such an analysis as these spillover into (given the firm-centric element of the resource-based approach) both meso and macro levels.

3. Defining capacity and its utilization within firms and within industries

In principle, the definition of utilization is non-controversial. Capacity is established or captured at a point in time, or for a period of time, for a given usage including shift patterns (Steindl, 1976, pp. 24-5). There is also a level of slack built-in as a precaution against fluctuations and inaccuracies in predicting demand, and to cover the behavioural, administrative and political consequences of planning and operating within a firm

considered as an on-going social organization (Andrews, 1949, 1964; Cyert and March, 1992). In historical analysis, it may be difficult to disentangle the effects of capacity change and utilization change, especially where utilization carries with it capacity expansion through unexpected or unpredictable learning by doing (Langlois 1991). This implies that there may be an autonomous component to utilization, or at least one beyond managerial deliberation. But when adopting the Marshallian convention of short and long runs, utilization is part of the short-run analytical framework, while changing capacity can be understood within the long-run analytical framework.

Even in principle, capacity is harder to define than utilization. We can begin with thinking of capacity as a set of connected and at least partially irreversible commitments brought together in order to realize a business plan or an (especially Schumpeterian) entrepreneurial vision. Means of connection can include designed and planned organizational and production systems (Lieberman, 1989; Henderson and Clark, 1990; Brusoni et al., 2001; Buensdorf, 2005) and also market-mediated or more durable relationships between companies (Richardson, 1972; Dubois 1998). Critical to commitment is irreversibility and so a loss of options values (Pindyck 1988). Research into capacity, whether of a macro or micro nature, has tended to focus on fixed capital, which as an expression of irreversible commitment is unified as units and separable from other units. Capacity becomes identified with fixed capital, with empirical analysis often concentrating on the manufacturing and extractive sectors of a national economy in which firms tend to use fixed capital intensively and also face many replacing, upgrading and marginal adjusting decisions (Lieberman, 1989; Driver and Shepherd, 2005).

We present a different framework for understanding capacity that questions the assumptions of unity and separability that are often implied where capacity is identified closely with fixed capital. Following Simon (1962), we question the implication that all units of capacity (including but not totally identified with fixed capital) are uniformly nearly decomposable. We take our lead from two authors whose work has yet to be well integrated into contemporary analyses of capacity and utilization, namely Steindl and Penrose. Steindl's (1945, 1976) framework centres on the industry, as a meso unit of analysis, with capacity commitments being made by heterogeneous firms with differential access to 'outsider' funds, and as a consequence to fixed capital. Steindl – along with contemporaries including Andrews (1949, 1964) and Downie (1958) – is attempting to pursue a Marshallian analysis of firms and industries in the context of imperfect competition rather than in the context of Marshall's "generally competitive" capitalism.

Steindl's starting point is that industries, as meso units of analysis, comprise heterogeneous firms with well-specified connections to external suppliers of finance and fixed capital, and to buyers. The suppliers and buyers are, in contrast to the firms in the focal industry, considered as groups. As Steindl makes few enquiries as to the composition of these groups, implicitly they may as well be homogenous. Further, the boundaries between groups (or other industrial systems) are represented as markets, which are upstream (suppliers of fixed capital and outside finance) and downstream (buyers) of the focal industry (Podolny et al., 1996; White, 2002). Strictly, the buyer market is imperfect because producer firms in the focal industry can influence demand through varying their selling costs, extending to what Steindl (1976, pp. 55-66) terms 'quality competition' with other firms in the industry. Explicitly, markets for external

funds (borrowing) are imperfect, with large firms being able to acquire “external” funds and small firms being excluded from these exchanges.⁵ There are significant barriers to expansion caused by differential abilities in acquiring outsider finance, and to a further extent by indivisibilities in fixed capital: ‘The basic fact with regard to the cost of borrowing is that the long term capital market is open to small firms only at prohibitive cost’ (Steindl, 1945, p. 19).⁶

Firms’ participation in markets for fixed capital is derivative of their participation in markets for external funds. Fixed capital itself is of a commodity or undifferentiated nature at any particular time. Firms will have different capabilities to adapt and apply fixed capital, and this may be related to absorptive capacity (which is speculative intellectual capacity and may prove to be redundant) (Cohen and Levinthal, 1990). Any absorptive capacity to apply fixed capital is secondary to the dominant and differential abilities of firms to borrow external finance in order to buy fixed capital.

The key point about capacity that we can carry over from Steindl is that it is identified with three types of capital: fixed capital, uncommitted or liquid insider finance in the form of retained earnings, and external uncommitted or liquid finance that only a subset of firms in the industry has access to. Capacity in the industry is then a function of firms’ fixed capital, their differential abilities to earn and then retain a proportion of earnings for reinvestment internally, their abilities to borrow outsider finance and a factor that is external to the individual firm, namely the firms’ imperfectly competitive interactions with other firms in the industry (Boone, 2001).

Each firm’s capacity, the site of irreversible commitments in different types of capital by different agents within the industry or connected with the industry, is

considered as part of an industry's capacity. There are oligopolistic interdependencies between firms' capacities. For Steindl, as with Schumpeter's "creative destruction", one firm's differential capability to install capacity can lead to the destruction of capacity elsewhere in the industry if one or more of the heterogeneous firms are producing the same product but at higher unit cost. Less dramatically, one firm's installation of capacity can reduce the value of other firms' capacities differentially, irrespective of the levels of utilization. We discuss value and capacity below, in Section 4.

