# Why Does Technology Advance in Cycles?

Ola Olsson*

Department of Economics,
Göteborg University

### Abstract

Long-run technological progress is cyclical because drastic innovations that introduce new technological opportunity are only profitable at times when repeated incremental innovation has nearly exhausted existing technological opportunity and driven entrepreneurial profit and income growth towards zero. The article presents a 'technological opportunity model' where endogenous drastic and incremental innovations interact with exogenous discoveries in an idealized metric technology space. New ideas are created by convex combinations of existing ideas. Diminishing technological opportunity results in lower profits and growth, which then makes costly and risky drastic innovations profitable again. This relationship between intense drastic innovation intensity and poor levels of economic growth receives some empirical support.

Keywords: technology, growth, long waves, cycles, technological paradigms, innovations.

JEL Codes: E37, O11, O31.

## 1 Introduction

Long-run technological progress has hardly ever been a smooth process. Periods of revolutionary advance have been succeeded by periods of stagnation. During just a short span of years around 1880, innovations such

*Adress: Vasagatan 1, P.O. Box 640, 450 30 Göteborg, Sweden. E-mail: Ola.Olsson@economics.gu.se. This research has been granted generous financial support from Jan Wallander and Tom Hedelius' Stiftelse. I am grateful for helpful comments from Marcus Asplund, Arne Bigsten, Wlodek Bursztyn, Henrik Hammar, Lennart Hjalmarsson, Susanna Lundström and from seminar participants at Göteborg University.

as electric light, the automobile, the telephone, the phonograph, the fountain pen, and the bicycle appeared in a burst of technological activity (Gordon, 2000). Another revolutionary wave of innovations took place in the 1930s when television, radar, synthetic rubber, colour photography, and FM radio were introduced (van Duijn, 1983). In contrast, not more than a handful major innovations were made during the two and a half centuries between 1500-1750 (Mokyr, 1990) and surprisingly few major innovations can be associated with the 1960s and 1970s. In the 1990s, it was once again felt that innovations within IT- and biotechnology were about to revolutionize production and possibly even lead to a 'new economy'. Why does technological progress proceed in this cyclical fashion? Considering the significant impact that technological knowledge is believed to have on standards of living and on economic development in general, it appears that a better understanding of the reasons behind the cyclical evolution of technology would be very useful.

The main idea behind this article is to put *technological opportunity* at the center of analysis. I develop a simple model where rational entrepreneurs exploit existing technological opportunity by making incremental, non-revolutionary innovations within the limits set by the prevailing technological paradigm. As technological opportunity becomes exhausted, profits and income growth rates diminish. Eventually, profits from incremental innovation fall below expected profits from highly risky and costly drastic innovations. Entrepreneurs then switch to drastic innovation, which introduces new areas of technological opportunity and a new technological paradigm. When technological opportunity once again is abundant, incremental innovation resumes and growth rates increase. In this way, development proceeds in long waves of varying duration and intensity. The fundamental determinants of the economy's behaviour are the capacity of a society to exploit existing technological opportunity and its system of rewards for drastic innovation.

Similar notions of long run development have been advanced by researchers in the 'long wave-tradition', notably Schumpeter (1934, 1942), Mensch (1979), Freeman (1983), van Duijn (1983), and Kleinknecht (1983). Schumpeter (1934, 1942) famously argued that discontinuous 'swarms of "new combinations"' (innovations), carried out by entrepreneurs, disrupted the 'circular flow' of the economy and induced a cyclical pattern of development that was characterized by 'creative destruction' of existing firms and monopoly profits. Schumpeter's theory was developed theoretically as well as empirically by Mensch (1979). In his 'depression-trigger' theory of development, Mensch argues that waves of drastic innovations tend to appear in depressions when investment inertia and lack of technological opportunity drive profits to zero. On the basis of a compilation of 111 'basic innovations' made between 1796 and 1955, Mensch shows that clusters of innovations indeed did appear around 1880 and 1930 in the wake of major downturns in the economy. Later works have refined Mensch's list of innovations (Clark et al, 1983; van Duijn, 1983; Kleinknecht, 1983), but this has not altered the basic pattern of two long waves that Mensch demonstrated.

Both Schumpeter and Mensch associate slowdowns in economic activity with the failure of firms to come up with new, drastic innovations. However, the somewhat vague and unformalized notions of the reasons behind slowdowns in innovation and economic activity, are probably the weakest links in the Schumpeter/Mensch-hypothesis.[1] The model in this article proposes that it is waning technological opportunity rather than a change in firms' or consumers' behaviour that cause declining profits and growth.[2] Following Olsson (2000), technological opportunity is explicitly modelled at the micro level as the nonconvex segments of a *technological knowledge set,* embedded in a multidimensional *technology space* of all possible ideas. Incremental innovations take the shape of linear, binary combinations of closely related ideas already in the technology set. When repeated combinations have made the technology set convex, technological opportunity is exhausted. Only a *technological paradigm shift* - i.e. the occurrence of a cluster of drastic innovations that combine ideas in the technology set with distant and isolated *discoveries* - has the potential to reintroduce nonconvexities and technological opportunity.

The model presented in this article is related to several previous theoretical contributions. In a model on long-run 'Schumpeterian cycles', Jovanovic and Rob (1990) distinguish between 'extensive search' for new inventions or major discoveries on the one hand, and 'intensive search' for refinements of a given invention on the other. It is shown that extensive search will only be carried out when the expected payoffs from this activity exceed expected payoffs from intensive search. Given specific parameter values for the level of technological opportunity and the costs of extensive search, a cyclical behaviour emerges where innovators switch repeatedly between search activities. Boldrin and Levine (2001) construct a similar model where agents switch from working on new 'generations' of capital (incremental innovations) to working on new 'types' of capital (drastic innovations) when all possible generations of

[1] In a famous review of one of Schumpeter's works, Kuznets (1940) criticizes the Schumpeterian model along the same lines.

[2] Caballero and Jaffe (1993) and Kortum (1997) provide empirical evidence suggesting that technological opportunity has fallen steadily since the 1950s. As will be shown below, this observation is well in line with the results of this model.

a certain type of capital have been found. A primary difference between these models and the one worked out below is that unlike in Jovanovic and Rob (1990) and Boldrin and Levine (2001), technological progress is seen as the result of convex combinations of existing ideas in a technology space. Furthermore, technological opportunity is endogenously determined and varies over time.

The article is also associated with the literature on *General Purpose Technologies* (GPTs) (David, 1990; Bresnahan and Trajtenberg, 1995; Aghion and Howitt, 1998; Helpman, 1998; Caselli, 1999). A GPT is essentially what this article defines as a drastic innovation. The date of arrival of a new GPT is often assumed to be exogenously given (Helpman and Trajtenberg, 1998) or to follow a random Poisson process (Aghion and Howitt, 1998). In addition, Aghion and Howitt (1998), Helpman and Trajtenberg (1998), and Boldrin and Levine (2001) all share the prediction that the onset of the new major technology causes a slowdown in the economy because of the necessary reallocation of resources from the old to the new sectors. Only after a certain critical level of adjustment activities, the new technology is fully implemented and a period of boom starts. This hypothesis that drastic technological change causes the economic downturn, stands in contrast to the 'depression-trigger' model of Mensch where it is the downturn that causes the wave of technological change.

The full model in this article, based on the insights from the set theoretic framework, makes the assumption that it is in times of economic depression when technological opportunity nears exhaustion that the wave of drastic innovations will occur. Unlike in the GPT models, the technological revolution is endogenously determined and happens when expected profits from incremental innovation fall below expected profits from drastic innovation. The actual profits that entrepreneurs make from drastic innovations in the short term are random and might deepen or ease the depressionary tendencies. However, the paradigm shift reintroduces technological opportunity so that entrepreneurs in the next period once again can make large profits from incremental innovation.

Simulations of the model suggest several different scenarios in terms of growth rates, cycle duration, and final output levels, depending on how the model is calibrated. In one scenario, long waves of about fifty years in duration is produced which seems to fit the general picture of economic development in the Western world during the last two hundred years. Comparing three indice on drastic innovative activity with average growth rates per capita for the United States, the UK and Germany for the period 1870-1968, it is demonstrated that the model roughly makes the correct predictions about the sequencing of events and the timing of the two clusters of drastic innovations around 1880 and 1930.

The article is organized as follows: Section 2 discusses the two basic premises of the article; that there are three kinds of technological progress and that innovation activity is asymmetrically distributed in time. Section 3 then outlines the set theoretic analysis of the interactions between drastic innovations, incremental innovations, and discoveries in a multidimensional technology space. Section 4 describes the details of the full model when technological opportunity is integrated into an economic growth framework. Section 5 presents simulations of the model and discusses the results in the light of long run estimates of GDP growth rates per capita for three leading economies. Section 6 analyzes the differences in long run growth predictions between the neoclassical model, the $AK$-model, and the technological opportunity model. The last section gives the concluding remarks.

## 2 Two Basic Premises

The technological opportunity model below rests on two basic premises: (i) That there are three kinds of technological progress. (ii) That innovative activity follows a cyclical pattern.

### 2.1 Three Kinds of Technological Progress

A fundamental premise for the argumentation below is that there are essentially three kinds of technological advance: (i) *Incremental innovations*, (ii) *drastic innovations*, and (iii) *discoveries*.

Incremental innovations are small, non-revolutionary changes in technology that are carried out by profit-oriented entrepreneurs.[3] This kind of innovations refine existing knowledge in a predictable fashion and are generated when entrepreneurs combine older insights that are in technological proximity. The costs and risks associated with this activity are relatively low. Incremental innovations are also highly path dependent, following specific technological trajectories (Dosi, 1988). The boundaries to incremental innovation are set by the prevailing technological paradigm which defines the opportunities of technological research at some point in time. Incremental innovation is the normal activity for a progressive entrepreneur.

