STRATEC

Competition in Increasing Returns and Network Industries

"Increasing Returns is incompatible with perfect competition as has been known since the work of Cournot" – Kenneth Arrow (1993)

Abstract.

This paper reviews and integrates in a conceptual model heretofore disjointed conceptual contributions that explicitly and implicitly report the notion of increasing returns (IR) to address dynamic polarization phenomena in business competition. In contrast to conventional views which ordinarily associate business competition (and success) with a very reduced number of variables and conceive the growth of the firm, industrial structure and technological competition as independent levels of analysis , the model suggested moves away from this decomposing approach to suggest that competition in IRIs is better characterized by a more comprehensive framework where a variety of IR mechanisms combine to explain simultaneously three likely features of IR competition: (i) big firms with non-negligible market shares cluster around concentrated industrial structures with skewed distribution of firm size and market share, (ii) there is a tendency of certain cases of technology competition to end up in a temporary monopoly,
(iii) the competition process usually follows a dynamic, cumulative pattern that starts symmetric and ends up asymmetric.

For an industrial policy toward promoting economic growth it would be of prior importance to complementarize IRIs which would enhance their value creation and competitiveness.

Keywords. Industrial Organization, Increasing Returns, Dynamic Competition, Network Externalities

1 Introduction to Increasing Returns

Most industrial sectors of highly industrialized economies are not perfectly competitive. They are usually formed by a small number of big firms with non-negligible market share.

Besides being prevalent in the economy, big firms cluster around concentrated industrial structures which exhibit a skewed distribution of firm size and market share. This situation may be brought about by the intrinsic potential of dynamic technological competition to end up in (temporary) technological monopoly, so in those cases industrial competition may start out symmetric but end up asymmetric.

In this paper we show how the competitive process proliferates in increasing returns industries (IRIs) where the total of all unit activities linked together yield a higher return than the sum of the individual unit activities operating separately. For this to be happening we must show that a variety of increasing returns mechanisms combine to enable the effect of an increasing returns industry.

We propose an integrated framework to provide tools and insights for explaining competition among skewed industrial structures. However, it is only a tentative step toward attempting to explain the path-dependent, indeterminate, suboptimal, locking-in nature of technological competition under increasing returns.

Because of this we partially review the literature on the dynamics of technological diffusion, substitution, and competition. The purpose of this review is to show that we cannot accurately understand industrial competition without taking into account the self-reinforcing nature of commercial success in most emerging markets.

We enrich increasing returns mechanisms by incorporating a set of stronger, yet neglected, increasing returns mechanisms -reputation effects, infrastructure effects and positive network externalities into a preliminary framework model. The resulting theoretical structure, we will argue, captures the interdependent and cumulative character of the three aspects of industrial competition: the number and size of firms, skewed industrial structures, and the nature of technological competition.

The increasing returns discussion in economics has provided important insights into the characteristics and dynamics of modern industrial economies. However, the discussion on policy applications has (mis)led some authors and policy analysts to conclude that a completely *new economy* is emerging and that it obeys a set of rules, which are totally different from those that apply to traditional sectors of the economy. While it is undeniable that the increasing return paradigms remain fairly new and revolutionary and while there is no doubt that this paradigm is key to our understanding of new industrial sectors,

and their sustaining role in productivity growth, we should clarify its proper role in industrial structure and growth of the economy. .At this stage we are most concerned about the catalytic role of technological competition in increasing returns industries. Increasing returns industries are nowadays most likely to be identified with high technology industries, in particular with information, communication and health care related industries (Gottinger, 2003). As an example, in a corporate context, how to unlock increasing returns in its global operations, consider General Electric (GE). It constantly evolves its portfolio to drive growth despite its large size and already significant presence in major markets. It encourages its executives and business units to take an expansive view of its markets as a means of unlocking growth initiatives that a product centric view would miss. Often, when its market share exceeds l0 percent, it seeks to redefine the market more broadly to include adjacent products or services . This continual questing lies behind successful moves from manufacturing to services , that has allowed it to keep growing in complementarizing given industrial markets .

For those industries Shapiro and Varian (1999) have suggested a combination of supply-side scale economies and demand-side scale economies to explain the intrinsic aspects of technological competition. It appears however that this way of seeing technological competition is too simple to capture the variety and complexity of real-world businesses in those industries Thus we suggest a general framework to describe technological competition in what we are going to call the *increasing returns economy.*

Here is an outline of what follows. Section 2 identifies supply-side scale economies as a major ingredient of increasing returns economies. Section 3 lists increasing returns properties as part of a Schumpeterian mechanism. The relationship of increasing returns and non-ergodic markets is explored in Section .4. Section 5 explores technological competitive paths subject to uncertainty of technological outcomes whereas Section 6 focusses on technological competition under standard setting. Section 7 relates the intensity of technological competition to network externalities, and Section 8 provides a summary of increasing returns factors in competitive settings. Section 9 provides an example of increasing returns, structural change and development conomics. Finally, Section 10 draws conclusions on the systemic connection of increasing returns and technological competition.

2. Supply-Side Scale Economies

A first source of increasing returns assuming constant technology identifies a concentrated industry structure as a result of supply-side scale economies. In many cases large firms are more efficient than smaller companies because of its scale: larger corporations tend to have lower unit costs. This efficiency in turn fuels further their growth. However, positive feedbacks based on supply-side economies of scale usually run into natural limits. Past a certain size companies find growth difficult owing to the increasing complexity of managing a large organizational structure. From then on, negative feedback takes over. As traditional supply-side economies of scale generally become exhausted at a scale well below total market dominance, large firms, burdened with high costs, never grow to take the entire market and smaller, more nimble firms can find profitable niches. Shapiro and Varian (1999) conclude that because of this most industrial markets are oligopolies rather than monopolies.

Negative feedback generated by the difficulties of managing large organizations (scale. diseconomies) indeed interrupts the growth of the firm and the level of industrial concentration. This situation, nevertheless, may be transient, because firms may be subject to other sources of increasing returns. Large firms that go through increasing returns mechanisms other than scale economies may increase their efficiency and overcome the negative aspects of overgrown organizations. Industries in which scale diseconomies are counterbalanced by other increasing returns mechanisms, then, may begin to head toward the extreme of 'winner-takes-most' situation. The increasing returns mechanisms capable to offset scale diseconomies are usually related to technological progress, so in the following sections we analyze other major causes of the growth of the firm, namely, the Schumpeterian loop, cost reducing learning, learning-by-doing, learning-by-using, and demand-side increasing returns.