Penrose (1959) provides a different perspective on capacity, and one that has been developed further within the contemporary research area of firms' dynamic capabilities (Teece et al., 1997; Foss, 1999; Loasby, 1999B; Dosi et al., 2000; Eisenhardt and Martin, 2000; Winter, 2003). By adopting a Penrosian perspective, we question the unity of capacity by re-assessing Steindl's clear-cut identification of fixed capital with capacity. Firms' abilities to execute plans or entrepreneurial visions are no longer separated from the process of developing plans and visions. Internally, a firm's ability to deploy direct capabilities as capacity is constrained by their indirect capabilities in making business plans and identifying resources that can be connected together. Any success in enacting plans can lead to a freeing of resources over time, forming a feedback effect. Formulating and implementing plans of productive activities and entrepreneurial visions is integral to firms' capacities and is installed permanently within the firm, especially in the form or role of a top management team. A further implication is that many employees also make significant and irreversible commitments to the implementation to firms' plans and entrepreneurial visions (Blair and Stout, 1999).

The boundary conditions of Steindl's industry system are the means of connection with other industries (including immediate buyers). The boundary conditions of Penrose's system are in the firm's ability to generate and re-generate resources – in part through learning by doing and through managerial experience – and in capturing these resources as direct productive services. “Productive services” is capacity currently connected with an implemented plan or entrepreneurial vision. Following Penrose, resources themselves (despite lending themselves to the “resource-based view”) are unconnected or awaiting capture and connection by way of plans, visions, and means of implementing these, so that upon connection they may be classified as yielding productive services. By analogy, for Steindl, firms' abilities to acquire and then retain funds internally for re-investment and to borrow external funds are means to capacity, or are an indirect capacity to create direct productive capacity.

Following Penrose, we can expand this “middle type” of firm-specific, but indirect, capacity to include the repeatable activities of reviewing and ranking established and proximate market opportunities, harnessing resources internally, and implementing and/or monitoring routines and standard operating procedures.⁷ The presence of the senior management team in particular can be interpreted as fixed but indirect capacity because of the difficulties in hiring new managers and introducing them into an established management group. This much is concealed in Steindl's analysis through his restrictions on firm heterogeneity, and so his fairly standardized view of firms seeking to install additional capacity through acquiring additional fixed capital.

4. Changes in capacity and in value

So far we have argued for the extension of a default understanding of capacity. We seek to assess the implications for autonomous and deliberative changes in utilization to include both direct capabilities in undertaking activities and indirect capabilities or dynamic capabilities. The latter involve firms envisaging, planning and implementing adjustments to capacity by drawing upon repeatable procedures.⁸ We mention default, as bounds are required on this rather complex vision of capacity in dynamic settings. Implicitly, different kinds of bounds are available to researchers, depending first upon whether they adopt an industry-centric or rather a firm-centric view of the acquisition, adjustment and utilization of capacity.

Steindl displaces the higher-order capacity to structural conditions of acquiring outsider finance in combination with retained and pre-committed insider finance in order to make commitments to a production plan by investing in fixed capital. While only the subset of “progressive” firms in an industry seeks to acquire external funds and new capital goods, the interfaces with finance and capital goods markets from the perspective of firms within the “progressive” subset are generally competitive. Furthermore, the capital goods implicitly have the characteristics of commodities, in the sense that their characteristics are described adequately by a contract of exchange, supported by market institutions of exchange. As components of overall plans and as realized in production or activity processes, the capital goods are by implication risk free in that their exchange is assumed to be captured and described fully through a spot contract. Firms in the focal industry do not need to have installed and maintain on-going absorptive capacity – a higher-order or dynamic capability (Cohen and Levinthal, 1990) – in order to source,

purchase and use capital goods.⁹ As Steindl identifies firms with an industry and with capacity, and capacity with fixed capital, he excludes from his explanation the role for a proportion of employees as similarly permanent of on-going members of a firm responsible for the development, maintenance and practice of indirect or dynamic capabilities in an on-going social organization.

Penrose, along with subsequent contributors to the resource-based view and to the dynamic capabilities approach, presents accounts of connections with other firms, industries, markets and organizations in fuzzy terms (Penrose, 1995, pp. 107-8). Objectively, and in common with Steindl's industry-centric analysis of firms' productive capacities, a focal firm may acquire fixed capital in supply markets non-problematically. Implicitly, fixed capital again has well-established characteristics that allow productive services to be acquired from fixed capital, considered as resources. Purchases are described by contracts, and their exchanges and guarantees are further secured by the conventions of established upstream markets. There are no improvised and on-going interactions between firms, implying that all firms have distinctive boundaries essentially that demarcate their activities from those of actual and potential supplying, rivaling, collaborating and purchasing firms and organizations.