As the label suggests, drastic innovations are radically new ideas that are reached after deliberate efforts at combining previously unrelated

[3] The definition of incremental innovations used here follow Helpman (1998), but there are several related concepts in the literature such as 'microinventions' (Mokyr, 1990) 'refinements' (Jovanovic and Rob, 1990) and 'secondary innovations' (Aghion and Howitt, 1998).

ideas. Therefore, they have no obvious antecedents. Drastic innovations are by nature associated with high risks and costs, but usually also with the possibility of significant financial rewards. As with incremental innovation, the entrepreneur is the main agent in this process and his sole motive is commercial success. The most important characteristic of a drastic innovation is that it opens up new fields of technological opportunity. The concept is therefore closely related to what Bresnahan and Trajtenberg (1995) and Helpman (1998) refer to as a 'General Purpose Technology' (GPT).[4] A drastic innovation is an 'enabling technology' in the sense that it makes possible the evolution of a whole new class of new technologies. Using the language of Kuhn (1962) and Dosi (1988), a cluster of drastic innovations gives rise to a new technological paradigm, a new outlook on the relevant problems and on the ways to solve them. When the new paradigm is generally accepted, a period of normal incremental innovations resumes along the trajectories defined by the new paradigm.

The third kind of technological advance, discovery, has some similarities to a drastic innovation. Like a drastic innovation, a discovery is a completely new piece of knowledge that is markedly different from the predictable results of incremental innovation. It does not fit into the prevailing pattern of established facts and is therefore, initially, looked upon as something of an anomaly to the ruling technological paradigm. However, the new idea is not the result of purposeful entrepreneurial combination processes. My definition of a discovery is meant to describe the accidental empirical finding of some phenomenon, often while conducting normal science with the aim of finding something completely different. Thus, it is not necessarily entrepreneurs that make the discovery. It might just as well be a scientist or a layman who happen to stumble upon something previously unknown.

The most important difference between a discovery and a drastic innovation is that the former is not an enabling technology or a GPT since it does not immediately introduce new fields of technological opportunity. Nor does it have any immediate commercial potential. For longer or shorter periods of time, discoveries might remain as separate islands outside the prevailing paradigm. Not until some entrepreneur deliberately makes a successful drastic combination between the ideas inherent in the older paradigm and the new discovery - i.e. a drastic innovation - technological opportunity increases and a new technological paradigm is eventually introduced.

[4]The phenomenon has also been described as a 'new combination' (Schumpeter, 1934), 'basic innovation' (Mensch, 1979), 'macroinvention' (Mokyr, 1990) and as 'fundamental innovation' (Aghion and Howitt, 1998).

In order to make the argumentation clearer, I will offer some historical examples. Beginning with incremental innovations, it should be noted that these make up the vast majority of innovations in all fields. Since the drastic innovation of automobiles driven by internal combustion engines was carried out in the 1880s by Karl Benz and Gottlieb Daimler, repeated incremental innovations have led to vast improvements. The first car that Benz built in 1885 was a three-wheeler, powered by a two-cycle engine, and probably much slower than a horse. The contrast to a modern car in terms of speed, energy use, reliability, and comfort is of course striking. Yet all these improvements had the character of incremental innovations, made possible by the drastic innovation of internal combustion engines. Specific incremental innovations are not often referred to in books on the history of technology but are nonetheless of paramount importance for technological and economic development.

The history of drastic innovations, on the other hand, is well-known and documented in numerous academic works. In this category are for instance such enormously important innovations as the wheel, writing, sedentary agriculture, the spinning wheel, the clock, the electric light bulb, the internal combustion engine, nuclear power, and the computer. The most famous innovation of all times is probably the *steam engine*. The development of the steam engine is a typical example of how major technological change proceeds and it also illustrates the close interdependence between discoveries and drastic innovations.[5]

In the seventeenth century, Galileo's student Evangelista Torricelli became the first man to create a sustained vacuum, which led to the realization that there existed an atmosphere. This remarkable insight was gained as a biproduct of trying to construct a more efficient vacuum pump. At first, no immediate use of this new empirical finding was found. Nor was it easily reconciled with the prevailing technological paradigm. However, the discovery of atmospheric pressure was picked up decades later by Thomas Newcomen who in 1712 constructed an atmospheric engine that was successfully used for pumping water out of coal mines. Finally, in 1769, James Watt claimed a patent for a steam engine based on the same principles as Newcomen's engine but with a separate condenser which greatly improved efficiency. The steam engine opened up technological opportunities in several fields and soon revolutionized transportation, mining, and the textiles industry. Yet it must be remembered that a prerequisite for Newcomen's and Watts' drastic innovations was the unintentional discovery of atmospheric pressure.

Examples of discoveries, serving as the main stimulus to subsequent

[5]The exposition on steam engines relies on Mokyr (1990).

technological progress, are indeed manifold. For instance, we can confidently assume that the human use of *fire* was initiated after innumerable horror-stricken but gradually more curious observations of the devastating effects of fires ignited by lightning. Once tamed, fire revolutionized prehistorical technology for cooking, housing, hunting, and warfare. Most *metals* and other natural materials that humans have learnt to use through history are also discoveries in the right sense of the word. Only long after the discovery of iron did humans learn how to process iron into more useable forms. Some of the most important drastic innovations of all times - like the telegraph, the electric motor, and the electric light bulb - could be made in the wake of the scientific discovery of *electricity* in the seventeenth and eighteenth centuries. In the beginning of the nineteenth century, however, it was not at all obvious to entrepreneurs or scientists what electricity could be used for (Lipsey et al, 1998). Lastly, Alexander Fleming's accidental discovery in 1928 of the dramatic effect of moulds on bacteria - paving the way for the drastic innovation *penicillin* - is probably one of the most clearcut cases of an unintentional discovery with pervasive effects on technology and human welfare.

## 2.2 Cyclical Innovative Activity

A second fundamental premise in this article is that the growth of technological knowledge is asymmetric in time. This notion is certainly not new. In older research on long waves, the unevenness of technological advance was seen as an important reason for the cyclic long-run behaviour of output and productivity.[6] Schumpeter (1934, p 223) argues that 'new combinations', i.e. drastic innovations carried out by entrepreneurs, '...are not, as one would expect according to general principles of probability, evenly distributed through time...but appear, if at all, discontinuously in groups or swarms.' This swarm-like appearance of drastic innovations and entrepreneurial activity occurs, 'Exclusively because the appearance of one or a few entrepreneurs facilitates the appearance of others, and these the appearance of more, in ever increasing number.' (Schumpeter, 1934, 228).

Based on empirical evidence on the timing of drastic innovations (see below), Mensch developed his 'depression-trigger' or 'metamorphosis' model where the long-run evolution of the economy is described in terms of logistic industrial development curves. Each industrial cycle is initiated by a cluster of drastic innovations. When the cycle has reached its maturity level where the curve flattens out, further incremental innovations fail to improve profit rates and stagnation sets in. Only during

[6] For overviews on long-run innovation theories, see van Duijn (1983) or Lipsey et al (1998).

such times are investors willing to accept the higher risks of drastic innovations. Stagnations or depressions thus evolve solely because of lack of drastic innovations, but they also trigger the appearance of a new cluster of innovations and a new cycle.

Aside from the long-wave tradition, the cyclical character of knowledge growth has been discussed also in other reseach areas. In one of the most influential works in the philosophy of science, Thomas Kuhn (1962) argues that the natural sciences have evolved in a process where normal science, characterized by incremental progress and 'puzzle solving-activities', is followed by a period of doubt when a pile of anomalies make researchers question the existing scientific paradigm. Eventually, a scientific revolution occurs and a new paradigm that better fits the accumulated facts is introduced.

In their works on economic history, Landes (1969) and Mokyr (1990) provide several examples of how periods of revolutionary change have been followed by periods of much slower innovative activity. In the productivity literature, Gordon (2000) argues that the many electricity-related innovation in the closing decades of the nineteenth century initiated a powerful wave of productivity growth that lasted even beyond World War II. The idea that assymmetric technological change might cause economic cycles has been picked up by the Real Business Cycle-literature where random technology shocks are seen as an important source of short-term variations in output (Kydland and Prescott, 1982). Jovanovic and Rob (1990) model 'Schumpeterian cycles' or 'long waves' of technological change. Bresnahan and Trajtenberg (1995) and Helpman (1998) coined the phrase General Purpose Technologies to analyze the longer term effects of drastic innovations in a growth theoretic framework.

What empirical evidence are there of cyclical technological change? In particular, do the data support the Schumpeter/Mensch-hypothesis that drastic innovations tend to appear in clusters? The rest of this section will deal with that issue. Mensch (1979, Tables 4.1-4.4) lists 111 'basic innovations' introduced between 1796 and 1955. Both the year of innovation and the year of invention are provided, where the former is the date when, '...a technological basic innovation...is being put into regular production for the first time.' (Mensch, 1979, p 123). The year of invention marks the date of basic discovery of the main principle behind the innovation. As an example, Mensch claims the year of invention of synthetic rubber (Neopren) to be 1906, when J.A. Nieuwland observed the acetylen reaction in alkali medium, and the year of innovation to be 1932 when the firm E.I. du Pont de Nemours and Company introduced Neopren into the market. Mensch's definitions of 'basic innovations' and 'inventions' are therefore very similar to the concepts 'drastic innovations' and 'discovery' developed above. The sample of basic innovations has been distilled mainly from Jewkes et al (1958).

Clark et al (1983) criticize the Mensch sample on the grounds that nearly a dozen of the case studies in Jewkes et al (1958) have not been included, nor have the ten additional cases in the second edition of Jewkes et al (1958) been used. Apart from this, Clark et al disagree with the dating of some innovations and inventions and have suggested alternative datings. When these changes to the Mensch-data are made, a Clark-Freeman-Soete sample (CFS) of 132 drastic innovations emerges.

In a similar fashion, van Duijn (1983, Table 10.1) presents a compilation of 160 drastic innovations made between 1811 and 1971. Van Duijn uses a greater number of sources and the drastic innovations that have been picked out often coincide with those of Mensch and CFS, but not always. Even when all three samples include a certain innovation, the stated years of innovation and invention might differ. For instance, Mensch maintains that the invention of the telephone occurred in 1854 and the innovation in 1881, while van Duijn's corresponding dates are 1860 and 1877.

Naturally, the determination of such dates is a complicated and somewhat arbitrary task. Among several difficulties with this kind of sampling, van Duijn mentions the following: (i) *Innovation heterogeneity in character*. The wide spectrum of different kinds of innovations in totally different fields makes measurement of importance along a single scale very difficult. Objectivity is hard to establish. (ii) *Innovation heterogeneity in impact*. The innovation of the telephone was most likely far more important than the innovation of DDT, yet both are included as drastic innovations. (iii) *Time bias*. The importance of an innovation can usually not be ascertained until many years have passed, often even decades. All else equal, there should thus be a tendency to underestimate the importance of recent innovations. (iv) *Past and present*. The perception of how important innovations are might change with time. DDT, which was mentioned above, would nowadays rather be seen as a menace to the environment than as a welfare-enhancing technological breakthrough.