3. Schumpeterian Mechanism

The most widely accepted theory of technological change in modern economics is Schumpeter's (1942). In the Schumpeterian world, scale economies are present as well, but technology is not a constant. Here the creative role of the entrepreneur allows for the introduction of new technologies capable to displacing the established ones. In the context of IRIs, Schumpter's main point has been that innovation competition leads to increasing returns economies triggering serial innovations inducing more IRIs (Freeman,2003). Most of Schumpeter's discussion stresses the advantages of concentrated market structures involving large firms with considerable market share. In his view, it is more probable that the necessary scale economies in R&D to develop new technologies be achieved by a monopolist or by the few large firms of a concentrated industry. Large size firms, besides, may increase their rate of innovation by reducing the speed at which their transient rents and entrepreneurial advantage are eroded away by imitators. In the absence of patent protection large firms may exploit their innovations on a large scale over relatively short periods of time -and in this way avoid rapid imitation by competitors- by deploying their productive, marketing and financial capabilities. Large firms may also expand their rate of innovation by imitating and commercializing other firms' technologies.

Schumpeter's thesis encouraged a large body of empirical literature in the field of industrial organization. Most of this literature focused on two hypotheses associated with Schumpeter's assertion: (1) innovation increases more than proportionally with firm size and (2) innovation increases with market concentration.

The most comprehensive review of the empirical evidence of the relationship between innovation and firm size and market structure is Cohen and Levin (1989). They observe that the empirical results on the Schumpeterian relation are accurately described as fragile. They note that the lack of robust results seems to arise in part from the inappropriate attention to the dependence of these relationships on more fundamental conditions. From their overview Cohen and Levin (1989) draw the basic methodological lesson that the omission of important and potentially correlated variables that influence innovation can lead to misleading inferences concerning firm size and concentration

Following Schumpeter's lead, Richard Nelson and Sidney Winter (1978, 1982) stand out for having formalized and completed many of Schumpeter's original intuitions. Whereas the connection between industrial structure and innovation has been viewed by Schumpeter as going primarily from the former to the latter, in Nelson and Winter (1982) there is a reverse causal flow, too. That is, there is clearly a circular causality suggesting a self-reinforcing mechanism between the innovations and the firm’s growth. Nelson and Winter (1982) stand out not only for having recognized the endogeneous character of innovation and market structure, but also for having pointed out and modeled the mutual causality between technical change and market structure. (Nelson, 1986).

Evolutionary economists (like Nelson and Winter) define innovation very broadly. It encompasses product and process innovation, opening up new markets, and acquisition of new sources of raw material. They also describe the nature of technical progress as succession of major discontinuities detached from the past and with quite transitory life span. This process of change is characteristic of certain industries, but it is not the sole kind of technological change. Technological change can also be *continuous.* That is to say, technologies improve constantly in absolute terms after their introduction. The view of technological progress as a continuing, steady accumulation of innumerable minor improvements and modifications, with only very infrequent major innovations, has two sources: (1) the accumulation of knowledge that makes possible to produce a greater volume of output from a given amount of resources and (2) the accumulation of knowledge that allows the production of a qualitatively superior output from a given amount of resources. The former source of technological progress is the result of a *cost reducing learning process,* while the second category is the result of what is known as *learning-by-doing* and *learning-by-using.* Given that both categories of technological progress are important determinants of the number and size of firms in a given industry, we analyze them in the next sections.

Cost Reducing Learning

An important aspect of technological change is costs reducing in nature. As we saw before, Porter (1980) and Henderson (1975), in the strategic field, pioneered the notion of experience curve as a source of cost reductions. In economics, Hirsch (1956) has underlined the importance of repetitive manufacturing operations as a way of reducing direct labour requirements, while Arrow (1962) has explored the consequences of learning by-doing (measured by the cumulative gross investment, which produces a steady rate of growth in productivity) on profits, investment, and economic growth. However, the historical study on the patters of growth and competitiveness of large corporations of Alfred D. Chandler (1990) is a major and detailed contribution to our understanding of the way firms grow by diminishing costs.

Large corporations, according to Chandler, along with the few challengers that subsequently enter the industry, do not compete primarily on the basis of price. Instead they compete for market share and profits through functional and strategic effectiveness. They compete functionally by improving their products, their processes of production, their marketing, their purchasing, and their labour relations. Big corporations compete strategically by moving into growing markets more .rapidly and effectively than do their competitors. Such rivalry for market share and profits make more effective the enterprise's functional and strategic capabilities, which, in turn, provide the internal dynamics for continuing growth of the enterprise. In particular, it stimulates its owners and managers to expand into distant markets in its own country and then to become multinational by moving abroad. It also encourages the firm to diversify and become multiproduct by developing and introducing products in markets other than the original ones.

Learning-by-doing

Some of the writings on industrial competition assumes that firms compete mainly in cost-reducing competitive advantages, especially those achieved through scale economies, scope economies (economies of joint production and distribution), and innovation in production and organizational processes. Here technical progress is implicitly treated as the introduction of new processes that reduce costs of producing essentially unchanging products. Beyond, there is a category of learning known as 'learning-by-doing' (Rosenberg, 1982) which enhances the qualitative aspects of final products.

Western industrial societies today, Rosenberg (1982) argues, enjoy a higher level of material welfare not merely because they consume larger per capita amounts of the goods available. They have also made available improving forms of rapid transportation, instant communication, powerful energy sources, life-saving and pain-reducing medications, and other goods that were undreamed of one or two centuries ago. Therefore, ignoring product innovation and quality improvements in products is to overlook what well has been one of the most important long-term contributions of technical progress to human welfare. Many products, such as beverages, toothpaste, soap, clothing, VCRs, TV sets can be subject to improvements. Such improvements, however, are marginal when compared with the amazing rate of development that other products and technologies can reach. Automobiles, aircraft, flight simulators, computers, and nuclear reactors are very complex technologies and, as a consequence of this, have a tremendous capacity of being enhanced. Consequently, the competitive behavior of the firms that produce these technologies consists not only of the innovative acts they perform to improve production, organizational, and distribution processes, but also from the efforts to improve constantly their products.

Learning-by-using

With respect to a given product, Rosenberg (1982) distinguishes between that kind of learning that is internal to the production process (learning-by-doing) and that which is generated as a result of subsequent use of that product (learning-by-using). The later category of learning begins only after a certain new product is used. In an economy where complex new technologies are common, there are essential aspects of learning that are a function not of the experience involved in producing a product but of its *use* by the final consumer.

The optimal performance of durable goods (especially complex systems of interacting components) often is achieved only after intensive and prolonged use. In the aircraft industry, for instance, the attainment of high standards of reliability is a major concern, in particular during the development stage. But it is only through extensive use of aircraft by airlines that faults are discovered and eliminated and detailed knowledge is gained about metal fatigue, weight capacity, fuel consumption of engines, fuselage durability, minimum servicing, overhaul requirements, maintenance costs, and so on.