Subjectively, resources that yield productive services, and are available from other firms and organizations through market-mediated exchanges, are subject to the capabilities of a firm's employees – especially its senior managers – of connecting these with an entrepreneurial vision and plan (*ibid.*, p. 77). This encompasses a perceived market opportunity, resources available currently with the firm, and resources that can be acquired. A resource, often fixed capital with well-defined productive services, becomes

combined with a subjective vision or chain of activities stretching across a focal firm and its connections within suppliers and buyers (Podolny et al., 1996; White, 2002).

The products of potential suppliers and the buying behaviour of potential buyers exist independently of a firm's plans. In contrast, the formulation of the plans is a subjective and imaginative process undertaken within firms, such that any "unified" external resources may be combined in different ways with respect to different firms' subjective plans and planning processes. In the first instance, learning by doing and more generally autonomous routinization of emerging and designed planning, implementing, reviewing procedures creates resources over time within the specific location of a firm in activities undertaken from direct and indirect capabilities.¹⁰ In this sense, the firm is not so much a general autopoietic system, but more specifically one that both exhibits and creates redundancy (Brusoni and Prencipe, 2001). In a second sense, although the created resources are free in that they jointly produced and reproduced alongside a plan of activities, they are awaiting connection with a new or amended plan of productive activities before they can be assimilated and adapted as resources yielding productive services. One set (and possibly type) of resources' productive services is required in order to acquire and connect another set (Scarborough, 1998).

Penrose (1995) describes these processes of generating and also capturing resources within firms as "economies of growth," and they are only available to firms that have undertaken a general commitment commensurate with being enterprising. Firms that make a general commitment to being enterprising do so irreversibly as they employ personnel, especially in management teams, to actively and deliberately "know more than they can do at any one moment in time" or with respect to a current plan or combination

of plans (Brusoni et al., 2001; Brusoni and Prencipe, 2001). While Penrose does not use these terms, the categories of direct and indirect capabilities are again useful here. Steindl's firms, by assumption or empirical proposition, only invest in their own industries (Steindl, 1976, p. 41). Penrose's firms are likely to re-invest in themselves and their current activities (there is no industry group) by virtue of accumulated experience among those involved in reproducing both direct and indirect capabilities. A proportion of capacity formed through processes of feedback remains as firm-specific capacity, rather than being translated into what are initially insider finances or earnings. Penrose's firms retain the possibility of diversification, alongside the process of industry consolidation foreseen in Steindl's analysis.

The subset of progressive or enterprising firms, as identified by their planned commitments to expanding their capacities and irrespective of whether considered within firm-centric or industry-centric frameworks, expand in order to create additional value. Changes in value are defined simply as changes in economic rent, and this may subsequently be shared or dissipated across and within the firms and other organizations connected into the range of productive activities, understood from the perspective of a focal firm and/or industry. In Steindl's (1976, pp. 46-50) argument, the prime entrepreneurial act is developing a firm's ability to supply existing products given the alternative actions of altering the firm's capital intensity or gearing ratio. In Penrose's argument, the entrepreneurial act is pervasive in seeking novel combinations for the firm's established and created resources, together with bought-in resources and market opportunities.¹¹

The source of a firm's value in expanding capacity for Steindl is in achieving operating costs that are below the average of the industry, allowing a progressive firm to take over the market share of a rival with higher than average operating costs. Expansion is realised through offering a standardized or commodity product at a lower price and/or increasing selling and marketing activities among actual and potential buyers in downstream markets. This is notwithstanding some spillover effects of progressive firms' marketing that may benefit the whole industry by expanding demand generally. If expansion is to be considered as an entrepreneurial act, then by extension – and in Knightian terms – the reactions of higher and also lower-cost rivals within the focal industry group are sources of uncertainty, as are the reactions of actual and potential buyers (Knight, 1921; White, 2002). By comparison the acts of buying additional fixed capital, financed by acquiring a combination of insider and external finance are not in themselves sources of irreducible uncertainty, given the planned expansion.

We can refer to Menger (1976, pp. 52-6) for a general account of how resources may acquire economic value through being connected to a stream of activities leading ultimately to final consumption. Expanding the means of economic organization can include the indirect satisfaction of wants through means of roundabout production, encouraging and encompassing levels of activities that are successively of higher order and so more distant from final consumption. While Menger identifies firms and households as units of analysis, so also as black boxes, he also identifies a process of commodification through and supported by markets (Hong, 2000). Menger identifies a process by which a set of “useful things” may successively be converted into “economic goods” by means of connection with what we could now call “value chains”. In static

terms and despite the emphasis on processes in Menger's argument, the distinction between economic goods and other useful things is easy to draw. Something has economic value if it is connected with a plan of activities, itself culminating in an act of final consumption, and in so being also exhibits properties of commodity and marketability. The problem is in the proximity of as yet unconnected things, where we can anticipate that something could feasibly be connected, so acquiring the status of economic good.

In Penrose's theory, economies of growth are available through individuals in firms enacting capabilities to connect what we might call proto-resources or anticipated resources with plans, thereby first imagining and then implementing the retrieval of productive services from resources. The act of planning is the entrepreneurial act of envisaging that resources can feasibly acquire economic value. Penrose's firms exhibit Knightian uncertainty in the act of planning and connecting resources with this plan, which is located within firms and to some degree among the activities drawn from indirect capabilities.