Despite these difficulties, some clear regularities in the data seem to be robust to most objections. Figure 1 shows the cumulative distribution of drastic innovations and discoveries 1800-1941 according to the Mensch sample. Several things are worth noting. First, the wave-like character of the drastic innovations-curve suggests that this kind of technological change is cyclical. At least, this seems to hold for the second half of the time period. Second, two waves in drastic innovations can be singled out; the first starting prior to 1880, the second taking off around 1930. Third, the figure suggests that the appearance of drastic innovations is more uneven and wave-like than the development of discoveries. Indeed, the curve for discoveries suggests a more or less random distribution of discoveries over time. The same waves can be observed also in the CFS and van Duijn samples. However, the seemingly random distribution of discoveries does not appear when graphing the time distribution of the van Duijn data points.

By calculating a seven-year moving average time series for the number of drastic innovations per year for each of the Mensch, CFS and van Duijn samples, we get a clearer picture of the trends in drastic innovation activity. Figure 2 shows the three resulting curves for the interval 1800-1968. To begin with, it should be recognized that the CFS-curve is an exact mirror image of the Mensch-curve until 1900 since the two samples are in fact identical until then. The notable feature in general is that the three samples give a very similar picture of cyclical behaviour. No great clustering of innovations can be detected before 1870. Both curves then make a dramatic turn upwards at the year 1876. This wave of innovation activity culminates in the middle of the 1880s whereupon there is a sharp decline. A much less distinct peak can possibly be supported around 1910, at least for the CFS and van Duijn samples. The really significant increase does not appear until the end of the 1920s. The CFS and, even more, the van Duijn-curve, then suggest a pervasive boom in R&D that lasts for several years. The Mensch-curve, however, falls sharply already before World War II. As noted above, because of the time bias, one should be careful about overinterpreting the years near the end of the interval.

## 3 The Technological Opportunity Set

The central question in this article is why technological progress is cyclical, i.e. evolving through periods with clusters of drastic innovations, followed by periods of lower activity. In this section, I will use the notion of technological progress through the recombination of existing ideas, suggested by Schumpeter (1934) and Weitzman (1998). The framework for the combinatory process will be based upon Olsson (2000) and assume that the *technology set* and its corresponding *technological opportunity set* are embedded in a metric *technology space*. The analysis will be structured around the two basic premises discussed in the previous section; that there are three kinds of technological progress - incremental innovations, drastic innovations, and discoveries - and that the time distribution of (drastic) innovations is asymmetric.

The context for the model below is that individual ideas are defined in metric technology space $\mathbb{T} \subset \mathbb{R}_+^k$ which is the universal set of all possible technological ideas in the past, in the present, and in the future. Each individual idea $i_n \in \mathbb{T}$ is a $k$-dimensional vector of positive, real-valued coordinates where $k > 1$ is the number of dimensions in human thinking. $k$ thus includes complexity, abstractness, mathematics, utility, etc.

$\mathbb{T}$'s metric is the *technological distance function* $d(i_m, i_n)$ where $i_m, i_n$ are individual ideas contained in technology space.[7] This function measures the objectively determined distance between any two ideas in technology space.[8] The distance function $d(\cdot)$ is meant to convey that ideas are more or less closely related. An idea like 'wheeled transport' is technologically close to the idea 'automobile'. Neither idea, however, has much in common with an idea like 'electric light bulb'. Hence, technological distance between 'wheeled transport' and 'automobile' is shorter than between 'automobile' and 'electric light bulb'. The distance function satisfies the necessary requirements for defining a metric space: (i) $d(i_m, i_n) \geq 0$ and $d(i_m, i_n) = 0$ *iff* $i_m = i_n$, (ii) $d(i_m, i_n) = d(i_n, i_m)$ $\forall$ $i_m, i_n \in \mathbb{T}$, and (iii) $d(i_m, i_n) + d(i_n, i_o) \geq d(i_m, i_o)$.

The technological knowledge set $\mathbb{A}_t \subset \mathbb{T}$ is the set of all known technological ideas at time $t$. It should be seen as the union of all technological ideas in the world rather than as describing any individual country's state of technology. In antiquity, this set included ideas like 'aqueduct', 'horse power', and 'the sail'. In late medieval times, $\mathbb{A}_t$ included 'the wind mill', 'the clock', and 'the printing press.' At the beginning of the twentieth century, $\mathbb{A}_t$ contained ideas like 'steam engine', 'telephone', and 'automobile'. It should be made clear, however, that although all the examples mentioned are drastic innovations, $\mathbb{A}_t$ also includes the numerous accumulated incremental innovations made in the time period $[0, t]$. There is no depreciation of technological knowledge which means that $\mathbb{A}_{t-1} \subseteq \mathbb{A}_t$ at all $t$. A piece of technology that once has been employed for the production of some good or service can always be used again, though with inferior efficiency. The complement of $\mathbb{A}_t$ is $\mathbb{A}_t^C$ which contains all technological principles that have still not been discovered at time $t$.

The technology set has a number of technical properties. To begin with, $\mathbb{A}_t \cup \mathbb{A}_t^C = \mathbb{T}$, i.e. the technology set and its complement is what

[7] The notion of a metric technology space with a corresponding 'distance function' can be found, for instance, in Jaffe (1986), Dosi (1988), and Kauffman et al (2000). See Olsson (2000) for a more thorough treatment of the set theoretic properties of a 'knowledge set'.

[8] It might be argued that technological distance between ideas rather should be modelled as *subjective*, depending maybe on an individual's knowledge and experiences, or that distance between any two given ideas possibly should be allowed to vary over time.

constitutes technology space. Furthermore, $\mathbb{A}_t$ is *infinite*, *closed* and *bounded*. Infinity implies that the number of ideas in $\mathbb{A}_t$ is always infinite and therefore irrelevant for analysis. However, since $\mathbb{A}_t$ is bounded in $\mathbb{T}$, the 'size' of $\mathbb{A}_t$ is finite. The assumption that $\mathbb{A}_t$ is closed implies that it includes all its boundary points. The boundary points of $\mathbb{A}_t$ might be thought of as the technological frontier at time $t$. Finally, I assume that the size of $\mathbb{A}_t$, or any other subset $\mathbb{Z}_n \subset \mathbb{T}$, can be captured by a size function $s(\mathbb{Z}_n) \in \mathbb{R}_+$ which satisfies the properties: (i) $\mathbb{Z}_1 \subset \mathbb{Z}_2 \Longrightarrow s(\mathbb{Z}_1) < s(\mathbb{Z}_2)$, and (ii) $s(\mathbb{Z}_1 \cup \mathbb{Z}_2) = s(\mathbb{Z}_1) + s(\mathbb{Z}_2) - s(\mathbb{Z}_1 \cap \mathbb{Z}_2)$. The size of the technology set at time $t$ will be denoted $s(\mathbb{A}_t) = A_t$.

The technology set and its surrounding technology space is depicted in Figure 3. Without loss of generality, I assume $k = 2$ in order to make the representation simple. Disregarding all the dots and straight lines for a moment, we have initially a set $\mathbb{A}_t$ that apart from all its other characteristics also is *nonconvex*.

Let us now see how the set of technological knowledge evolves. The crucial assumption in this setting is that new ideas in the form of incremental and drastic innovations are formed through linear, binary combinations of existing ideas, undertaken by entrepreneurs, whereas individual discoveries appear independently from other ideas. In Figure 3, incremental innovations are described as the linear combinations of technologically close ideas in the nonconvex parts of $\mathbb{A}_t$ (to the left and right). For instance, combining the two ideas $i_l$ and $i_m$, both contained in $\mathbb{A}_t$, results in the new idea $i_n$ which was not previously included. The new idea's location in technology space becomes $i_n = \lambda_n i_l + (1 - \lambda_n) i_m$ where $0 < \lambda_n < 1$. $i_n$ is then used for subsequent linear combinations, which yield yet other ideas that can be used, and so on. This process is similar to Kuhn's view of how normal science evolves. Nothing radically new results from this piecemeal advance of technology. Progress is continuous and proceeds in a predictable fashion along the expected routes.

It is easy to see that linear combinations of ideas can be made only as long as $\mathbb{A}_t$ is nonconvex. When the limits to the left and right in the figure are reached, the set is convex and incremental innovations are no longer possible. Thus, the sections of $\mathbb{A}_t^C$ that border the nonconvex sections of $\mathbb{A}_t$ might be described as the areas of *technological opportunity*. In more formal terms, we can use the following definition:

**Definition 1** *The technological opportunity set* $\mathbb{B}_t$ *is the smallest set that satisfies the requirement that* $\mathbb{A}_t \cup \mathbb{B}_t$ *is convex.*

Technological opportunity thus depends on the level of nonconvexity in $\mathbb{A}_t$. If we let $s(\mathbb{B}_t)$ be the size of the technological opportunity set, then $s(\mathbb{B}_t) = B_t \in \mathbb{R}_+$ is a real-valued measure of nonconvexity or of technological opportunity. A high $B_t$ implies a great deal of technological opportunity and vice versa. Furthermore, we also have that during periods of 'normal' incremental innovations, $B_{t-1} \geq B_t$ at all $t$.[9] $B_t$ will play an important role in the next section.

Obviously, progress comes to an end when technological opportunity is nearly exhausted, i.e. as $B_t$ goes to 0. How can there be any progress beyond this limit? The answer lies in an interplay between discoveries and drastic innovations. As was discussed above, entrepreneurs, researchers or laymen occasionally stumble upon some completely new and unexpected insight that does not fit into the prevailing technological paradigm. Torricelli's discovery of atmospheric pressure and Fleming's accidental finding of the principle behind penicillin, are examples which were mentioned above. Such findings are not the result of linear combinations of existing ideas, purposefully made by entrepreneurs. Instead, they have the character of 'new island formations', isolated from the coherent mass of technological knowledge. Some empirical evidence suggest that discoveries appear more or less continuously in time (see Figure 1). In Figure 3, discoveries are represented by the isolated black dots in technology space. To begin with, not many entrepreneurs pay any attention to these anomalies to the existing paradigm. As long as there is plenty of normal technological opportunity, no one bothers to undertake the highly risky and costly endeavour of trying to understand the technological potential of the isolated discoveries.