Demand Side Increasing Returns

In the economy there are increasing returns mechanisms that come from the demand side of the market, not just from supply side. For the average (risk adverse and imperfectly informed) consumer it becomes more attractive to adopt a widespread technology or product. Minimizing the risk of purchasing a defective technology or the cost of searching for an adequate one introduces a reputation or informational feedback that may produce a disproportionately high selection of the best-selling option. Informational or reputational feedback effects occur in various situations that could be reinforced through network externalities. First, when the complexity of the technology or product in question is such that consumers try to reduce uncertainty by asking to previous purchasers their experience with these technologies (Arthur and Lane, 1993). Second, in other situations the source of uncertainty is not the complexity of the technology, but the large quantity of options the consumers face. One is bound to choose, and the best way to do so is by confining one's attention to the best-assessed items in the consumer report. Third, in a market where the quality or value of a product is defined on the basis of arbitrary and short-living conventions, rather than strictly on the basis of lasting objective value, consumers usually tend to follow the expert's opinion. This kind of easy-to-manipulate, reputation-driven market success is typical of markets for highly symbolic products (e.g. many art markets, fashion wear and luxury items) , which also will result in a disproportionately high selection of one of the options.

Finally, the most preeminent and common kind of reputation effects in the economy, arise plainly as a result of a well-timed and very aggressive advertising campaign. This self-reinforcing mechanism -and the lasting market dominance that it causes- might be quite unrelated to relative added value, but it certainly might produce an excessive predilection for one of the options.

By moving beyond the Schumpeterian hypotheses and focus on a more complete model of industrial competition we have identified other fundamental determinants of technological change that affect the mutual link between firm size and market structure (Aghion and Howitt, 1998).

These determinants -which in our analysis take the form of increasing returns mechanisms- are usually studied as if they work independently from the other. But there are not many cases of industries where one single mechanism acts in isolation from the other sources of increasing returns. Therefore, the growth of the firm and the evolution of skewed industrial structure, more than the result of a single self-reinforcing mechanism, are the effect of the combination of several sources of increasing returns, which overlap and feed back upon one another..

As depicted in Figure 1, the unification of the resource-based loop, the

Schumpeterian loop, scale economies, the different categories of learning, and demand-side increasing returns (reputation) – loops A, B, and C, respectively, in Fig. 10.1—constitutes a simple but useful model capable to explain endogeneously the number and growth of firms in a given industry, and in a wider context, the gap of economic performance in a given industrial sector.

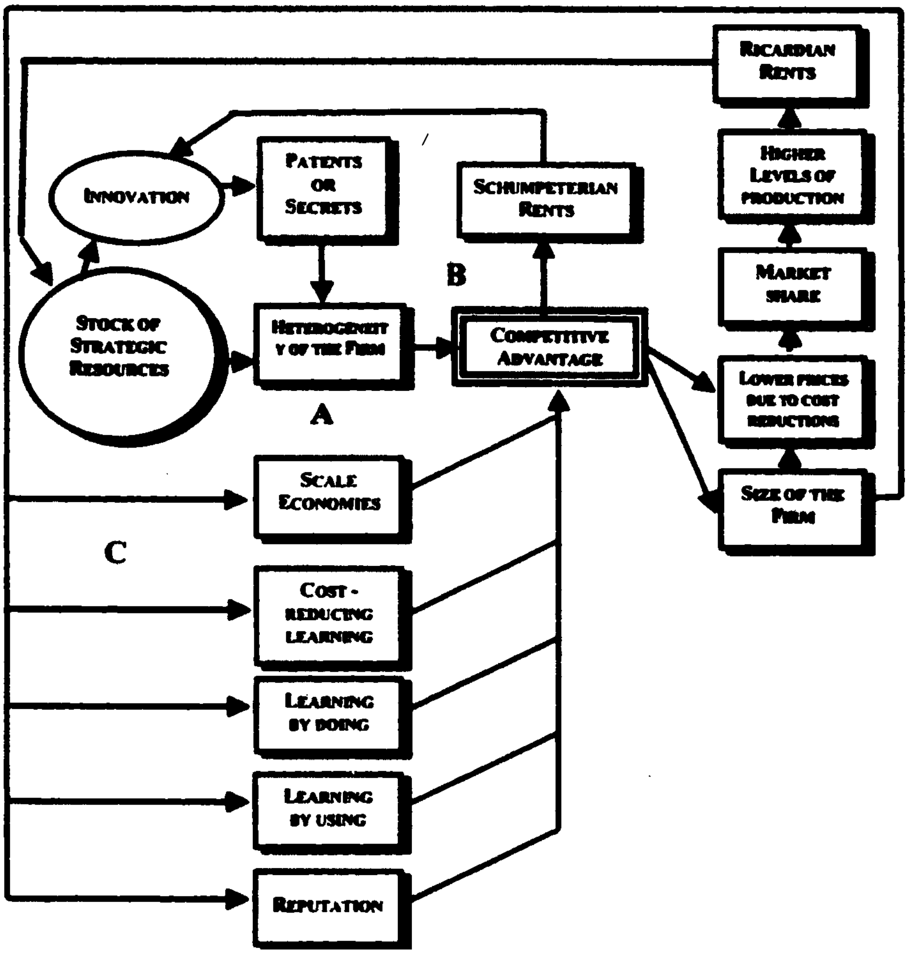

Figure 1 Increasing Returns Mechanism: A Qualitative Model of Industrial Competition

In the model sketched in Figure 1 the positive relationship that runs from industrial structure to efficiency operates through the accumulation of rare resources, innovations, scale economies, reputation, and the different aspects of learning. This dynamics, over time, makes costs fall as learning accumulates, new technologies are developed and improved, and firm-specific factors are amassed and exploited due to output increases. As a result of this mutual causality, market share and production levels increase, price falls, profitability rises, and with which relatively profitable firms expand continually while unprofitable ones contract uninterruptedly.