From a behavioural perspective, it is the accumulation of capacity perceived to be surplus or under-utilized capacity that lowers the average value of the expanding firm's capacity, which leads firms to embark upon lowering prices and greater selling efforts. This assumes that companies have their main focus of attention in downstream markets, populated by actual and potential buyers (White, 2002). It could be the case that firms have a greater focus of attention in the connections upstream of their main activities. In which case, additional efforts are required in sourcing additional projects, lowering "hurdle rates" below that which are customary for the industry on average for accepting

new projects, and in industrial or business-to-business marketing efforts. Higher-cost firms then become starved of new productive activities and projects to which their capacities may otherwise be employed.

Given Steindl's industry-centric form of analysis, firms are agents in creating additional capacity within their own boundaries, and destroying the industry capacity that is located within the boundaries of higher-cost firms. There need not be Schumpeterian innovation involved in this process of destroying the value of capacity through disconnecting it from chains of activities, but there may well be incremental Marshallian innovation in the industry relating directly to products. When an innovation is available to all as embodied in fixed capital goods, such as with Salter's (1960) vintage-capital model, it is an external economy, conditional upon all firms first acquiring capacity in the forms of internal and external finance. The characterization of the source of technical change in Steindl's analysis and its implications for the dynamics of competition, and hence the creation and destruction of value, are discussed in Bloch (2005).

So much for extending capacity and evaluating its growth as a process. We know from Steindl, Schumpeter and from contemporary contributors to the dynamic capabilities approach – including Henderson and Clark (1990) and Flaherty (2000), though not explicitly in Penrose (1995) – that capacity can be destroyed because its value is reduced. In effect, resources are disconnected from a value-chain. As part of this process, capacity may be utilized at rates that are lower than customary or as envisaged during processes of corporate review and planning. Variations in (under-)utilization can be assessed over business cycles in the macroeconomic assessments of capacity (Driver

and Shepherd, 2005). In micro and meso terms, the “un-valuing” of capacity is perhaps traced more effectively through the changing connections of a focal firm’s activities.

If a product or a production system is decomposable, the principles of modularity, of architecture, interfaces and standards (Langlois, 2002, p. 23), may be threatened, disconnecting the way in which resources yield productive services with respect to a plan. Disconnected resources as direct capabilities are themselves at risk of devaluation if the principles of connection are accidentally destroyed or re-arranged. Capacity that is under-utilized may be so because, strictly, it has ceased to become capacity and has reverted to – in Menger’s terms – “other useful things.”

5. Bottlenecks: A broadly evolutionary approach to empirical analysis

Steindl (1976, p. 14) argues that, ‘In order to establish the capacity of a plant we have in practice to confine our attention to one particular piece of equipment which forms a decisive bottleneck and ... occupies a strategic position in the plant ...’ For Steindl, this is a pragmatic, empirical and almost throw-away remark.¹² It is an admission that the underlying process is complex, but can only be captured partially and by an empirical proxy (Thompson, 1967). Steindl’s advice conforms with his view of industries comprising firms, the heterogeneity of which is determined by differential access to imperfect ‘outsider’ capital markets, rather than any significant differential access to types of capital goods. We can propose that the bottleneck is a stable means of comparison across firms in particular industries for specified time periods given imperfect competition, providing a means of comparing capital intensity and capital

utilization. With the benefit of historical records, we can capture “normal” capacity usage, variation over time for individual firms, and make cross-sectional comparisons.

In this section, we seek to develop the idea of bottleneck from Steindl’s pragmatic concerns, with the bottleneck as the flipside of surplus capacity (Langlois, 1992; Langlois and Robertson, 1995, p. 24). In so doing, we investigate whether the concept of bottleneck is more general than process manufacturing, possibly extending to the subjective activities related to indirect capabilities of firms that Penrose envisages.

5.1 Bottlenecks among fixed capital and in other productive services

Bottlenecks may be understood as both wasteful, as a problem with clear symptoms that are to be resolved, and as a feature of an established operating system. For a bottleneck to signify waste within a firm, a member of that firm must be capable of undertaking search to discover it and to formulate means of its resolution. Someone needs to imagine how things could be different, that is to perceive that there is some as yet unused (potential) capacity that can be captured. This is the essence of Rosenberg’s (1976, p. 125) observation, and also of the case study analyses presented by Nightingale et al. (2003). While Rosenberg writes of novel solutions and innovative activities, Nightingale et al. undertake an analysis of gaining access to unused capacity in large technical systems through installing ‘a class of capital goods called control systems’ (Nightingale 2003, p. 478). Langlois (1992) and Langlois and Robertson (1995) take a more conceptual approach and point out that a bottleneck also implies an anti-bottleneck (unused capacity), which can be accessed by “spinning-out” and so reorganizing different production activities into separate firms.

Resolving waste associated with bottlenecks tends to involve grappling with the consequences of indivisibilities in fixed capital, coordinating fixed capital with varying degrees of indivisibility, and also responding to fluctuating and perhaps irreducibly uncertain demand. Resolution also implies a transition. The bottleneck begins as a normal and potentially undetected feature of a firm's established capacity and as a consequence of adapting capacity to an expected level of demand. The bottleneck's identification involves the recognition of there being some surplus, unutilized or even unconnected capacity. Finally, we can anticipate an articulation and instigation of some plan to identify and resolve the bottleneck through technical or organizational means.