As technological opportunity approaches exhaustion, it becomes clear to entrepreneurs that progress will soon come to an end. Since the expected profits from further incremental innovation are lower than the expected profits from drastic innovation, the entrepreneurs move their attention to the unexplored, isolated discoveries of technology space and start trying to make combinations between these discoveries and already familiar ideas in $\mathbb{A}_t$. After considerable efforts in terms of time and money, entrepreneurs eventually succeed in making such 'long-distance' combinations and a cluster of drastic innovations appear. The wave of electricity-related innovations in the late 1800s might serve as an example. The combination of existing technologies for lighting, interpersonal communication, and sound reproduction with the discovery of electricity, gave the world such drastic innovations as the electric light bulb, the telephone, and the record player.

When the majority of the discoveries in this way have been connected

[9] Several earlier works have suggested that the decline in research outputs per worker that has been evident since (at least) World War II probably is a result of diminishing technological opportunity (Caballero and Jaffe, 1993; Kortum, 1997).

to the technology set, the period of drastic innovation comes to an end. The drastic innovations have created new areas of nonconvexity, i.e. the size of the technological opportunity set, $B_t$, has increased. I will refer to such a phenomenon as a *technological paradigm shift*:

**Definition 2** *If $B_t > B_{t-1}$, then a technological paradigm shift has occurred at $t$.*

The technological paradigm shift induces a wave of incremental innovations that tie up the loose ends that the drastic innovations have left in their wake. The more than hundred years of gradual improvements in automobiles since Benz' three-wheeler in 1885, is an example of this process. Innovative activity continues to be intense as the vast new areas of technological opportunity are exploited, but progress is now once again predictable and nonrevolutionary. While entrepreneurs make profits from normal science, new discoveries continue to appear like new islands in technology space. However, as long as there is plenty of technological opportunity in the prevailing paradigm, no one will pay these isolated and distant ideas any attention. And so the technological long-wave pattern repeats itself with decades of incremental normal progress and random discoveries, followed by years of clusters of drastic innovations which introduce new technological opportunities and paradigm shifts.

# 4 The Full Model

The previous section showed how the different kinds of technological knowledge interacted and evolved over time in the technology space of ideas. I this section, I present an extended but simple 'technological opportunity model' model of long-run economic growth. The carrying theme will be that drastic innovations are undertaken in periods of declining entrepreneurial profit from normal activity. Entrepreneurs are the main agents in this process and perform both incremental and drastic innovations and then also reap the profits. In this sense, the model follows in the spirit of Schumpeter (1934), Mensch (1979), and Jovanovic and Rob (1990), but the focus on technological opportunity makes it quite different from these and most other efforts.

As before, the size of the technological opportunity set for a 'world leader' country at some time $t$ is $B_t = s(\mathbb{B}_t) \geq 0$.[10] There are two sources of change in $B_t$: (i) Incremental innovations that decrease $B_t$

[10] The modelling of a 'world leader' in technology allows us to disregard all aspects of technological diffusion. See Olsson (2000) for a model that explicitly treats the diffusion of ideas.

and (ii) technological paradigm shifts, caused by drastic innovations, that reintroduce nonconvexity and increase the level of $B_t$. The two growth processes are described in (1). Incremental innovation - i.e. the convex combination of technological ideas that are technologically close - is the normal situation. The change in technological opportunity is then $B_t - B_{t-1} = -B_{t-1} \sum_{i=1}^{N} \delta_i$ where $\delta_i$ is the fraction of technological opportunity at $t-1$ that is successfully turned into new technological knowledge (i.e. increases in $s(\mathbb{A}_t) = A_t$) by individual entrepreneur $i$ and where $N$ is the number of entrepreneurs. Assuming for simplicity that $\delta_1 = \delta_2 = ... = \delta_N$, we have that $\sum_{i=1}^{N} \delta_i = N\bar{\delta} = \delta < 1$. $N$ and average entrepreneurial capacity $\bar{\delta}$ are exogenously given parameters.[11] Hence, at any time $t$, only $(1-\delta) B_{t-1}$ is available to the $N$ entrepreneurs for exploitation. In the long-run setting of this model, the time elapsed between $t-1$ and $t$ should not be thought of as a single year but rather as a decade.

$$B_t = \begin{cases} B_{t-1} - \delta B_{t-1} & \text{if } E_{t-1}\left(\Pi_t^I\right) > E_{t-1}\left(\Pi_t^D\right) \\ B_{t-1} + D_t & \text{if } E_{t-1}\left(\Pi_t^I\right) \leq E_{t-1}\left(\Pi_t^D\right) \end{cases}. \qquad (1)$$

Decades of normal technological activity are also associated with a zero-sum relation between the sizes of the sets for technological opportunity and knowledge: $B_{t-1} - B_t = A_t - A_{t-1} = \delta B_{t-1}$. If, on the other hand, there is a paradigm shift at $t-1$, then $A_t - A_{t-1} = 0$. Thus, the technology set grows only through incremental innovations. The relatively small number of drastic innovations that appear during paradigm shifts do not increase $A_t$ directly, but since such innovations reintroduce technological opportunity, they have a crucial indirect effect on the level of technological knowledge.

Entrepreneurs form their decisions on the basis of rational expectations about next decade's profits. If next decade's expected entrepreneurial profit from normal incremental innovation, $E_{t-1}\left(\Pi_t^I\right)$, is lower than or equal to the expected profit from drastic innovative activity, $E_{t-1}\left(\Pi_t^D\right)$, all entrepreneurs switch from incremental to drastic innovation at $t$.[12][13] The change in technological opportunity is then $B_t - B_{t-1} =$

[11] As argued by Baumol (1990) and others, the allocation of entrepreneurial talent to productive activities probably depends on social institutions and on the incentive system in general. By assuming a fixed $N$, this model disregards all sectoral allocation decisions.

[12] Both $\Pi_t^I$ and $\Pi_t^D$ should be thought of in terms of *profits per capita*, that is profits divided by the whole population in the economy. Scaling profits in this way is necessary because the discussion below will deal with income per capita. Note that the analytical use of profits per capita as the key variable does not give different results from using the more intuitive profits per entrepreneur.

[13] This inequality-condition for innovative activity resembles the condition for 'ex-

$D_t > 0$. In this sense, a trough in the economy triggers a technological paradigm shift. $D_t$ reflects the size of the technological opportunity shock that is created by the wave of drastic innovations and by the accompanying paradigm shift.

$D_t$ is a random variable whose expected value $E_t(D_t) = g(t, \delta, A_t)$ depends positively on the random number of accumulated discoveries at the time of the paradigm shift and on $\delta$ and the level of $A_t$. All else equal, the greater the entrepreneurial capacity and the greater the body of existing technological knowledge, the greater the expected technological opportunity shock. Thus, whereas the evolution of $B_t$ is highly predictable during normal innovative activity, decreasing by a percentage of $\delta$, the increase in the wake of a paradigm shift is not.

The stochastic level of profit per capita from drastic innovation, $\Pi_t^D$, is determined by two factors; the *cost* of combining an existing idea with a distant discovery and the *revenue* from that same undertaking. The cost for this kind of operation is always substantial, involving advanced machinery and effort by several people with expertise both on existing knowledge and on the rare discovery. Let us assume that the cost of a drastic innovation is $c > 0$ with certainty. Revenues are influenced by forces of demand which are more difficult to get a grip of for the individual entrepreneur. Let total revenues $R$ be uniformly distributed $(R \sim U(\cdot))$ in the closed interval $[0, r]$ with the probability of total failure $(R = 0)$ being just as likely as the probability of the maximum revenue $(R = r)$. Hence, profits per capita $\Pi_t^D = R - c$ is a random variable which assumes values in the interval $[r - c, -c]$. It can also be easily shown that the density function for some $x \in [0, r]$ is simply $f(x; r - c, -c) = \frac{1}{r}$ and that the expected level of profits for drastic innovations is

$$E_{t-1}\left(\Pi_t^D\right) = \frac{(r - 2c)}{2} \quad \text{for all } t. \tag{2}$$

$E_{t-1}\left(\Pi_t^D\right)$ is thus constant across time. Note that a requirement for drastic innovations ever to be undertaken is that (2) is positive, which implies that we must have that $r > 2c$.[14]

Profits per capita from incremental innovation are a function of technological opportunity in the previous period:

$$E_{t-1}\left(\Pi_t^I\right) = \Pi_t^I = \omega^E \delta B_{t-1} = \omega^E (A_t - A_{t-1}) \tag{3}$$

tensive' versus 'intensive' search in Jovanovic and Rob (1990). In reality, it is never the case that *all* entrepreneurs switch from incremental to drastic innovation simultaneously. The assumption is made to keep the model simple.

[14]It should also be remarked that according to the specifications above, $\Pi_t^D$ is independent of $D_t$. What this means is that the profitablility of drastic innovations is uncorrelated with the size of the paradigm shift that they induce.

Unlike $\Pi_t^D$, $\Pi_t^I$ is not a stochastic variable. Since $B_{t-1}$, $\delta$, and $\omega^E$ are all known at $t-1$, entrepreneurs know with certainty what profits they can make in the next period from incremental innovation. At all $t$ characterized by normal R&D activity, a fraction $\delta < 1$ of existing technological opportunity is exploited and the profits show up with a lag in the next period. $\omega^E > 0$ is an entrepreneurial productivity parameter that indicate to what extent exploited technological opportunity is transformed into profits. $\delta$ captures the inherent capacity of the 'world leader'-society to exploit intellectual opportunities. This crucial determinant of development is modelled as a constant but should ideally be seen as time varying and probably a function of a country's human capital and societal institutions.[15]

The lag between the extraction of ideas at $t$-1 and the profits at $t$ is a result of the time needed for entrepreneurs to bring the new idea to the market and for consumers to get acquainted with it. The costs for incremental innovations are lower than the costs of drastic innovations.[16] For simplicity, I assume that the costs are zero and that profits are directly proportional to the exploited fraction of technological opportunity. Since it was shown before that $\delta B_{t-1} = A_t - A_{t-1}$, it follows that $\Pi_t^I$ is also proportional to the increase in technological knowledge.

By substituting in the expressions for $B_{t-1}$ from (1), we can further see that we must have either $\Pi_t^I = \omega^E \delta (1-\delta) B_{t-2}$ or $\Pi_t^I = \omega^E \delta (B_{t-2} + D_{t-1})$, depending on whether $t-1$ was characterized by normal activity or by drastic innovations. Since $B_t$ shrinks during normal activity, the standard case is that $\Pi_{t-1}^I > \Pi_t^I$. Profits continue to decrease until the critical lower level $E_{t-1}\left(\Pi_t^D\right)$ is reached.