A relevant aspect of the structural determinants of the number and size of firms in an industry suggested in this model is that, when one of them is exhausted, causing a slowdown in the growth of the firm, the other mechanisms may be activated, which may allow for a further period of continued rapid growth. When the firms of a given industry are capable to accumulate firm-specific resources, innovations, costs reducing learning, qualitative product innovation based on learning-by-doing and learning-by-using, and reputation, these firms usually use them as strategic weapon. In doing so, they are capable not only to neutralize but also to overwhelm the negative effects of complex, overgrown, hard-to-manage organizational structures that arise from their constant growth. The process can take a long period of time, but eventually the sources of increasing returns can drive markets toward increasingly skewed industrial structures.For instance, in the commercial aircraft industry competition principally involves considerable development costs, continuous improvements in aircraft models, technology and product support, so this industry exhibits substantial scale economies, large scope for learning-by-doing, learning-by-using, and reputation effects. Because of this, the commercial aircraft industry has been characterized by an increasing skewed industrial structure. Recently, the structure of this industry, after the acquisition of McDonnell Douglas by Boeing, was reduced to a monopoly in the United States. In the world aircraft market Boeing only competes with European Airbus. It is obvious that the merge of the two main manufacturers of the American aircraft industry should have brought about some gain in efficiency, which counterbalanced the diseconomies owing to the managing a more complex organization. Otherwise, the merger would not have taken place or would have been the result of irrational behavior.

The structure of some industries does not seem to head toward monopoly. However, over time, their level of concentration has increased substantially. The world automobile industry, for instance, in 1970 was composed of at least 40 major competitors. In 1998, with some mergers and acquisitions, the number of the main manufacturer was reduced to 17. Because of large possibilities to accumulate cost reducing learning and the large scope for qualitative product improvements in the world automobile industry, both the number and the homogeneity of the firms competing in this industry are expected to decrease even further in the future. Here, again, benefits due to both costs reducing learning and qualitative product innovations brought about by merges and acquisitions are larger that any cost created by scale diseconomies. Another interesting aspect of this model is that it also offers an endogenous explanation of the number and size of firms. In contrasts with the traditional economic views -that see industrial structure (number of firms) as an exogenous variable and assume homogenous firms- and the strategic paradigms -which are focused first and foremost in explaining heterogeneity among firms within an industry-,this model recognizes that the strategic choices and acts of the firms have an effect not only on the performance and size of the firm itself, but also on the structure of market. In summary, industrial structure is caused by a combination of various increasing returns mechanisms. Here, then, the combination of accumulation of resources, product innovation, scale economies, cost reducing learning, learning-by-doing, learning-by-using or reputation enhances the performance of the firm and determine, to a great extent, the level of skewness of the structure of the industry where it competes.

4. Increasing Returns and Ergodic Markets

Conventional economics has tended to portray most economic situations as something analogous to a large Newtonian system, with a unique and stable equilibrium solution predetermined by a given pattern of resources, preferences, and technological possibilities. Brian Arthur and his group (cf. Arthur, 1994b; Arthur et al., 1987), however, have shown that this conventional way of seeing economic reality overlooks important and frequent economic situations where increasing returns are conspicuous. In order to distinguish economic situations characterized by decreasing returns from those where increasing returns are dominant, Arthur, Ermilov and Kaniovski

(1987) developed the theory of *non-linear Polya processes,* which describes the long-run self-organizing structures that emerge from dynamic processes where proportions are involved. The general non-linear Polya scheme can be pictured by imaging an um of infinite capacity to which balls of several colors are added .In the simplest case, where decreasing and constant returns prevail, the probability of a ball of a given color to be chosen the next time is independent of proportions of colors at the moment of the addition. In this simple sequential process, the strong law of large numbers predicts that, over time, the proportion of balls of color i has a fixed probability, where $\Sigma_i q(i) = 1$. Therefore, it has a unique, predetermined outcome. outcome. Sequences of choices in these simple cases are important at the beginning of the process. However, as the process advances, different sequences of choices are averaged away by the economic forces, which are subject to constant or decreasing returns. So, no matter the sequences of choices, the system will always -- with probability one – end up into the same pattern. For instance, in a coin-tossing experiment the event "head" is independent of previous tosses, then the expectation of a "head" in each toss is 0.5 no matter how many times the experiment is repeated. Likewise, the proportion of 6's in a dice-casting experiment will tend to 1/6. The process by which firms in an industry concentrate in different regions is like the coin-tossing or the dice-casting experiment, if the geographic preferences of each firm is not modified by the preferences of the other firms.

In more general cases -where increasing returns are present- the dynamics is completely different and the standard strong law is inapplicable. In this regime, the next ball to be added into the urn is not known, but the probability of adding one ball of specific colour depends on the present proportions of colours in the urn. In other words, the probability of an addition of the colours becomes a function of the proportions of ball of each colour at each time of choice. The case of firms deciding where to settle down illustrate this kind of non-linear Polya processes. Here increasing returns can be incorporated within the model by introducing agglomeration effects. Because of agglomeration effects, additions to a specific region are not independent of previous locational choices and firms are added incrementally to regions with probability exactly equal to the proportions of firms in each region at the time. Under increasing returns, then, the process becomes *path-dependent.*

Arthur et al. (1987) show that at the outset of the process proportions are not stable, but once the industry settles into a vector of proportions, locational patterns become constant at that vector with probability one. However, the constant vector is selected randomly from uniform distribution over all possible shares that sum to 1.0. This means that each time this locational process is rerun under different historical events, it will in all likelihood settle into a different pattern. Therefore, it is possible to predict that the locational pattern will tend toward a constant proportion, but it cannot be foreseen at *which* proportion it will settle down.

The interpretation of economic history is different under different regimes[3]. Under constant and diminishing returns, the evolution of the system is *ergodic.* Ergodic structures emerge when repeated random events -that are drawn from the same distribution and are independent from previous ones- have a long-term average that approach their expected value. While other results might be possible, they have probability zero to occur. The typical example of an ergodic system is coin tossing. If a fair coin is tossed indefinitely, the proportions of heads varies considerably at the outset, but settles down to 50 per cent with probability one. The evolution of an ergodic system, therefore, follows *a convex* probability function, which has expected motions that lead toward a unique, determinate outcome. In this regime "historical chance" cannot influence the evolution of the systems so history is reduced to mere deliverer of the inevitable and the known.

Under increasing returns, by contrasts, the process is *nonergodic,* because small historical events become magnified by positive feedback. A nonergodic system follows a nonconvex probability function, so two or more outcomes are possible and "historical chance" determines which of these is ultimately selected. History becomes all-important. There are some cases of nonergodic systems in which , from the multiplicity of structures that may emerge, there are some "corner solutions" with a single option monopolizing the choices. In this specific kind of nonergodic systems, while information on preferences, endowments and transformation possibilities allows locating and describing the various possible corner equilibria, it is usually insufficient to determine which one will be selected. In these cases, as Arthur (1994b: 13) has pointed out 'there is an *indeterminacy* of outcome'.