Capturing or releasing any additional capacity requires additional problem-solving activities drawn from indirect capabilities, requiring additional investment in capacity, and some organizational innovation, so changes a state. This is more problematic to capture and measure in Penrose's case than in Steindl's case. For Steindl, the additional capacity is "more of the same". One (progressive) company's capture of its surplus capacity leads to the accidental or deliberate creative destruction among other (marginal) firms in the industry. For Penrose, any potential capacity that is unavailable due to bottlenecks remains as unconnected resources, and cannot become productive services until some additional change is made to the firm's operations.

Bottlenecks could be crucial factors in theoretical accounts of imperfect competition that occur partly through capacity fluctuations in all of micro, meso and macro levels of analysis. They can also be comparable metering points among firms by virtue of firms being in a common industry, once it has also been established that a group of firms are connected through competitive relations by selling in the same market. The

identification of bottlenecks requires in-depth case-study research, as would the further identification of groups of firms involved in competitive relationships in the same market.

Steindl's interest in bottlenecks is in measurement, in the context of capacity being identified with fixed capital and, implicitly, in manufacturing activities. In evolutionary terms, our interest extends to considering how bottlenecks are perceived and resolved within firms and also among firms at the level of an industry. As argued above, this leads us into considering variations in tacit and embedded procedures and shared or comparable routines of practice among firms as indirect capabilities alongside variations in fixed capital. This practical research task potentially provides us with a measure, however difficult to articulate, codify, categorise and compare, which is capable of supporting broad principles of evolution.

Rather than capture heterogeneity along a single index – in bimodal terms as the categories of “marginal” and “progressive” – we can arrange observations around dimensions of a bottleneck defined in the context of a group of competing companies. We can then analyse companies using some evolutionary principles, we hold out the potential of organizing observations again around the notion of how companies perceive, manage and address bottlenecks. The bottleneck anchors the firm over a period of time, being in part a constraint and in part a means of stability.

5.2 Codifying variety and variation

It is a straightforward matter to observe variation across firms in a group at a point in time, irrespective of whether the group is defined tightly as an industry, or less formally as a network of firms with overlapping capabilities and markets once some common

reference point has been established. It is also straightforward to attribute this variation in general terms to idiosyncratic paths of development among corporate bodies and among personnel. It is more difficult to provide a simple explanation of the processes that affect variation and to capture these in sufficient dimensions so as to enable comparison without compromising the inherent complexity. Social scientists have tended to draw upon Lamarckian approaches, in which agents can seek to alter their behavioural tendencies through deliberative selection.

Deliberative and necessarily imperfect imitation is hence a significant source of variation, and one that is appropriate to our analysis of the consequences of capacity growth. Imitation is necessarily imperfect because the imitator has to unravel some observed practice from another context and then adapt this imitation to her own context (which might be simply a matter of having higher unit costs). If routines within firms are to be analogues of a genotype, there is probably 'too much variation' (Andersen, 1994). Routines are continually, but inaccurately reproduced in action, which is imperfectly and irregularly subject to conscious review and selection within companies and competitive and (perhaps) near-autonomous selection between firms in industry-market settings.¹³

We need a way of capturing significant variation as implied by Steindl, but lost in his simple codification strategy of adopting an ordinal scale, and as set out only in a firm-specific manner by Penrose. Bottlenecks are a site of variety and variation, which in principle can be measured, related to direct and indirect capabilities, and also adopted as part of a broadly evolutionary explanation of the interactions between firms and industries.

Steindl has a clear selection process through his 'ideal' pattern of competition. Penrose has no explicit general means of selection across firms because her firms are idiosyncratic and, if successful, are so presumably because they develop niches, comparable with Chamberlinian "mini-monopolies". Steindl's model is an example of survivor selection (Knudsen, 2002). This is mainly because areas of variation, for example in adopting and absorbing new capital equipment (available only to a subset of firms in an industry), and in undertaking successful marketing and selling efforts, are reduced in his model to budgetary choices. Further, and implicitly, these are connected to spending proportions of accumulated capital (or "insider" money), depending on the overall state of capacity in the industry in response to an overall and exogenous level of demand. Simply, the tendency is for companies with lower cost to be selected, given an ordered, unequal but unproblematic access to innovative fixed capital. This is activated once the growth in accumulation among some, or all, progressive firms exceeds that of market demand, leading to surplus capacity among these firms and at the level of the industry.

In straying beyond Steindl's model and including accounts of selection and retention too, we would need to know something of two or more firm-level behavioural strategies, such as "be a progressive firm" or "be a marginal firm". Following Steindl, there is no scope for a Lamarckian interpretation because the scope for entrepreneurs or managers choosing among these phenotypes is strictly one off. Further, the choice is really determined by the availability and scale of both 'insider' and 'outsider' financial capital, and the composition of the industry at any point in time for entry.¹⁴

In attempting to present Penrose's argument in terms of broadly evolutionary principles of selection, we are confronted with a complex series of idiosyncratic processes, combining instances of autonomous and deliberative selection, which may be simultaneous. These are also nested or loosely hierarchical, as with the co-presence of dynamic or indirect capabilities and direct capabilities. The success of a firm depends partly on the resources available to it at any moment, either in-house or close-at-hand, and the capabilities in firms of capturing these resources. These accumulate from past performance and past selection, but without some stability through retention, this is difficult to compare and generalize across firms. Managers also draw upon imagination, improvisation and accumulated knowledge to make judgments as to how to translate some resources into productive services, and also to combine these productive services in alignment with some perceived (and uniquely perceivable) market opportunity.