By using (3) and (2), we can therefore define the *actual* level of entrepreneurial profit as:

$$\Pi_t = \begin{cases} \Pi_t^I = \omega^E \delta B_{t-1}, & \text{if } E_{t-1}\left(\Pi_t^I\right) > E_{t-1}\left(\Pi_t^D\right) \\ \Pi_t^D \sim U\left(\frac{(r-2c)}{2}, \frac{r^2}{12}\right) & \text{if } E_{t-1}\left(\Pi_t^I\right) \leq E_{t-1}\left(\Pi_t^D\right). \end{cases} \quad (4)$$

The break-even point for expected profits, i.e. the level where an entrepreneur is indifferent between continued incremental innovation and a switch to drastic innovative activity, is thus $\Pi_t^I = \omega^E \delta B^* = E_{t-1}\left(\Pi_t^D\right) = \frac{1}{2}(r-2c)$, implying a critical level of technological opportunity, $B^* = \frac{1}{\omega^E \delta}(r-2c)$. When $B_{t-1}$ is equal to or lower than this level, all entrepreneurs switch to drastic innovative activity at time $t$. As noted above, expected profits will then be $\frac{1}{2}(r-2c)$, but the actual

[15] Olsson (2000) includes a lengthy discussion of the factors behind technological creativity. See also Mokyr (1990) and Baumol (1990) for influential accounts on the same issue.

[16] The same assumption is made in Jovanovic and Rob (1990).

level might of course end up below the expected value and below last period's profits from incremental innovations. At $t+1$, incremental innovation once again becomes profitable since technological opportunity has increased by $D_t$ during $t$. Thus, a new period of normal innovative activity is initiated.

Lastly, I will propose an extremely general function for logged aggregate income per capita:

$$\ln y_t = \omega^M z_t = \ln y_{t-1} + \Pi_t \tag{5}$$

$z_t$ stands for a broad interpretation of capital per capita, including both human and physical capital, and $\omega^M > 0$ is a parameter. Together they form an aggregate production function whose output equals logged income per capita in the previous decade plus profits per capita at $t$. Entrepreneurial profit per capita $\Pi_t$ is either equal to realized profit from incremental or from drastic innovation. The growth rate in this economy is therefore simply given by $\ln y_t - \ln y_{t-1} = \Pi_t$, i.e. in the long run, it is entrepreneurial profits from innovative activities that drive economic development. Note also that during years of normal innovative activity, $\ln y_t - \ln y_{t-1} = \omega^M (z_t - z_{t-1}) = \omega^E (A_t - A_{t-1})$, implying that the increase in technological knowledge is transformed into increases in capital according to $z_t - z_{t-1} = \frac{\omega^E}{\omega^M} (A_t - A_{t-1})$.

# 5 Analysis

The five equations in the section above do not allow for any straightforward analytical solutions or results. However, the system of equations can be used to simulate the behaviour of the model under different assumptions. Table 1 shows the results from five such simulations.[17] By trying different parameter values for $\delta$ and $E_{t-1}\left(\Pi_t^D\right)$, we receive different values for the endogenously determined relative levels of output in the final period ($y_{20}/y_0$) and technological knowledge ($A_{20}/A_0$). We also receive a varying number of cycles. As a general conclusion, the lower the levels of $\delta$ and $E_{t-1}\left(\Pi_t^D\right)$, the lower the levels of ($y_{20}$) and ($A_{20}$) and the fewer the number of paradigm shifts. For instance, an 'extraction' rate of 5 percent ($\delta = 0.05$) and a critical lower level of output growth rate rate per capita of 0.1 percent ($E_{t-1}\left(\Pi_t^D\right) = 0.001$), result in an end-period output level that is roughly 1.8 times the initial level and a

[17] In all cases, I have used $\omega^E = 0.001$, $B_{-1} = 100$, $y_0 = 1000$, $A_0 = 1000$, and No.of periods = 20. $D_t$ is a random, uniformly distributed variable computed according to the formula $RANUNI * A_t/20*(1+\delta)+30$ where $RANUNI$ is a computer-generated number between 0 and 1. As mentioned in the text, $D_t$ is thus a positive function of $A_t$ and $\delta$. Likewise, $\Pi_t^D$ assumes values according to a uniform distribution in the interval $E\left(\Pi_t^D\right) \pm 0.05$.

technological knowledge that has increased by a mere 6.1 percent. During the twenty periods, there is not a single paradigm shift. In contrast, in Case V, the higher levels of $\delta$ and $E_{t-1}\left(\Pi_t^D\right)$ result in six paradigm shifts, a more than 154 times higher relative level of end-period output ($y_{20}/y_0 = 154.503$) and a more than 46 percent greater level of technological knowledge ($A_{20}/A_0 = 1.461$).

Table 1: Model simulation results

| | Case | | | | |
|---|---|---|---|---|---|
| | I | II | III | IV | V |
| $\delta$ | 0.05 | 0.1 | 0.2 | 0.3 | 0.3 |
| $E_{t-1}\left(\Pi_t^D\right)$ | 0.001 | 0.005 | 0.01 | 0.01 | 0.02 |
| $y_{20}/y_0$ | 1.839 | 4.834 | 18.651 | 29.516 | 154.503 |
| $A_{20}/A_0$ | 1.061 | 1.145 | 1.273 | 1.283 | 1.461 |
| No. of shifts | 0 | 2 | 4 | 4 | 6 |

The time series for Cases I and IV are plotted in Figures 4-5. The figures show the simulated patterns of profits/output growth rates and paradigm shifts over twenty periods where the size of the technological opportunity shock is downscaled 200 times so that the two curves can be illustrated in the same diagram. As was mentioned above, one period should not be thought of as being equivalent to one year but rather as a ten-year period. Thus, the figures might be interpreted as covering a two hundred year-era, in particular the period 1800-2000.

Figure 4 shows the time series for Case I when the values of $\delta$ and $E_{t-1}\left(\Pi_t^D\right)$ are relatively low. During the whole period, the growth rate lingers between 0.5 and 0.18 percentage points with an average of 0.31 percent. In a very slow manner, the growth rate converges towards the critical level 0.1, but even in period 20, the attainment of the critical level lies far ahead in the future. Case I illustrates a society with a stable, non-dynamic development where improvements in standards of living are hardly recognizeable during an average human life.

In Figure 5, both the opportunity extraction rate $\delta$ and the critical level of $E_{t-1}\left(\Pi_t^D\right)$ are higher (0.3 and 0.01 respectively). The result is a much higher average growth rate (1.8 percent) but also a much more volatile time series (standard deviation is 1.5 percentage points). Four cycles with a duration of four to five decades are endogenously created, implying the familiar 'long wave-pattern' of 40-50 year cycles. One serious depression in the fourth decade occurs as a result of the technological paradigm shift. During all other periods with paradigm shifts, the profit from drastic innovation exceeds the expected level, significantly so in the ninth decade. In general, Case IV describes a dynamic development process that is far from the stable, undramatic scenario in Figure 4.

From (4) and (5), we know that the growth rate of income per capita, and hence the whole development process, might be summarized in a general function

$$\ln y_t - \ln y_{t-1} = \Pi_t = g\left(r, c, \delta; B_{t-1}\right) \tag{6}$$

where $r, c$ and $\delta$ are exogenous parameters and where $B_{t-1}$ is an endogenously determined variable.[18] Let us discuss each component in turn.

The levels of $r$ and $c$ determine the expected profits from drastic innovation, $E_{t-1}\left(\Pi_t^D\right)$. The maximum revenue from drastic innovations, $r$, depends crucially on a society's system of intellectual property rights. Without property rights, a new idea is a nonexcludable good which anyone can utilize freely. The most important institution for intellectual property rights is presumably *the patent*. An enforceable patent gives its holder a temporary monopoly to some idea and the patentee is guaranteed the profits from commercial exploitation. The patent is a recent institution, dating back to laws passed in early seventeenth century Britain (Mokyr, 1990). However, a patent to a drastic innovation is a necessary but not a sufficient condition for a high $r$. Equally important is probably the existence of a *free market* and a *demand for new goods*. As Landes (1969) notes, at the time of the Industrial Revolution, Western Europe had already moved away from subsistence levels of consumption and could thus afford the luxury of new goods.

Profits from drastic innovations of course also depend on the costs incurred. Costs arise from the employment of physical capital and labour and should generally increase with the technological distance between the discovery and the existing idea to be combined. Innovative activities are risky endeavours and as discussed by Schumpeter (1934) and later writers, an entrepreneur's profit often hinges on his or her availability of credit. In this respect, banks and financial markets play a crucial role. A well-functioning credit market should imply lower levels of $c$.

But even if revenues are high and costs of drastic innovation are low, the cluster of technological breakthroughs becomes useless unless people are able to effectively utilize the technological opportunity created. This is where the $\delta$-parameter fits in. As discussed above, $\delta = N\bar{\delta}$ where $N$ is the number of entrepreneurs and $\bar{\delta}$ is their average capacity to exploit the potential of technological paradigms. Obviously, an exogenous increase in the number of entrepreneurs $N$ leads to a proportional increase in exploitation capacity. $\bar{\delta}$ probably depends to a great extent on the level of human capital and on the underlying institutions of a society. General levels of literacy, technical proficiency, and life expectancy should

[18] I choose to ignore the parameters $\omega^{\mathsf{M}}$ and $\omega^{\mathsf{E}}$ in this analysis.

have positive effects on $\bar{\delta}$. Likewise, institutions such as corporate laws, universities, freedom of technological inquiry, and a favourable public attitude towards rationalism and scientific curiousity, are all important explanations for the level of $\bar{\delta}$. The rise of commercial R&D departments within big firms in the late nineteenth century led to a routinized innovation process and was a major watershed in the history of technology that probably greatly increased both $\bar{\delta}$ and $N$ (Schumpeter, 1942; Peretto, 1998).

The prime source of growth in the model above, however, is technological opportunity. Except for being heavily influenced by the level of the parameters $r, c$ and $\delta$, $B_t$ also depends crucially on the size of the random technological opportunity shocks $D_t$. It was discussed earlier that the level of $D_t$ depends on the random and exogenously given appearance of discoveries. In some eras, discoveries happen to be more frequent than in other eras. Consequently, the technological opportunity shock induced by drastic innovations varies from one paradigm shift to another, with obvious effects on the ensuing growth rate (see Figures 4-5). This also means that there is a substantial random element in the whole development process.