Adoption of technologies that compete under diverse regimes can be appropriately modeled as a nonlinear Polya process (cf. Arthur, 1989, 1990a, 1990b; Arthur et al., 1987). In the simplest regime, when technological competition is characterized by constant and decreasing returns, the probability of a technology of being chosen depends on its current market share. As each adoption is independent of the previous one, market share should converge to a point where they equal probability. Therefore, under constant and decreasing returns two technologies or products performing the same function will end up sharing the market according to each technology's intrinsic value and technical possibilities. Therefore, markets characterized by constant and decreasing returns can be called *ergodic markets*.

Under increasing returns to adoption, the probability of adoption depends on the numbers of adoptions holding each technology at a particular time. Markets of this kind can be called *nonergodic* markets. Within this kind of markets there are those where increasing returns may drive the outcome toward a single dominant technology, with small events early on selecting the technology that takes over. This particular type of nonergodic markets can be termed *tipping or indeterminate markets.* This indeterminacy relates to the "selection problem" – how one allocation outcome is "selected" over time by small historical events when there are several possible long-term results.

5. Technological Competition under Uncertainty and Inertia

In high technology markets the commercial success of emerging new technologies is both highly uncertain and inertial. As regards uncertainty, in addition to the problem trying to discern the true potential of a new generic technology, there is also the difficulty of foreseeing the precise direction in which the said technology will evolve. Indeed, as depicted in the Figure 2, the emergence of a new generic technology generally opens the door not just to one specific technological path, but rather to a whole variety of possible trajectories in product design and process technology.

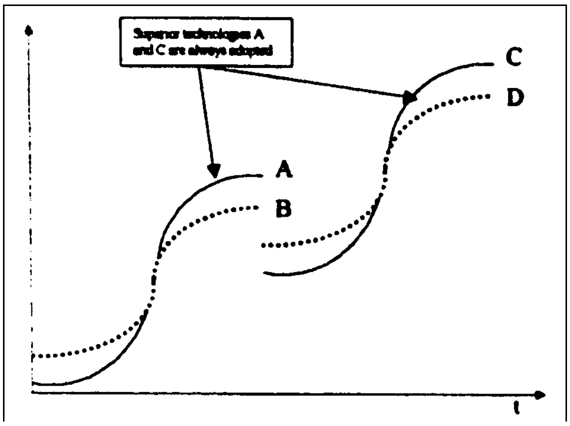

Figure 2. Uncertain and Inertial Technological Trajectory: superior technologies A and C are always adopted

On the other hand, the inertial forces unleashed by commercial success are a lot more powerful than the classical models of diffusion suggests. Abernathy and Utterback (1978), Abernathy and Clark (1985), Teece (1986) and Hughes (1994), among others, have rightly underlined that in addition to the rigidities which may affect individual workers or machines and individual intermediaries or users, there are *systemic rigidities* of a much greater scope and importance. Every successful generic technology has a complex web of complementary technologies woven around its core. Once such an integrated and expensive, in terms of purchasing and using, technological system is in place, its momentum becomes enormous. Consequently, once a specific new technology becomes part of a dominant system, it will become increasingly difficult to dislodge, even by more worthy alternatives. This is depicted in Figure 2 where technology B (the inferior one) is foreclosed by the entrenchment of technology A (the superior one), which will only be displaced by the much superior technology C.

Uncertainty and inertia can combine to cause decisive first-mover advantages, which may grant an unassailable market dominance to an early technological trajectory. And yet, these authors do not go far enough in recognizing that in the real world, optimal technological cycles, trajectories and discontinuities, such as options A and C in Figure 2, are not inexorable realities. They largely ignore and or minimize the self-reinforcing (and not simply inertial) nature of commercial success and the consequent unpredictability of technological evolution in general and of "dominant designs" in particular. With this, they also overlook that a technology's success is tributary to the competitive decisions (often arbitrary and myopic) of the major players in an industry, as they are to any set of exogenous technical parameters.

Abernathy and Utterback did not believe in the research lab as an optimal selector of new technologies and they did question the optimality of selection by the market, but only to a small degree. The early articles written together by Abernathy and Utterback are thoroughly ergodic markets-oriented (Abernathy and Utterback, 1978). Nowhere in them is it hinted that *a dominant design* might not be optimal or that its lasting power might not be inevitable. Later on, Abernathy and Clark (1985) made a strong case for contingency in the maturity and decline of technologies and industries. They thus rejected in no uncertain terms the deterministic view of technological life cycles. Paradoxically, however, their emphasis on historical contingency did not extend to the emergent phase of a new generic technology. They continued to suggest or imply that within a given generic technology, a specific "dominant design" will be chosen strictly on the basis of its relative merit. As for Utterback, in his more recent writings (Utterback, 1993; Utterback and Suarez, 1994), he fully acknowledges that indeterminacy characterizes both the emergence and the decline of a generic technology.

Our observations on technological competition have shown that markets, in the presence of *increasing returns to adoption*, tend to become very unstable and tipping -i.e., to discriminate sharply between winners and losers-, often on the basis of minimal, perhaps almost random, market share differences among the various offers and regardless the relative merit and potential of a new technology. From two comparable competing technologies A and B (see Figure 3) in a market characterized by unbounded increased returns to adoption, only one will win the race for dominance (lock-in). But, *a priori,* it is hard to determine which technology will tip the market (indeterminacy). Furthermore, it is not always sure that the market will select the superior option (sub-optimality). Then, the market is not necessarily an optimal selector of optimal technologies. Another aspect the work on technological competition under increasing returns has shed light on is related to the implications of market instability for the management of risk.

Ergodic models tacitly assume that new technologies are cheap to develop, hard to improve, stand alone, easy to appraise, easy to use, and strongly protected through patents. This implies technologies arrive at the market full developed before diffusion and that the process of development and the diffusion itself can be separated from each other. Ergodic models, then, introduce a very limited level of uncertainty.

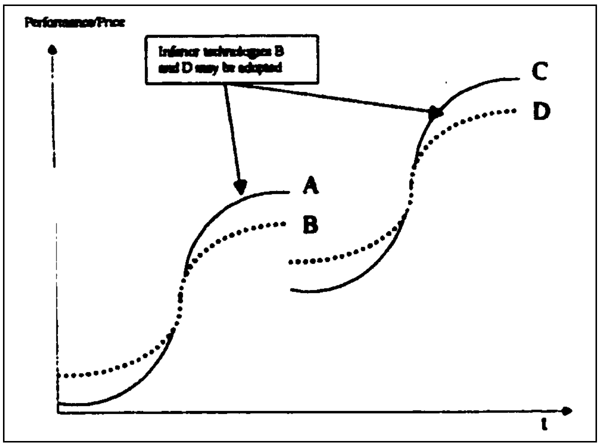

Figure 3 Indeterminate Technological Trajectory: Inferior Technologies B and D may be adopted

In the models of indeterminate technological change, in contrast, new technologies are expensive to develop, subject to further improvements, often due to systemic nature, difficult to appraise and use and with weak patent protection. Therefore, their adoption become self-reinforcing not only for the *reputation effect* of market success, but also for the significant improvements that technologies accumulate during their spreading As the *diffusion process confers value* to technologies and not only conspicuousness, a technology that initially did not deserve being chosen may end up meriting it. Consequently, according to Foray (1989) , the market becomes not only a selector of adequate technologies, but a creator of dominant, superior ones. Under these circumstances, technological sponsors do not face an information problem. but an indeterminate scenario, as Arthur (1994) has indicated. Thus the most effective way to manage totally contingent and unpredictable results in unstable environments is to invest aggressively in market share as the market takes off.