6. Concluding remarks

In the introduction to the 1976 edition of *Maturity and Stagnation in American Capitalism*, Steindl reflects on developments in the period since the original publication in 1952. He notes that the period since 1952 could be characterized by robust growth, rather than the stagnation he suggests might be expected with the maturity of many American manufacturing industries. He also suggests that his ideas have changed somewhat in two respects that are relevant to our discussion above. First, he notes the expansion of forms of competition by oligoplistic firms, including sales outlays, product innovation and direct investments abroad. Second, he at least partially recants on his

denial of the role of exhaustion of a long technological wave in explaining the decline in American capitalism in the 1920s and 1930s.

Our discussion suggests Steindl's definition of capacity is too narrow to deal with the evolving nature of modern capitalism. We expand the concept of capacity by noting that bottlenecks which constrain the expansion of the activities of firms might occur in a number of different forms, not limited to inadequate plant or capital equipment. We point to the work of Penrose in *The Theory of the Growth of Firms* as providing a higher-order concept of capacity in the form of managerial capability, but suggest even this broader concept is artificially restricted in terms of allowing for the role of imagination and enterprise in overcoming bottlenecks. Instead, we argue for an evolutionary and empirical approach to the study of capacity.

Further, we argue that it is inadequate to consider capacity only at the level of the firm. Instead, we adopt a micro-meso-macro approach. The expansion of capacity by an individual firm (micro level) impacts on competition in the industry (meso level). This, in turn, affects the environment in which firms make investment decisions in plant and equipment, research and development, marketing and managerial resources, which add up to determining the course of demand and supply growth for the aggregate economy (macro level). We follow Steindl in suggesting the industry (meso level), particularly the pattern of competition in the industry, is the critical link in understanding evolution of capitalism as a whole. What is required is the working out of the pattern of competition using the broader set of forms of competition that Steindl identifies and allowing for the role of imagination and enterprise in firms that will perhaps lead to yet further forms of competition.

References

- Andersen, E.S. 1994. *Evolutionary Economics. Post-Schumpeterian Contributions*, London, Pinter
- Andrews, P.W.S. 1949. *Manufacturing Business*, London, Macmillan
- Andrews, P.W.S. 1964. *On Competition in Economic Theory*, London, Macmillan
- Blair, M. and Stout, L. 1999. Team production theory in corporation law, *Virginia Law Review*, vol. 85, 248-328
- Bloch, H. 2000. Schumpeter and Steindl on the dynamics of competition, *Journal of Evolutionary Economics*, vol. 10, 311-28.
- Bloch, H. 2005. Steindl on imperfect competition. The role of technical change, *Metroeconomica*, forthcoming
- Bowles, S. and Gintis, H. 1993. The revenge of homo economicus. Contested exchange and the revival of political economy, *Journal of Economic Perspectives*, vol. 7, 83-102
- Brusoni, S. and Prencipe, A. 2001. Unpacking the black box of modularity. Technology, products and organizations, *Industrial and Corporate Change*, vol. 10, 179-205
- Brusoni, S., Prencipe, A. and Pavitt, K. 2001. Knowledge specialization, organizational coupling and the boundaries of the firm. Why do firms know more than they can make? *Administrative Science Quarterly*, vol. 46, 597- 621
- Buenstorf, G. 2005. Sequential production, modularity and technological change, *Structural Change and Economic Dynamics*, vol. 16, 221-241
- Chamberlin, E.H. 1933. *The Theory of Monopolistic Competition*, Cambridge, MA., Harvard University Press

- Chandler, A.D., Jr. 1962. *Strategy and Structure, Strategy and Structure, Chapters in the History of the Industrial Enterprise*, Cambridge, MA., MIT Press
- Cohen, W. and Levinthal, D. 1990. Absorptive capacity. A new perspective on learning and innovation, *Administrative Science Quarterly*, vol. 38, 128-152
- Cyert, R.M. and March, J.G. 1992. *A Behavioral Theory of the Firm*, second edition, Oxford, Blackwell (first published, 1963)
- Downie, J., 1958. *The Competitive Process*, London, Duckworth
- Driver, C. and Shepherd, D. 2005. Capacity utilization and corporate restructuring. A comparative study of the US, UK and other European countries, *Cambridge Journal of Economics*, vol. 29, 119-140
- Dopfer, K., Foster, J. and Potts, J. 2004. Micro meso macro, *Journal of Evolutionary Economics*, vol. 14, 263-280
- Dosi, G., Nelson, R.R. and Winter, S.G., eds 2000. *The Nature and Dynamics of Organizational Capabilities*, Oxford, Oxford University Press
- Eisenhardt, K.M. and Martin J.A. 2000. Dynamic capabilities: What are they? *Strategic Management Journal*, vol. 21, 1105-21
- Eliasson, G. 1991. Modeling the experimentally organized economy. Complex dynamics in an empirical micro-macro model of endogenous growth, *Journal of Economic Behavior and Organization*, vol. 16, 153-182
- Flaherty, M.T. 2000. Limited enquiry and intelligent adaptation in semiconductor manufacturing, in Dosi, G., Nelson, R.R. and Winter, S.G., eds *The Nature and Dynamics of Organizational Capabilities*, Oxford, Oxford University Press, pp. 99-123