Does the model then give a reasonable description of economic history? The rest of this section will discuss that issue. I would argue that Figure 4 gives a roughly accurate picture of world economic development before 1800. As has been suggested by Maddison (1982), Fogel (1999) and others, per capita income levels have remained near subsistence levels during most of world history. Some estimates indicate that growth of GDP per capita was nonexistent until about 1500 AD when average income levels started to grow by a modest 0.1 percent a year (Maddison, 1982, Table 1.2).

The model and the simulations above can favourably be reconciled with these facts. Before about 1800, the absence of a world demand for new goods and an efficient system of intellectual property rights implied that expected revenues from drastic innovation ($r$) were very likely to be lower or only slightly above twice the costs ($2c$) so that no or very few technological paradigm shifts were endogenously generated. Furthermore, low general levels of education, a low esteem of production-related knowledge, and poorly functioning corporate and private property laws probably meant that $\delta$ was hovering near zero. Thus technological progress was slow and incremental. Analogously, growth of per capita income was also slow or nonexistent and largely noncyclical. All these circumstances greatly resemble the situation depicted in Figure 4. With the introduction of public education, intellectual property rights, a world market for new goods, and social norms towards technological advance, nineteenth century Britain was the first country to experience a shock in the levels of $r$ and $\delta$ that was sufficiently strong to set the endogenous innovation-machinery in motion. Germany and the United States then followed by the end of the 1800s.

With the shock in the levels of $r$ and $\delta$ that occurred after 1800, the Western world entered the capitalist era. By 1820, US income per capita (in 1990 US dollars) was 1,287 (Maddison, 1995). It is well known that the subsequent era was characterized by relatively high average growth rates, a greater degree of variability and of cyclic behaviour, greater profits for entrepreneurial activity, and faster technological progress. All these ingredients are present in Figure 5. The 'capitalist engine' quickly exploits technological opportunities so that only about fifty years after a paradigm shift takes place, another one is introduced. Growth rates change by up to six percentage points from one decade to another (see periods 4 and 5). In the simulation, the level of output increases from 1,000 in the starting period to 29,516 in the end-period. This twenty nine-fold rise in income is roughly in line with the Maddison data where US income increased from 1,287 in 1820 to 22,569 in 1994. In comparison to earlier epochs, the capitalist era is a roller-coaster ride of creation and destruction that, in Schumpeter's (1942, p 83) words '...incessantly revolutionizes the economic structure *from within*, incessantly destroying the old one, incessantly creating a new one.'

Apart from being in line with these common generalizations of economic history, how well does the model survive a more careful investigation of available data? Figure 6 shows drastic innovation activity and GDP per capita growth for the time period 1870 to 1968. The drastic innovation curve is a composite index of the seven-year moving averages of the Mensch, van Duijn, and CFS time series in Figure 2. The growth curve shows the seven-year moving average of the average annual GDP per capita growth rate in United Kingdom, United States, and Germany. These countries were arguably the three world leaders at the time in terms of innovation activity. The calculations use data from Maddison (1995). The inclusion of Germany makes the time series highly volatile, especially during the period 1918-1950 when both post-war depressions and subsequent recoveries are more intense in Germany than in the less badly hurt UK and the US. In general, it should be acknowledged that the numerous economic, political and institutional influences that the three sample countries were subjected to during this era, make causal interpretations a complicated task.

Nevertheless, when the two curves are presented jointly, some interesting regularities appear. Among the nineteenth century growth rates in the period covered, 1876 has the lowest rate (0.99 percent).[19] This is also the date when the intensity in drastic innovation takes off in a remarkable increase, reaching a peak in 1882. Late nineteenth century growth rates, however, did not reach a peak until 1898. Relating these facts to the model, one might argue that the low growth rates around 1875 triggered the electricity-related paradigm shift that unfolded during the following years. The growth payoffs from the technological opportunity shock that drastic innovations such as the telephone, the electric light bulb, the gasoline motor and electric heating introduced, did not materialize until the last years of the 1800s. This lag from the peak of the drastic innovation wave and the increase in growth, is certainly well in line with the predictions of Figure 5.

From 1898, it is clear that despite at least two business cycle booms, a long downward trend is initiated that lasts until the end of World War I. A significant post-war boom then ensues with a peak in 1925, only to be followed by the worst non-war related depression in recorded economic history. One might therefore speak of a downward trend in growth rates that lasts until 1929. An interesting fact is that drastic innovation activity starts picking up again from very low levels around 1916. The really sharp increase, however, does not occur until about 1928. From 1932 to 1937, the Mensch-sample records as many as sixteen drastic innovations, for instance the radar, the rocket, television, helicopters, titanium, and the kodachrome camera. Activity peaks in 1934 and then declines.[20] The height of output growth is reached five years later, in 1939. If we disregard the war years and the post-war depression in war-ridden Germany, we see that the 'normal' growth process seems to start again in the beginning of the 1950s, this time at substantially higher growth levels. There is then once again a long-run downward trend that continues into the 1990s.

Even this twentieth century picture might be regarded as supporting the main hypotheses advanced in this article. The technological paradigm shift of the 1930s appears after a serious depression and after three decades of a falling trend in growth rates. After World War II, growth rates stabilize at unprecedented high levels, no doubt largely because of the economic integration of Western Europe, but probably also as a result of the new technological opportunity created by the drastic innovations of the 1930s. As technological opportunity wanes, growth rates

[19] The annual growth rates for each country in 1876 are -0.09 percent for the UK, -0.96 for the US, and -1.77 for Germany.

[20] The CFS-sample indicates that the real decline does not start until 1943, whereas the van Duijn-sample suggests a decline starting as late as 1950. All of them, however, records a sharp peak in 1934.

gradually decline during the late 1960s and 1970s.

It would certainly be incorrect to claim that my interpretation of Figure 6 provides conclusive evidence in support of my model. However, my two main hypotheses - that a cluster of drastic innovations should appear after a longer period of falling growth rates and that growth rates a decade or so after the paradigm shift should be relatively high - do indeed receive some support from the data. A more careful econometric analysis should be able to shed further light on the causal relationships between technological revolutions and economic growth.

## 6 Growth Theory

The technological opportunity model in section 4 departs in significant ways from most previous economic growth models. The primary difference is probably that it is technological opportunity $B_t$ rather than the stock of technological knowledge $A_t$ that is the crucial factor for long run growth. Regardless of the size of $A_t$, $B_t$ must be greater than zero if there is going to be any growth in the economy. Furthermore, the endogenously created technological paradigm shifts gives fundamentally different implications for the behaviour of growth in the very long run. In this section, I will briefly compare the implications of my model for long-run growth with the implications of the neoclassical and the endogenous growth models.

Let us consider the following two models: (i) The *neoclassical* model of Solow (1956) where income per capita equals $y = Ak(t)^{\alpha}$ and where the growth rate of $y$ is $g_y = \alpha\frac{\dot{k}}{k} = \alpha\left(sAk\left(t\right)^{\alpha-1} - \gamma\right)$. As usual, $A$ is a technology parameter, $k\left(t\right)$ is capital per capita, $s$ is the saving rate, and $\gamma$ is the depreciation rate of capital. (ii) The linear $AK$-model where $y = Ak(t)$ and $g_y = \frac{\dot{k}}{k} = sA - \gamma$.[21]

As is well known, with a $k$ initially below its steady-state level and in the absence of policy shocks, the neoclassical model yields positive growth rates that decrease with time until the rate converges to zero. Hence, the neoclassical model implies a dynamic pattern similar to that described in Figure 4, i.e. with low levels of $\delta$ and $E_{t-1}\left(\Pi_t^D\right)$. At the steady state level, only changes in the parameters $s, A$, and $\gamma$ can make the economy advance. For instance, an exogenous increase in $A$ would temporarily increase the growth rate and lead to a higher steady-state level of output, but the effects on growth would only be transitory since

[21] The $AK$-model has been treated by several growth theorists and is often associated with Romer (1986). In this category, I would also include contributions such as Romer (1990) and Jones (1995).

the equilibrium growth rate would still be zero. Although the issue is far from generally agreed upon, there appears to be empirical evidence in favour of such a 'conditional convergence'-hypothesis for the period 1960 to the present (Barro, 1991). In the long run perspective, however, it is apparent that the world economy has not witnessed diminishing returns and stagnating growth. On the contrary, average growth rates have steadily increased from the sixteenth century onwards (Maddison, 1982). Repeated exogenous shocks to the levels of $s$, $A$, and $\gamma$ can hardly be regarded as satisfactory explanations for this development.

The endogenous growth models of the $AK$-type are better suited for explaining development in the long run. Unlike the neoclassical model, the $AK$-model implies a constant positive growth rate which is determined by the level of the parameters. An increase in a parameter such as $A$ results in a permanent increase in the growth rate. This conjecture might indeed be considered as a reasonable description of the increase in growth rates that the Western world witnessed from 1500 onwards and that was most evident after the Industrial Revolution. However, the slow but persistent decreases in growth rates that the UK, the United States and Germany experienced both after 1898 and after 1950 are not easily reconciled with the $AK$-model.

The technological opportunity model in this article has the advantage of explaining both falling growth rates in the medium term and increasing average growth rates in the very long run. Increases in $\delta$ and $r$ that took place in the nineteenth century led to a more pronounced cyclic behaviour of development but also to higher average growth. Periods of declining entrepreneurial profits and GDP growth rates were succeeded by a decade or so of intense drastic innovative activity, resulting in an increase in technological opportunity and in average growth rates. This pattern appears to have repeated itself at least during the clustering of drastic innovations in the 1880s and 1930s.

## 7 Concluding Remarks

The technological opportunity model presented in this article provides one explanation to the puzzle of why technological progress is cyclical. The proposition is simply that although incremental, non-revolutionary innovation is the normal activity for profit-maximizing entrepreneurs, the gradual exhaustion of technological opportunity implies that the time will always come when expected profits from further incremental innovations fall below expected profits from highly risky and expensive drastic innovation. For a brief and intense period, entrepreneurs will therefore switch to drastic innovation and try to combine existing technological knowledge with the previously isolated and technologically distant discoveries. The wave of drastic innovations that follow causes a new technological paradigm to appear and reintroduces technological opportunity. Entrepreneurs then switch back to riskless and once again profitable incremental innovation. The prediction that drastic innovations will appear in bad times is supported by the data for the two innovation clusters around 1880 and 1930. There is also some evidence that average growth rates after the two waves of innovations were relatively high.