6 Standards and Increasing Returns

Early models of technology diffusion explicitly recognized the general notions of adoption externalities and self-reinforcing dynamics. The epidemic model specified that the diffusion of a new technology, much as that of a potent virus, would be a self-reinforcing process since every user of the respective technology would turn into one of its ardent promoters. What these models, however, failed to recognize was that diffusion can increase considerably the value , and not simply the reputation, of the new technology. In other words, the early models were too focused on the

information externalities and neglected to take into account another source of self-reinforcing in diffusion, namely, *increasing returns to adoption.*

In recent years, technological competition and the emergence of a technological monopoly over a whole market has been the privileged topic of the literature on standards (David and Greenstein, 1990). This literature has identified three processes by which technological standardization can be attained through (a) government regulation through mandated standards, (b) voluntary agreements through formal or standardization committees, and (c) market competition. The first two processes, often called *de jure* standardization, usually result in a standard with public-good characteristics. Standards selected through the market -*de facto* standards-, on the other hand, are usually owned by a firm, which can therefore exclude other firms from its use. Every standardization process has its own theoretical interest, but here we will focus exclusively on *de facto* standardization, which is identified with the economic literature on technological competition under increasing returns.

7. Network Externalities

At the basis of what we know about technological competition is the literature on network externalities (Katz and Shapiro, 1992,1994; Economides, 1996) in which market size relates to increasing returns, and benefits grow with the size of competing networks.

Just as scale economies, learning, or reputation effects, positive network externalities are a self-reinforcing diffusion dynamic. Network effects, however, differ from the other self-reinforcing mechanisms in several important respects: First, while the benefits of scale economies, learning and (some) reputation effects can only be reaped at the time of purchasing the product in question, most of the benefits accruing from network externalities can be enjoyed well past the point of purchase and throughout the entire life cycle of the product. Second, network effects are considerably *forward looking* and *less bounded* and therefore more powerful than scale and learning effects. In fact, because they cast a shadow into the future, network effects can combine with reputation effects to create extremely powerful and lasting self-reinforcing dynamics in market success. Since most of the benefits accruing from network externalities can be enjoyed throughout the full life cycle of a product, new users faced with a multiplicity of competing technical options will have to make their choices not simply in function of what the majority of post purchasers have already chosen, but also in function of what the majority of *inter-face* users are likely to choose. Interestingly, while very pessimistic user expectations about an overall emerging market can be self-dampening, optimistic expectations about the success of a specific technical format in a battle of standards could easily become *self-fulfilling.* The greater the number of people who think that a given technical option is likely to become dominant, the more new users will side with it and the more likely and pronounced the dominance will in fact become. Third, while scale economies and learning can only be a source of increasing returns to adoption and while users' learning costs (or switching costs) can exclusively be a source of inertia, both reputation effects and network externalities, in contrast, can act as both strong inertial and increasing returns to adoption.

8. A Summary on Increasing Returns in Industrial Competition

The analysis so far done has gradually recognized the central role of increasing returns mechanisms in generating and sustaining dominant firms and technologies. Clearly, while scale economies, the resource-based loop, the Schumpeterian (innovation-based) loop, reputation, and the different categories of learning help to explain some of the most basic occurrences of dominant firms and slanted industrial structures, only increasing returns to adoption – a notion intrinsically connected to Schumpeterian economics - can explain most of the instances of technological dominance which we see in contemporary high-technology markets. Thus the integration of all these increasing returns mechanisms, as Figure 10.1 shows, results in a quite complete explanation of industrial competition.

Such an explanation combines the self-reinforcing loops based upon resources and innovations (loops A and B), scale economies, learning, and reputation (loops C) with the loops based upon increasing returns to adoption. These loops are of two kinds. A set of further loops, composed of a mesh of scale economies, learning, reputation effects, infrastructure effects, and network effects, links increasing competitive advantages with increasing returns to market share. A last set of loops indicates that, if increasing returns to adoption are present and considerable, market share becomes a strategic asset well worth investing on in an aggressive manner through vigorous production capacity expansion, price reductions, infrastructure development, and alliances with manufacturer of complementary technologies, products, and services.

As a model of increasing returns mechanism in industrial dynamics, the one described here seems to be rather comprehensive. It can explain the polarized outcomes that are common in most industrial sector and describe business competition as a dynamic and cumulative process. A final, but not less interesting, feature of this model of industrial competition is that it is general, in the sense that it is capable to describe simultaneously the three levels of industrial competition. It elucidates how technological adoption, the number and growth of firms, and industrial structure combine and cause each other. In other words, the general model of industrial competition gives a picture of how industrial competition is a process in which technological competition affects the size of the firms competition in a given industry and how the growth of the firms, in turn, influences the structure of that industry. That is why technological adoption goes between conduct and performance in the chain of causality that leads from the size of the firms to industrial structure.

9. Increasing Returns, Structural Change and Development Paths: An Example

The seminal article formalising the "big push" theory of industrialisation is that of Murphy, Shleifer and Vishney (1989). In their model, firms choose between a constant returns and an increasing returns of technology based on their expectations of demand. However, these choices spill over into aggregate demand creating a strategic interaction among sectors in their technology adoption decisions. Thus, under certain conditions, there exist two equilibria: with all firms choosing the constant returns or all choosing the increasing returns technology. Clearly, in the latter equilibrium, all households are better off.

While the Murphy, Shleifer and Vishny model shows how increasing returns (and a wage effect) aggregate to strategic complementarity among sectors, it does not lend itself readily to the debate concerning the degree of balance in industrialisation policy. First, the static content leaves open the question of whether the intervention should take the form of anything more than indicative planning. Second, the most commonly discussed policy instrument in the industrialization debate is the subsidisation of investments. However, in the Murphy, Shleifer and Vishny example, use of this instrument biases one toward a more unbalanced policy. To see this, observe that it is the role of the government to facilitate a move to the industrialising equilibrium. This means that the government must subsidise a sufficient amount of investment to make it profitable for all sectors to adopt the modern technology. Given the binary choice set, there then exists some minimum critical mass of sectors that must be targeted to achieve a successful transition. A greater range of successful industrialisation policies might be more plausible, however, if firms had the choice of a wider variety of technology to choose from. One might suppose that targeting a large number of sectors to modernise a little and targeting a small number of sectors for more radical modernisation might both generate a big push. Thus, to consider the balanced approach properly, a greater technological choice space is required.

