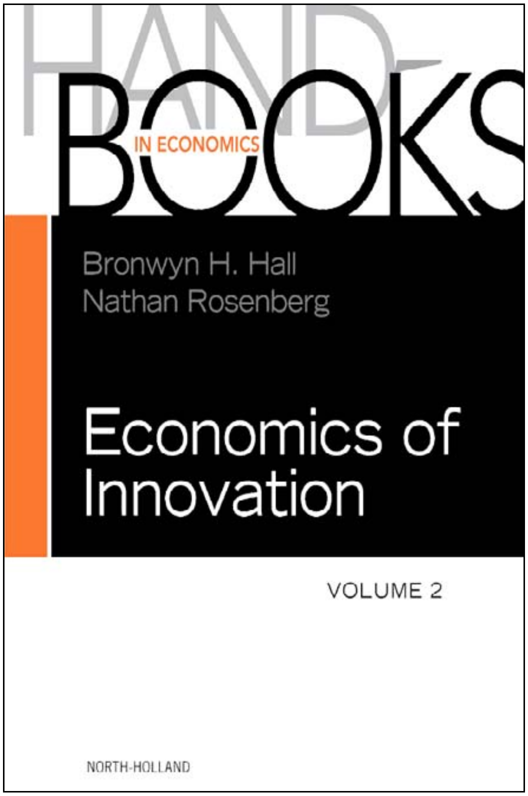

From: Timothy Bresnahan, General Purpose Technologies,
In Bronwyn H. Halland and Nathan Rosenberg:
Handbook of the Economics of Innovation- Vol-II,
Burlington: Academic Press, 2010, pp.761-792.
ISBN: 978-0-444-53609-9,

Chapter 18

GENERAL PURPOSE TECHNOLOGIES

TIMOTHY BRESNAHAN
Department of Economics
Stanford University, Stanford
California, USA

Contents

Abstract

This chapter selectively surveys the literature on general purpose technologies (GPTs), focusing on incentives and aggregate growth implications. The literature on classical GPTs (steam, electricity, computers) and on classical great economic transformations (industrial revolutions, the information age)

Handbooks in Economics, Volume 02

DOI: 10.1016/S0169-7218(10)02002-2

are linked to the theoretical and empirical literatures. The implications of GPT analysis for understanding the history of productivity growth in the late twentieth century are taken up on the concluding remarks.

Keywords

general purpose technology, innovation, technical change

JEL classification: L0, L1, O3

1. Introduction

The concept of "general purpose technologies" (GPTs) came into the economic analysis of technical change and growth nearly two decades ago with a number of distinct but linked research goals.[1] One goal lies in growth macroeconomics, to provide an explanation of the close link between whole eras of economic growth and the innovative application of certain technologies, called GPTs, such as the steam engine, the electric motor, or computers. Another goal is in the microeconomics of technical change and proceeds by differentiating between innovations of different types. The incentives and information related to the invention of GPTs themselves, for example, may differ from those related to the invention of applications; another example would be the incentives and information related to an established GPT with successful applications in contrast to earlier stages. A third goal links the macro and the micro. Can we understand the linkages between aggregate economic growth and the incentives and information structures related to particular inventions and to their application in particular uses and sectors?

Each of these goals has received considerable attention, with a different mix of theoretical, empirical, and historical methods. There has been a great deal of progress, though issues remain open about how to test the most interesting and important ideas. Some criticisms of the approach have emerged as well.[2] That makes this a good stage to revisit the original goals as altered by new research. A number of overviews of the literature have been written, giving me in this chapter the opportunity to look forward.[3]

1.1. Basic structure of GPT

The original motivation for the idea of GPTs arose in part from the history of economic growth, by the study of such key technologies in the past as steam or electricity and by the observation by economic historians that pointed to certain technologies as having a central role in growth.[4] Historical work also pointed to the importance of complementarities between innovations in different "technologies" as engineers understand that term.