The model might be extended in a number of directions. Perhaps the most fundamental determinant of cycle duration and growth rates of income is the societal capacity to exploit technological opportunity, $\delta$. In this setting, $\delta$ captures the quality of a number of diverse factors such as human capital and social institutions. A possible extension would be to let $\delta$ be determined endogenously. It would also be possible to let the maximum revenue from drastic innovation, $r$, be the solution to an optimization problem where the benefit from property rights to ideas is weighed against the benefit of completely free ideas.

The most obvious candidate for further research is probably a more elaborate econometric analysis of the relationship between drastic innovations and economic growth. The model and the data presented in this article have suggested that an important long-run relationship might be present, a relationship that the empirical growth literature, mainly using data from 1960, possibly has missed.

# References


[1] Aghion, P. and P. Howitt (1998) *Endogenous Growth Theory*, Cambridge, Mass: MIT Press.

[2] Barro, R.J. (1991) "Economic Growth in a Cross Section of Countries", *Quarterly Journal of Economics* 106(2): 407-443.

[3] Baumol, W. (1990) "Entrepreneurship: Productive, Unproductive, and Destructive", *Journal of Political Economy* 98(5): 893-921.

[4] Boldrin, M. and D.K. Levine (2001) "Growth Cycles and Market Crashes," *Journal of Economic Theory* 96(1): 13-39.

[5] Bresnahan, T. and M. Trajtenberg (1995) "General Purpose Technologies: 'Engines of Growth?" *Journal of Econometrics* 65, 83-108.

[6] Caballero, R. and A. Jaffe (1993) "How High Are the Giants' Shoulders?: An Empirical Assessment of Knowledge Spillovers and Creative Destruction in a Model of Economic Growth", Working Paper No. 4370, NBER.

[7] Caselli, F. (1999) "Technological Revolutions", *American Economic Review* 89(1): 78-102.

[8] Clark, J., C. Freeman and L. Soete (1983) "Long Waves, Inventions, and Innovations", in Freeman, C. (1983)(ed), p 63-78.

[9] David, P. (1990) "The Dynamo and the Computer: An Historical Perspective on the Modern Productivity Paradox", *AER Papers and Proceedings* 80(2): 355-361.

[10] Dosi, G. (1988) "Sources, Procedures, and Microeconomic Effects of Innovation", *Journal of Economic Literature* 26(September): 1120-1171.

[11] van Duijn, J.J. (1983) *The Long Wave in Economic Life*, London: George Allen & Unwin.

[12] Fogel, R.J. (1999) "Catching Up with the Economy", *American Economic Review* 89(1): 1-22.

[13] Freeman, C. (1983)(ed) *Long Waves in the World Economy*, London: Butterworths.

[14] Gordon, R.J. (2000) "Does the 'New Economy' Measure Up to the Great Inventions of the Past?", *Working Paper* 7833, NBER.

[15] Helpman, E. (1998)(ed.) *General Purpose Technologies and Growth*, Cambridge, Mass: MIT Press.

[16] Helpman, E. and M. Trajtenberg (1998) "A Time to Sow and a Time to Reap: Growth Based on General Purpose Technologies", in Helpman (1998)(ed.), 55-85.

[17] Jaffe, A. (1986) "Technological Opportunities and Spillovers of R&D: Evidence from Firms' Patents, Profits, and Market Value", *American Economic Review* 76(5): 984-1001.

[18] Jewkes, J., D. Sawers, and R. Stillerman (1958) *The Sources of Invention*, London: Macmillan.

[19] Jones, C. (1995) "R&D-Based Models of Economic Growth" *Journal of Political Economy* 103(4): 759-784.

[20] Jovanovic, B. and R. Rob (1990) "Long Waves and Short Waves: Growth Through Intensive and Extensive Search", *Econometrica* 58(6): 1391-1409.

[21] Kauffman, S., J. Lobo and W.G. Macready (2000) "Optimal Search on a Technology Landscape", *Journal of Economic Behaviour and Organization* 43: 141-166.

[22] Kleinknecht, A. (1983) "Observations on the Schumpeterian Swarming of Innovations", in Freeman, C (1983)(ed), p 48-63.

[23] Kortum, S. (1997) "Research, Patenting, and Technological Change", *Econometrica* 65(2): 1389-1419.

[24] Kydland, F.E. and E. Prescott (1982) "Time to Build and Aggregate Fluctuations", *Econometrica* 50: 1345-1370.

[25] Kuznets, S. (1940) "Schumpeter's Business Cycles", *American Economic Review* 30: 257-271.

[26] Kuhn, T. (1962) *The Structure of Scientific Revolutions*, University of Chicago Press.

[27] Landes, D. (1969) *The Unbound Prometheus*, Cambridge University Press.

[28] Lipsey, R.G., C. Bekar, and K. Carlaw (1998) "What Requires Explanation?", in Helpman (1998)(ed.), 15-54.

[29] Maddison, A. (1982) *Phases of Capitalist Development*, Oxford University Press.

[30] Maddison, A. (1995) *Monitoring the World Economy*, Paris: OECD.

[31] Mensch, G. (1979) *Stalemate in Technology*, Cambridge, Mass: Ballinger Publishing Company.

[32] Mokyr, J. (1990) *The Lever of Riches*, Oxford University Press.

[33] Olsson, O. (2000) "Knowledge as a Set in Idea Space: An Epistemological View on Growth", *Journal of Economic Growth* 5(3), 253-276.

[34] Peretto, P. (1998) "Technological Change, Market Rivalry, and the Evolution of the Capitalist Engine of Growth", *Journal of Economic Growth* 3(1): 53-80.

[35] Romer, P. (1986) "Increasing Returns and Long-Run Growth", *Journal of Political Economy* 94(5): 1002-1037.

[36] Romer, P. (1990) "Endogenous Technological Change", *Journal of Political Economy* 98: 71-102.

[37] Schumpeter, J. (1934) *The Theory of Economic Development*, London: Transactions Publishers.

[38] Schumpeter, J. (1942) *Capitalism, Socialism, and Democracy*, New York: Harper Perennial.

[39] Solow, R.J. (1956) "A Contribution to the Theory of Economic Growth", *Quarterly Journal of Economics* 70: 65-94.

[40] Weitzman, M. (1998) "Recombinant Growth", *Quarterly Journal of Economics* 113(2): 331-360.

Figure 1: Cumulative distribution of discoveries and drastic innovations according to the Mensch sample, 1800-1941.

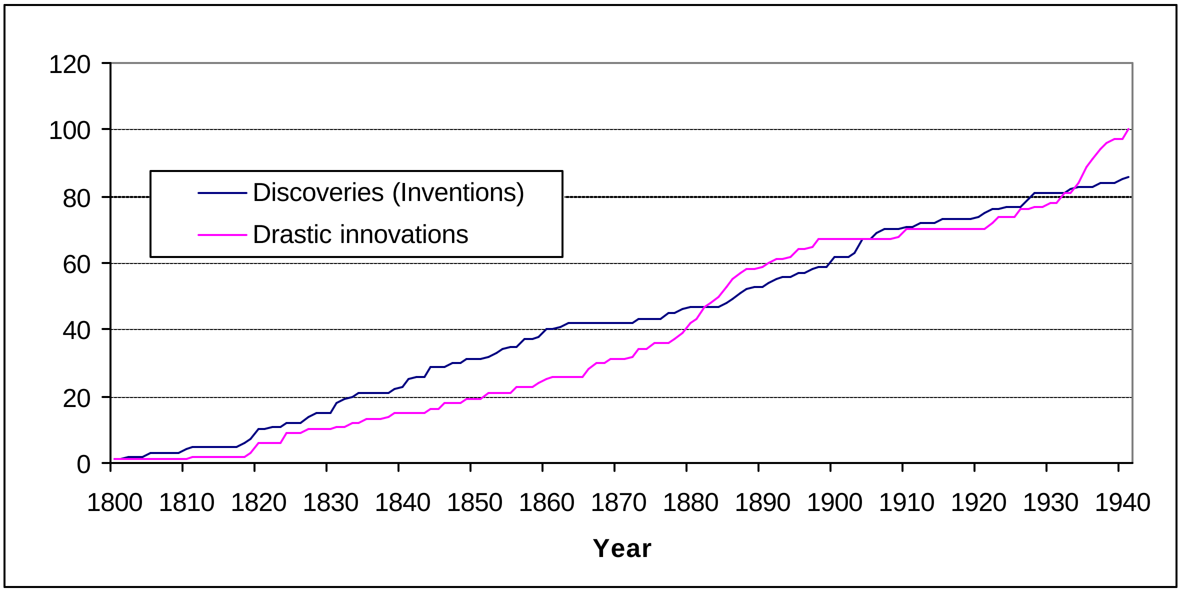
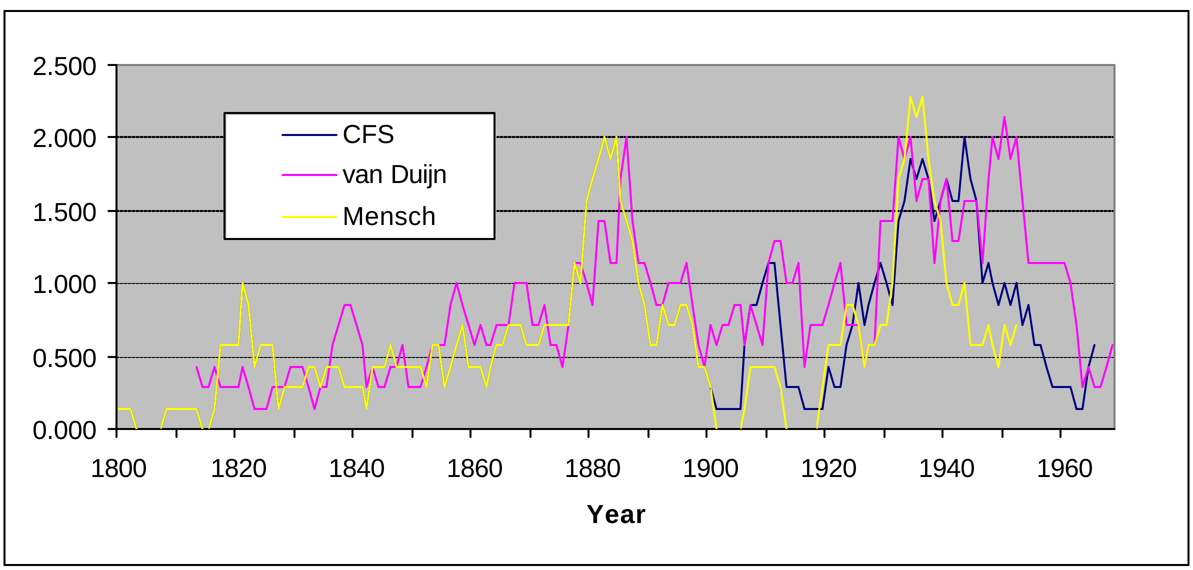



Figure 2: Drastic innovation activity, 1800-1968.