What would be the choice variables available to the government provided it would be able to pick up what is likely to be increaing returns industries in the future? First, in each period, the government can choose the set of firms that it targets for structural change. Second, for each targeted firm , the government can choose a target level for 'increasing returns industry' modernisation in the period. Along this vein, the government could choose to target the same number of firms in each period but induce those firms to modernise gradually over time. Or in contrast, the government chooses a single level of modernisation to occur across all firms and all periods. It then targets a mass of firms each period for entry and modernisation. This means that industrialisation policy is solely characterised by the critical mass of sectors targeted , and the target level of modernisation. Given a parametrised development path , the most significant parameter represents the strength of increasing returns in the technology adopted by industrial sectors which generates a rationale for 'big push ' intervention.

10. Conclusions: Increasing Return Mechanisms and Technological Competition

The strategic importance of increasing returns to technology adoption is unquestionable. In a strictly ergodic market technological options eventually obtain the share of the market they deserve in proportion to their value and technical possibilities. In non-ergodic situations, in a tipping market, on the other hand, the winner takes all or most and the losers (no matter how worthy and how many of them there are) loses all or much. Because of this , the introduction of factors causing tipping markets determines the outcomes of technological competition. This framework not only captures the interplay of institutional arrangements, resources and network of firms and industries in industrial competition, but also delineates very concrete regularities, which can provide us with a simple but powerful tool to explain endogeneously, and in a dynamic way the firm's growth, industrial structure , and technological competition.

This chapter suggests a Schumpeterian model of industrial competition. In contrast to the approaches that underline a specific aspect of industrial competition and/or base their explanation on a reduced number of factors and against conventional economics that overlooks increasing returns mechanisms, this model links the increasing returns mechanisms that determine endogenously inter-firm asymmetries and the kind of industrial structure which emerges during this competition process. This framework also emphasizes the fact that the emergence of dominant firms and the evolution of industrial structure are strongly intertwined with the process of technological change and diffusion.

One of the most important contributions of the work on increasing returns is its having shown that the emergence and persistence of technological monopolies is not an exogenous datum, largely determined by scientific and technical parameters, but is strongly influenced by strong market forces stemming from self-reinforcing mechanisms. In the presence of strong, global and long lasting increasing returns, the actions and omissions of the main actors in the industry in question affect considerably the final result of technological competition. To the extent to which these actors are capable of fully perceiving and exploiting strong increasing returns in emerging markets. they can ensure the entrenchment of their technology as the industrial standard by investing in those strategies that bring about market share. Once entrenched, and to the extent they are capable to exploit inertial forces, established firms can ensure the persistence of their technologies well beyond the time warranted by their relative technical value.

This chapter showed that technological monopoly should be attributed to increasing returns in general, but it is network externalities in particular what has caused some important markets to be dominated by a technological monopoly. In fact, at the level of technological competition, the only thing the other sources of increasing returns to adoption do is to exacerbate the implications of network effects, but they do not turn, isolated from network effects, a market tipping. The most obvious and direct reason for technological monopoly is that the components of a given network are compatible and constantly interconnected. The telephone and the fax are examples of networks where physical interconnection and compatibility have led to technological monopoly.

Winner-takes-all markets are associated with cases where there is (often intense) competition in innovative activity but the future market is such that competition in it is, over a reasonable timeframe, not sustainable. Thus firms compete to attain a position of dominance.

Perhaps the most famous example of a winner-takes-all market is that for operating systems for desktop PCs. It is instructive to recognise that this market benefits from massive economies of scale in production protected by IP rights (very low marginal cost of supply compared with very large fixed costs of initial product development) and substantial economies of scale in consumption (due in large part to the network effects associated with the relationships between the operating system market and the related applications software markets).

Identification of such markets is important because it affects the focus of competitive concerns.

Most obviously, if there are strong grounds to believe that a future market is a winner-takes-all market, it is perhaps not appropriate for a competition authority to block a merger or agreement between firms on the basis that this will create a dominant position or lessen competition in this future market. By definition, the nature of the market is such that its existence guarantees that a firm will be dominant on it, at least in the medium term. (This illustrates an important point relevant to wider issues in competition policy: it is typically better to have a situation where a firm is dominant in a relevant market than for that market not to exist at all.) Instead, any intervention must be based on the premise that the merger (agreement) lessens or distorts competition on some other, perhaps related market , or in competition in the innovative activity associated with the winner-takes-all market.

Similarly in dominance cases, if we anticipate that a market is subject to winner-takes-all properties, then it is difficult to establish a case that a firm has abused its dominant position in monopolising this market - the market is naturally prone to monopolisation. Rather, analysis of an alleged abuse of dominance associated with this market should focus on how a dominant position in a related market (perhaps an access market) could be used to distort competition in the innovative activity associated with the winner-takes-all market, or how a dominant position in the winner-takes-all market could be abused to maintain that position - in effect used to distort competition in the innovative activity associated with the future generation of that market.

Indirect network effects may also tilt the market in favor of one of the competing technologies. In the videofilm industry, for instance, because of strong, long lasting, and global network externalities, technological monopoly emerged and product cycle has lasted about 20 years. The strength, duration and scope of increasing returns in the videofilm markets are the direct consequence of particular technical characteristics of the competing VCR technologies. These technologies were quite similar and mature, the usage cycle of their compatible content -videofilms-- is very short, and their potential substitutes have not had large enough advantages so as to replace. them. In the video game industry, in contrast, because of weak, short lasting, and local network externalities due to incomparability of videogame technologies, long usage cycle of videogames, and the proliferation of new, more powerful new formats, different technologies have tended to share the market according to their intrinsic value.

A comparison between the videogame industry and the videofilm industry, then, allowes us to show that compatibility constraints are not a sufficient condition for a virtual network technology to become s monopoly: it is necessary. besides, that usage life of content is short and core technologies are incomparable. But when the usage cycle of content is long and core technologies are comparable, any virtual network technology becomes less systemic. As the strength, duration, and scope of network externalities are reduced considerably because of long-lasting usage cycle and technology comparability, competing virtual network technologies end up sharing the market in proportion to their value.