- Foss, N.J. 1999. Edith Penrose, economics and strategic management, *Contributions to Political Economy*, vol. 18, 87-104
- Gintis, H. 2000. *Game Theory Evolving. A Problem-Centered Introduction to Modeling Strategic Behavior*, Princeton, NJ., Princeton University Press
- Henderson, R.M. and Clark, K.B. 1990. Architectural innovation. The reconfiguration of existing product technologies and the failure of existing firms, *Administrative Science Quarterly*, vol. 35, 9-30
- Hong, H. 2000. Marx and Menger on value. As many similarities as differences, *Cambridge Journal of Economics*, vol. 24, 87-105
- Kay, N.M. 1999. Hercules and Penrose, *Contributions to Political Economy*, vol. 18, 67-86
- Knight, F.H. 1921. *Risk, Uncertainty and Profit*, New York. Houghton Mifflin
- Knudsen, T. 2002. Economic selection theory, *Journal of Evolutionary Economics*, vol. 12, 443-470
- Langlois, R.N. 1992. Transaction cost economics in real time, *Industrial and Corporate Change*, vol. 1, 99-127
- Langlois, R.N. 2002 Modularity in technology and in organization, *Journal of Economic Behavior and Organization*, vol. 49, 19-37
- Langlois, R.N. and Robertson, P.L. 1995. *Firms, Markets and Economic Change. A Dynamic Theory of Business Institutions*, London and New York, Routledge
- Lieberman, M.B. 1989. Capacity and utilization. Theoretical models and empirical tests, *European Journal of Operational Research*, vol. 40, 155-169

- Loasby, B.J. 1998. The organization of capabilities, *Journal of Economic Behavior and Organization*, vol. 35, 139-160
- Loasby, B.J. 1999A. *Knowledge, Institutions and Evolution in Economics*, London and New York, Routledge
- Loasby, B.J. 1999B. The significance of Penrose's theory for the development of economics, *Contributions to Political Economy*, vol. 18, 31-45
- Loasby, B.J., 2002, The evolution of knowledge. Beyond the biological model, *Research Policy*, vol. 31, 1227-1239.
- Marshall, A. 1919. *Industry and Trade*, London Macmillan
- Marshall, A. 1920 *Principles of Economics*, eighth edition, London, Macmillan (first published, 1882)
- Menger, C. 1976. *Principles of Economics*, translated by J. Dingwall and B.F. Hoselitz, New York, New York University Press (first published, 1871)
- Moorman, C. and Miner, A.S. 1998. Organizational improvisation and organizational memory, *Academy of Management Review*, vol. 23, 698-723
- Nelson, R.R. and Winter, S.G. 1982. *An Evolutionary Theory of Economic Change*, Cambridge MA., Belknap Press of Harvard University Press
- Nightingale, P., Brady, Y., Davies, A. and Hall, J. 2003. Capacity utilization revisited. Software, control and the growth of large technical systems, *Industrial and Corporate Change*, vol. 12, 477-517
- Penrose, E.T. 1952. Biological analogies in the theory of the firm, *American Economic Review*, vol. 42, 804-819

- Penrose, E.T., 1956. Foreign investment and the growth of the firm, *Economic Journal*, vol. 66, 220-235
- Penrose, E.T., 1960. The growth of the firm – a case study. The Hercules powder company, *Business History Review*, vol. 34, 1-23
- Penrose, E.T., 1995. *The Theory of the Growth of the Firm*, third edition, Oxford, Blackwell (first published, 1959)
- Pindyck, R.S. 1988. Irreversible investment, capacity choice, and the value of the firm, *American Economic Review*, vol. 78. 969-985
- Podolny, J.M., Stuart, T.E. and Hannan, M.T. 1996. Networks, knowledge and niches. Competition in the worldwide semiconductor industry, 1984-1991, *American Journal of Sociology*, vol. 102, 659-699
- Raffaelli, T. 2004. Whatever happened to Marshall's industrial analysis, *European Journal of the History of Economic Thought*, vol. 11, 209-229
- Richardson, G.B. 1972. The organisation of industry, *Economic Journal*, vol. 82, 883-896
- Rosenberg, N. 1976. *Perspectives on Technology*, Cambridge, Cambridge University Press
- Salter, W.E.G. 1960. *Productivity and Technical Change*, Cambridge, Cambridge University Press
- Shepherd, D. and Driver, C. 2003. Inflation and capacity constraints in Australian manufacturing industry, *Economic Record*, vol. 79, 182-195
- Steindl, J. 1945. *Small and Big Business. Economic Problems of the Size of Firms*, Oxford, Blackwell