Note: The figure displays three series of seven-year moving averages of drastic innovations per year on the basis of the CFS, Mensch, and van Duijn samples.

Figure 3: The technology set

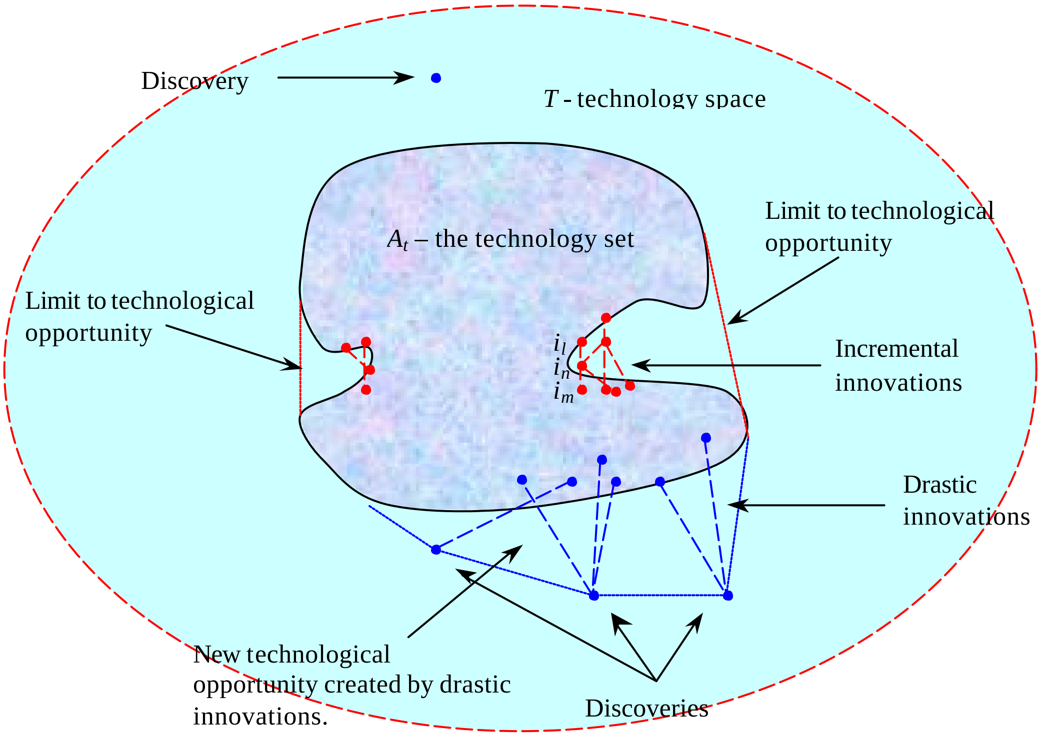
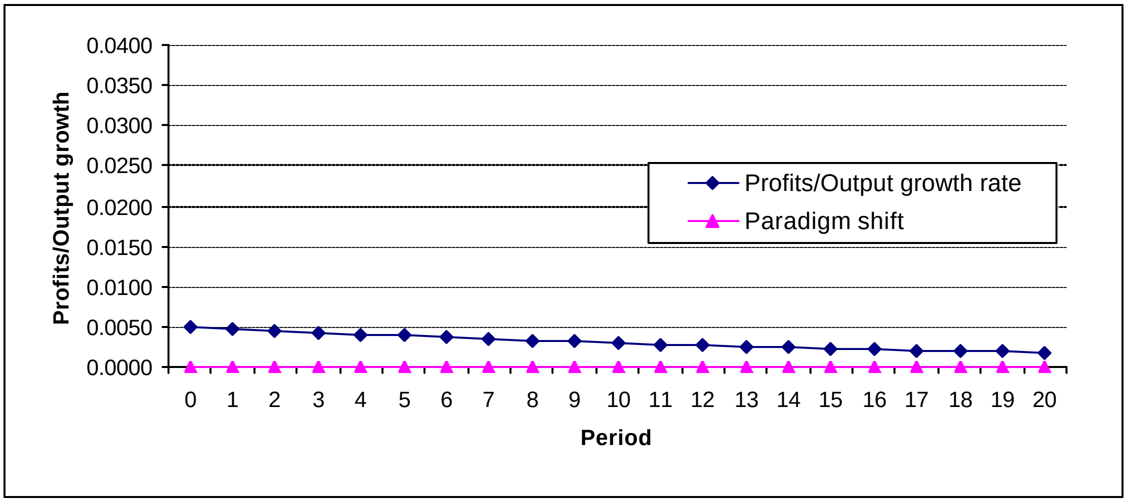
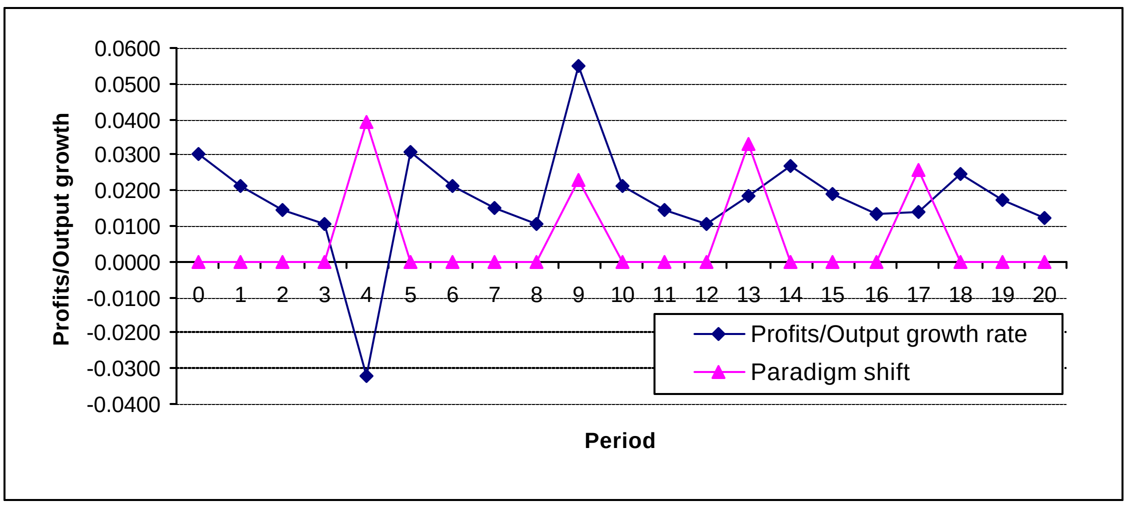


Figure 4: Simulated pattern of profits/output growth rates and paradigm shifts, Case I.



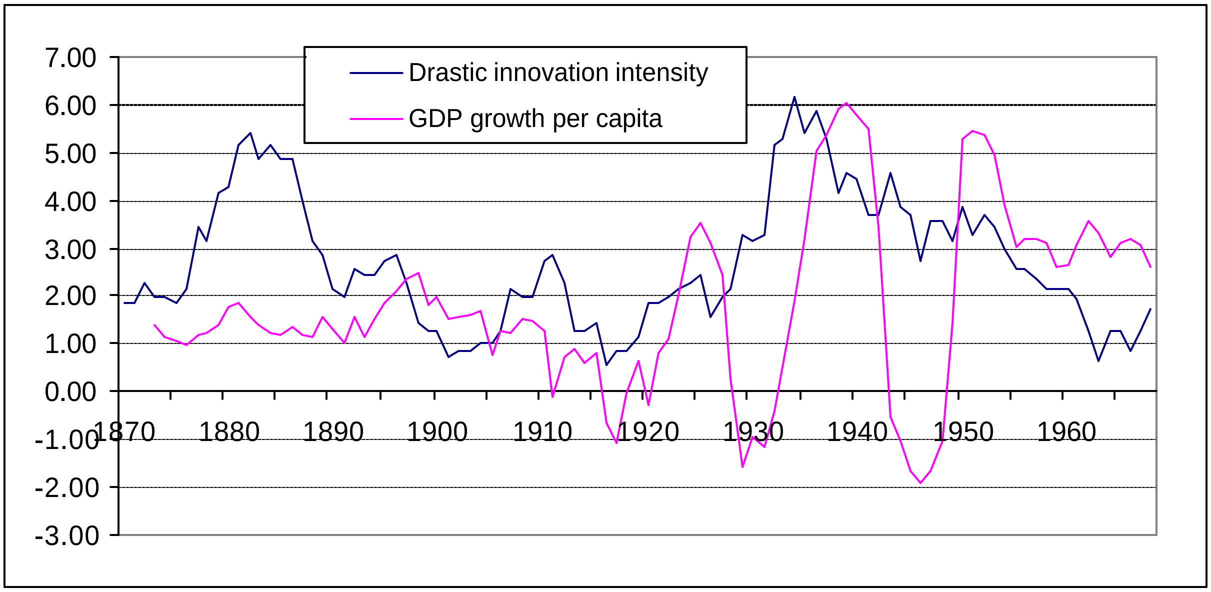

Note: δ=0.05, $E(\Pi)$=0.001. The paradigm shift-curve is downscaled 200 times.

Figure 5: Simulated pattern of profits/output growth rates and paradigm shifts, Case IV.




Note: δ=0.3, $E(\Pi)$=0.01. The paradigm shift-curve is downscaled 200 times.

Figure 6: Drastic innovation activity and average GDP per capita growth, 1873-1968.




Note: The GDP growth-curve is obtained by calculating a seven-year moving average for the average annual growth rates of UK, the US, and Germany. The drastic innovation-curve shows a composite seven-year moving average index constructed from the CFS, Mensch and van Duijn samples.