My turn to the GPT idea with Manuel Trajtenberg was also motivated by our studies of the contemporary economy linking computers, a GPT, to applications in a wide number of sectors. Manuel Trajtenberg had done work quantifying the benefits from product innovation in CT scanners, a computer-based medical diagnostic tool.[5] I had worked on quantifying the benefits from computerization in financial services.[6] Each of us was convinced that the gains to computerization were already large and spreading out across applications and sectors.[7] Each of us had noted that the availability of computers had enabled complementary innovation, one in health care, one in large organizations'

[1] See Bresnahan and Trajtenberg (1995).
[2] See, for example, Field (2008).
[3] The macroeconomics and growth literature on GPTs is well surveyed in Jovanovic and Rousseau (2005), while the micro-theoretical literature and much of the historical work has been summarized in Lipsey et al. (2005). Of course, anyone interested in this area should read Helpman (1998), especially the introduction.
[4] Such as Landes (1969), Rosenberg (1976, 1982).
[5] See Trajtenberg (1990a).
[6] See, for example, Bresnahan (1986).
[7] I will return to the role of computers as a GPT below.

accounting applications. An obvious question was how to assess the benefits from product innovation in a particular sector or market, given that a substantial part of what's going on is that the innovative product is benefiting from innovation in one of its key components, say computers (or semiconductors). In other words, how much can be causally attributed (in terms of the source of benefits) to innovation in CT scanners or in finance company operations *per se*, as opposed to innovation in the computers that go into it? A related but distinct question was how and why GPTs emerge. Do important general principles, perhaps from science, create technological opportunity that can be widely exploited? Does work on critical demand needs induce technical progress of general importance? Or is the process of market invention that leads to GPTs more complex than either of those views? Together with the historical observations that problems of coordination and of slow diffusion (perhaps because of the need for complementary investments) this led us to GPTs.

These dual motivations led to a basic definition of GPTs with three parts: a GPT (1) is widely used, (2) is capable of ongoing technical improvement, and (3) enables innovation in application sectors (AS).[8] The combination of assumptions (2) and (3) is called "innovational complementarities" (IC).

More precisely, IC means that innovations in the GPT raise the return to innovations in each AS and *vice versa*. Label the technological level in the GPT as T_G, and its rate of change as $\dot{T}_G$. Similarly, label the technological level in a typical AS, *a*, as T_a, and use the same dot notation for its rate of change. We can then write the rate of change in the social return to innovation realized in applications sector *a* as a function of GPT and AS technological change plus other causes (*X*): $\dot{V}_a(\dot{T}_G, \dot{T}_a, X)$. I focus only on the social return, and leave implicit the underlying components of cost reduction in the *a* sector, improvements in product quality or variety in the a sector, and so on.[9] Further, because the $(\dot{T}_G, \dot{T}_a)$, are fixed costs, if we make the economy larger by increasing all production and consumption by the ratio μ, the rate of growth of social benefits in each AS increases to $\mu\dot{V}_a(\dot{T}_G, \dot{T}_a, X)$.

The first point to note about this is that if the functions $\dot{V}_a$ have increasing differences in $(\dot{T}_G, \dot{T}_a)$, a present discounted aggregate welfare measure will have social increasing returns to joint technology investment in the GPT and all the AS.[10] Over the relevant range, the social returns to the entire GPT cluster, $\Sigma_a \int[(\dot{V}_a(\dot{T}_G, \dot{T}_a, X)e^{-rt}dt)]$, will be larger if all of $(\dot{T}_G, \dot{T}_a)$ are increased together, and will be larger if all of them are increased in a coordinated fashion than if there is not a coordinated increase. What is important here is that the social increasing returns to scale (SIRS) arise across the entire *cluster* of technical change in the GPT and technical change in the AS.

A related macroeconomic growth point arises when the span of the applications sectors in a GPT cluster covers much of the economy. Then the SIRS associated with a GPT are economy-wide increasing returns. Thus GPT models fall within the class of models which can have sustained aggregate growth.[11]

[8] This definition has been refined and improved. See Helpman and Trajtenberg (1998a). A far longer statement of this definition and careful thought about boundaries can be found in Lipsey et al. (2005). The question of what precise time period is the "age" associated with a GPT is taken up in Jovanovic and Rousseau (2005) who also consider definitional alternatives.

[9] Fully specified macro growth models derive $\dot{V}$ from the aggregate consumer welfare measure of the entire economy. In a slight change of notation, I bridge between the micro and macro notation by emphasizing the part of aggregate welfare determined in sector *a*.

[10] The function $\dot{V}_a(\dot{T}_G, \dot{T}_a, X)$ has increasing differences in $(\dot{T}_G, \dot{T}_a)$ if for all X, $t > t'$ and $u > u'$, then $\dot{V}_a(t', u', X) - \dot{V}_a(t, u', X) \geq \dot{V}_a(t', u, X) - \dot{V}_a(t, u, X)$.

[11] As Romer (1986) points out, only with increasing returns can an aggregate economy grow at a constant or increasing rate. He emphasizes technical change as a source of increasing returns.