- Steindl, J. 1976. *Maturity and Stagnation in American Capitalism*, second edition, New York and London, Monthly Review Press (first published, 1952)
- Teece, D.J., Pisano, G.P. and Shuen, A. 1997. Dynamic capabilities and strategic management, *Strategic Management Journal*, vol. 18, 509-33
- Thompson, J.D. 1967. *Organizations in Action*, New York, McGraw-Hill
- White, H.C. 2002. *Markets from Networks. Socioeconomic Models of Production*, Princeton NJ. and Oxford, Princeton University Press
- Winter, S.G. 2003. Understanding dynamic capabilities, *Strategic Management Journal*, vol. 24, 991-995
- Woodward, J. 1965. *Industrial Organization in Theory and in Practice*, Oxford, Oxford University Press

Notes

1. We have been assisted in writing this paper by Curtin Business School awarding Visiting Research Fellowships in 2003 and 2005 and by the University of Aberdeen College of Arts and Social Sciences awarding a Visiting Scholar grant in 2004. Earlier versions have been presented at the conferences of the European Association for Evolutionary and Political Economy in Maastricht, November 2003 and the International Schumpeter Society in Milan, June 2004, and at a seminar at Lancaster University Management School in November 2003. We thank participants at these seminars, and also Virginia Acha, Louis Araujo, Michael Dietrich, Steve Fleetwood, Debbie Harrison and Robert McMaster for their comments and criticisms.

2. We use the phrase “industry systems” in preference to “industries” in order to emphasize even in abstract terms the discrete and also ongoing nature of the industry as our meso unit of analysis. The problem of drawing boundaries undermines many readers’ appreciation of Chamberlin’s (1933) exposition of imperfect competition

3. Penrose’s approach lends itself to case studies in which individual companies form the focus of inquiry, as in her studies of the Hercules Power Company and of Holden Motors (Penrose, 1956, 1960; Kay, 1999).

4. Loasby (1999A) also argues that the firm can be seen as a system of options. Chandler (1962) makes a comparable argument, with firms being useful institutional forms for

generating and then using up fixed capital assets, which are risky to hold for long periods prior to pay back.

5. Bowles and Gintis (1993) describe these borrow-lender exchanges as contestable exchanges. In conditions of asymmetry and uncertainty, exchanges within the institution of markets can be closed by additional means, such as lenders dividing potential borrowers into short- and long-term structures.

6. Steindl sets out a table with supporting data showing the differences in the cost of floating securities and another showing the cost of borrowing for loans (short-term credit) decreases with the size of the loan.

7. Following Winter (2003), repeatability is an essential feature of indirect or dynamic capabilities. Repetition itself implies that indirect or dynamic capabilities can be shared, so are not characterized by personal and subjective knowledge, and have the character of large-grained corporate routines

8. Following Winter (2003) and Moorman and Miner (1998), improvisation is much more a matter of personal knowledge and is difficult to attribute to a firm's dynamic, indirect or even higher-order capabilities as significant features of a firm's capabilities are repeatability and being shared.

9. And by implication, sellers of capital goods face simple – as opposed to complex – tasks in selling their products to firms in the focal industry, in contrast to the functional requirement of short and long-sider relationships in markets governing exchanges of external finance.

10. We do not mean to imply any primacy of a higher-order corporate activity, or any particular ordering. These activities may be undertaken in different connections and intensities. The Penrosian point is that routinization releases resources, and these can potentially yield further productive services given additional plans and patterns of connection.

11. Penrose (1952, p. 809) argues for caution in the application of biological analogies: ‘It can be admitted that to some extent firms operate automatically in accordance with principles governing the mechanism constructed, but to abandon their development to the laws of nature diverts attention from the importance of human decisions and motives, and from problems of ethics and public policy, and surrounds the whole question of the growth of the firm with an aura of “naturalness” and even “inevitability.”’ There is no inevitability as growth depends on the connection of resources in the form of productive services to entrepreneurial visions. The “limited extent of automatic operation” implies at least two levels of capabilities, aside from a firm’s capacity to generate and capture new resources and entrepreneurial visions.

12. This is of interest for two reasons. First, it allows us to think of capacity within a plant, establishment or firm to do with bottlenecks and anti-bottlenecks (Langlois and Robertson 1995). Second, it allows us to expand our analysis to different types of production process following the industrial sociology tradition of Woodward (1965) and Thompson (1967). Lanlgois (1992) advises that some bottlenecks are organizational in origin and can be ameliorated by spinning-off stages of production and redrawing corporate boundaries.

13. There is no necessity in characterizing external relations with customers and rival producers in markets as autonomous selection. There is a good deal of deliberation, especially in business-to-business relationships, although a cluster of connected firms may be jointly selected by processes better characterized as near autonomous (Loasby, 2002). Downie (1958) introduces the notion of an “innovation mechanism” as a means for firms to modify routines that are relatively inefficient.

14. A set up of two phenotypes or strategies is consistent with an analysis drawing upon the technique of evolutionary game theory (Gintis, 2000). In this context, it is worth thinking about the advantages of being a “marginal firm” relative to being a “progressive firm.” It may be in terms of entrepreneurs wanting to use up old fixed capital, or being risk averse and not wanting to sink “insider” finance.