In this chapter we also advanced our understanding of the technical and institutional factors which are likely to affect the nature of technological competition. In doing so, we add to the dimensions of *strength* and *duration,* the dimension of *scope* of increasing returns. The distinction between strength, duration and scope is useful to realize that, contrary to popular and academic literature, a market for virtual network technologies with content dimensions not necessarily will end up with a technological monopoly. With this distinction in mind and against those who think that strong indirect network externalities always act as tipping mechanisms, we also can show that strong indirect network externalities are compatible with fairly ergodic market dynamics, if the scope of such externalities happens to be rather narrow.

By taking the telecommuncations or the videofilm industry, for example, we can show that strong network externalities are *necessary* but not *sufficient* conditions to produce technological monopoly. Short usage life of content and technology incomparability are technical and necessary conditions for technological monopoly in software intensive virtual network technologies to happen. But these technical aspects of virtual network technologies are not a sufficient condition to produce technological monopolies. In chapter 4 we showed that network externalities require not only to have high levels of strength. but also to be global in scope. Under certain institutional condition strong indirect network externalities may be rather localized, which leads to very ergodic market results. In these conditions markets are shared by the competing technologies according to their intrinsic value.

Technical and institutional factors causing different levels of strength. duration and scope of increasing returns to adoption are relevant to determine whether a market is tipping or ergodic. This has some implications. A first implication is that not all network technology is equally systemic. If there are strong network externalities but with a local scope, the systemic nature of a network becomes rather limited. In this case the systemic nature of the VCR network would be rather local compared with the actual network, which is global in nature.

Strength, duration, and scope of increasing returns are also useful to determine in a more detailed way the nature of *cooperation.* When network externalities are strong and global, content-intensive virtual network technologies become rather systemic. In these circumstances, the main sponsors of the competing technologies may produce some components of the system. but the rest of it may be out of their reach. For instance. a PCs producer may be incapable to produce software or microprocessors; and microprocessor producers may not be able to produce software nor hardware. VCR producers cannot produce films, in the same way as film producers cannot produce VCRs. Consequently, technological competition in markets characterized by strong and global increasing returns is more in connection with complex networks of firms than with conventional industrial array of firms producing homogeneous products. In contrast, in markets with weak and local network externalities, competition takes place mainly between firms than between networks of firms. This is so, because in this kind of markets products are not systemic.

References

Abernathy, William J. and James M. Utterback (1978), "Patterns of Innovation in Technology". *Technology Review* 10, June-July, 40-47.

Abernathy. William J. and Kim B. Clark (1985). "Innovation: Mapping the Wind of Creative Destruction, *Research Policy* 14, 3-22.

Aghion, E. and E. Howitt (1998), Endogeneous Growth Theory. Cambridge, Ma.: MIT Press

Arrow, Kenneth (1962), "The Economic Implications of Learning by Doing". *Review of Economic Studies* vol. 29, 155-173.

.Arthur, W. Brian (1989), "Competing Technologies, Increasing Returns and Lock-in by Historical Events". *The Economic Journal.*, Vol. 99. March, 116-131.

Arthur, W. Brian(1 990a), "Positive Feedbacks in the Economy" *Scientific American*, February 22, 92-99.

Arthur, W. Brian (1990b) "'Silicon Valley' Location Cluster. When Do Increasing Returns Imply Monopoly" *,Mathematical Social Science*. 19, 235-51.

Arthur. W. Brian. (1994) *,Increasing Returns and Path Dependence in the Economy.*

Ann Arbor: The Univ. of Michigan Press

Arthur. W. Brian, *Yu. M.* Ermoliev, and Yu. M. Kaniovski (1987), "Path-dependent Processes and the Emergence of Macro-structure". *European Journal of Operational Research.* 30, 294-303.

Arthur, W. Brian and David A. Lane, 1993, "Information Contagion". *Structural Change and Economic Dynamics 4,* 81-104.

Chandler, Alfred D.(1990), Scale and Scope: Dynamics of Industrial Capitalism. Cambridge. Ma.: The Belknap Press of Harvard University Press

.

Cohen, Wesley M. and Richard C. Levin. (1989). "Empirical Studies of Innovation and Market Structure". in Richard Schmalensee and Robert D. Willig. eds., Handbook of Industrial Organization. Amsterdam: North-Holland

David Paul and Shane Greenstein (1990), "The Economics of Compatibility Standards: an Introduction to Recent Research", *Economics of Innovation and New Technology 1. pp. 3-41.*

Economides. Nicholas (1996), "The Economics of Networks". *International Journal of Industrial Organization,25, 1-18*

Foray, D. (1993),"Standardisation et Concurrence: Des Relations Ambivalentes", Rev. d'Economie Industrielle, 63, 30-45

Freeman, Chris (2003) ,'A Schumpeterian Renaissance ? ',Science and Technology Policy Research Unit, Univ. of Sussex, Brighton, SEWPS #102, July

Gottinger, H.W., Economies of Network Industries (2003), London :Routledge

Henderson, Bruce D. (1975), "The Market Share Paradox", in Y. Brozen, ed. *The Competitive Economy*, Morristown:General Learning Press, pp.286-287

Hirsch, Werner Z. (1956), "Firm Progress Ratios", *Econometrica* 24, 136-143

Katz, Michael and Carl Shapiro (1992), "Product Introduction with Network Externalities", *The Journal of Industrial Economics* 40, 55-84

Katz, Michael and Carl Shapiro (1994), "System Competition and Network Effects", *Journal of Economic Perspectives* 8, pp.93-115

Murphy, K.M., A. Shleifer, and R.W. Vishny (1989), "Industrialization and the Big Push", Journal of Political Economy 97, pp.1003-1026

Nelson,Richard and Winter, Sidney G. (1982), *An Evolutionary Theory of Economic Change*, Cambridge, Ma.: Harvard Univ. Press

Nelson, Richard (1986), "Evolutionary Modeling of Economic Change", in Stiglitz, Joseph and G. Frank Matthewson, eds., *New Developments in the Analysis of Market Structure*, Cambridge, Ma.: MIT Press

Porter, Michael E. (1980), *Competitive Strategy*, New York: Free Press

Rosenberg, Nathan (1982), *Inside the Black Box: Technology and Economics*, Cambridge :Cambridge Univ. Press

Schumpeter. Joseph A.. (1942), *Capitalism. Socialism. and Democracy,* New York:Harper

Shapiro, Carl and Hal R. Varian (1999), *Information Rules,* Boston: Harvard Business School

Teece,D.(1986),'Profiting from Technology Innovations', *Research Policy* 15(6), 285-305

Utterback,J.M.(1994), *Mastering the Dynamics of Innovation*, Boston; Harvard Business School Press

Utterback,J.M. and Suarez,F.F. (1993), Innovation, Competition and Industry Structure, *Research Policy* 22, 1-21