If the technological opportunity associated with $\dot{T}_G$, $\dot{T}_a$ can be sustained over a range of increases in the technological levels and thus sustained over a period of time, a GPT can continue to provide the backbone of that growth. If the GPT itself or any of the associated $\dot{T}_a$ are embodied in capital goods or are complementary with capital investment, then this technical progress will be associated with continued growth of the economy's capital stock. This growth, $\dot{K}$, arises from a technical-change induced increase in the marginal value product of capital and thus is itself associated with resolving the limits to growth in constant-returns models.

As we shall see below, the key elements relating the cluster of a GPT and innovation in its applications to aggregate growth are a mixture of the obvious and the much less obvious. Like many models of innovation, a GPT cluster can overcome diminishing returns because innovation is inherently an increasing returns activity. Obviously, if a GPT has economy-wide scope, the relevant increasing returns also matter at the aggregate level. Less obviously, a GPT can trigger sustained innovation over a period of time because of the positive feedback between GPT and AS.

1.2. Industry structure, organization, and incentives

The basic structure of a GPT leaves unspecified a large number of details of how GPT and different AS are supplied, of their economic characteristics, and of the nature of interactions between GPT and AS innovators.

The basic GPT structure could be mapped into goods and markets in any of a number of ways. The GPT could be disembodied knowledge (as in the example of the factory system or mass production), or it could be embodied in a good or service that is purchased by the applications sectors (as in computing). If it is embodied in a capital good, that good could be bought by the applications sectors (like a computer or an electric motor) or, alternatively, services of that capital good could be sold by a GPT firm to each AS (like railroad tracks). These alternatives are related to but distinct from the question of how invention in the GPT and the AS are financed. The GPT can be in the public domain, controlled by a single firm with a patent or trade secret, or supplied by a large number of different firms each of which has distinct versions. The same set of alternatives applies to the AS; the applications technology can be disembodied (or not) protected by patents or trade secrets (or not) and supplied to the AS by a specialist firm or firms (or not.) The invention in the AS can be undertaken by each firm in the AS, or a specialist may emerge to supply a technology-bearing good.

There is a parallel question of timing of investment. Consider the difference between a railroad and a steam engine. A railroad line must be invested in before any customers can be served; the corresponding investment in a steam engine occurs on a customer-by-customer basis. More generally, there could be any of a wide number of relationships of the timing of initial investment and the sunkness of investment in both AS and GPT. As we shall see, these timing distinctions have important implications for the social return to a GPT and for the role of GPTs in long-term swings in productivity growth.

The relationship of the GPT to the AS' invention can also take many forms. There could be a technical spillout from the GPT to each AS, as in the case of a GPT which is an inventor's tool. There could be the enabling of technical opportunity for the AS, as in a GPT like electricity which created opportunities for plant-floor industrial engineering, even though industrial engineering and electricity are separate bodies of knowledge. There could be the enabling of market opportunity in many AS (potentially requiring innovation there) as in the case of the widespread use of the Internet. The original general idea could be invented for a specific purpose, then made general and abstract for reuse elsewhere. I emphasize the variety among potential relationships in order to point out the obvious but sometimes overlooked point that implications of specific model of interaction are not the same as implications of the basic GPT structure.

Similarly, the basic GPT structure could be mapped into a number of distinct vertical markets or organizational structures. There could be an *ex ante* contract among potential inventors in the AS sector and the GPT. Or that contract could be less formal, in that a GPT innovator "evangelizes" co-invention in the AS. Or the AS and GPT can be disconnected, as when a GPT emerges from one or two specific ASs. The potential externality associated with public R&D outcomes in the GPT sector itself can lead to creation of GPT specialist firms in order to gain some appropriability. Other structures are also possible.[12]

To move from the social returns to innovation to the private returns depends upon the detailed models of markets, of the economic institutions supporting appropriation by innovators, and of the dissemination of information about innovations. In the GPT context, this covers a great deal of ground, as there are multiple markets and multiple innovations in play.

However, there are some basic incentives which underlie any of these different structures. Accordingly, while some authors have emphasized those details, I shall work with a summary model of the returns to innovators from their innovations. Let λ_a be the fraction of the social value created in sector a to a-innovators. Similarly, let a fraction λ_G of the value created in each AS go to g-innovators, and a fraction λ_c to consumers. It is always the case that $\lambda_a + \lambda_G + \lambda_c \leq 1$, and the more typical case is the strict inequality <1 so that there is some social waste caused by the use of patents, secrecy, or some other appropriability device in either the AS or the GPT or both. I shall leave that in the background in what follows, however, as I do not treat the mechanisms determining λ_a and λ_G. The fixed-λ assumption corresponds most closely to a model in which each innovator gets a patent of fixed length and charges the price associated with the market power conveyed by that patent for the term of it. It rules out more complex structures that might result from complex economy-wide contracting among inventors in the GPT, multiple AS, and consumers, and thus corresponds to the case of a market economy. I assume that there is an efficiency frontier $E(\lambda_a, \lambda_G, \lambda_c)$ and (to avoid constant repetition) that society's innovation system is on this frontier, so that an increase in λ_a holding λ_c constant, for example, must reduce λ_G.

1.3. Social increasing returns and related externalities

This treatment of the private return to inventors in a GPT sector and in a number of application sectors has the advantage of bringing the duality of social increasing returns to the fore. Consider the private returns to invention in applications sector a, which are given by $\int \lambda_a \dot{V}_a(\dot{T}_G, \dot{T}_a, X)\mathrm{e}^{-rt}\mathrm{d}t$. A higher rate of technical progress in the GPT sector, higher $\dot{T}_G$ increases both private return to the innovator in a and the marginal return to increases in with $\dot{T}_a$ (the latter because of the increasing differences, i.e., the IC). This has dual implications.

[12] These issues have a large literature associated with them. An excellent overview can be found in Arora et al. (2001).

First, as a positive-economics matter, there are clear implications for any model in which inventors in different sectors act independently. Increasing differences mean that increases in $\dot{T}_G$ will increase the incentive for innovators in AS a to make increases in $\dot{T}_a$. Similarly, increases in $\dot{T}_a$ will increase the incentive for GPT innovators to make increases in $\dot{T}_G$. These increasing differences can overcome diminishing returns to investment in technical change only of each type, potentially over a wide range of improvements.

There are other implications of the same set of assumptions that come to the fore within the microeconomic approach. The basic assumption which makes a GPT analysis interesting is that all of the different sectors of the economy, and all the different subprocesses within the production process in each sector are quite heterogeneous. Diagnosing brain tumors and tracking/collecting accounts receivable, for example, are extremely different production subprocesses important in very different sectors of the economy. The innovation cost function of a large, heterogeneous economy can be lowered in the aggregate if there is a mechanism to share the fruits of innovative effort across some of these diverse sectors and subprocesses. The diagnosis and accounting example—the motivating example for the idea of GPTs—illustrates how technical progress in computing could, if combined with co-invention in medicine and in finance, be spread across a wide number of sectors of the economy. More generally, GPT models assume that specific intermediate inputs can be made very cheap through continued technical advance, and that those inputs are easily made useful in a wide variety of sectors and subprocesses. Ideas and thus innovations, in general, are characterized by zero marginal costs of reusing an idea. Any specific innovation may quickly run out of places in the economy where it has a positive marginal value product, that is, runs into diminishing returns. The GPT structure creates a wide scope of applications for GPT innovations, and thus a large level of social increasing returns, by using AS co-invention to avoid the problem of diminishing returns.

However, there is also an externality. While an increase in $\dot{T}_G$ gives an incentive to AS inventors to increase their innovative efforts, they pick $\dot{T}_a$ to maximize $\lambda_c \dot{V}_a$. If they were maximizing all of producer returns, they would pick the even higher level of $\dot{T}_a$ that would maximize $(\lambda_a + \lambda_G)\dot{V}_a$. This external effect is above and beyond the spillouts to consumers not internalized by inventors (which the GPT literature treats as an unavoidable cost of invention.)

This externality is symmetric. Consider the private returns to inventors in the GPT, which is summarized by $\lambda_G \sum_a \int \dot{V}_a(\dot{T}_G, \dot{T}_a, X)\mathrm{e}^{-rt}\mathrm{d}t$. A higher rate of technical progress in any a sector, $\dot{T}_a$, increases both private return to the innovator in G and the marginal return to increases in with $\dot{T}_G$ (because of IC.) So the symmetrical results hold: there is the positive prediction that increases in any $\dot{T}_a$ will causally lead innovators in the GPT to increase $\dot{T}_G$. There is also the external effect. The increase in $\dot{T}_G$ maximizes $\lambda_G \sum_a \int \dot{V}_a(\dot{T}_G, \dot{T}_a, X)\mathrm{e}^{-rt}\mathrm{d}t$, that is, it is less than the amount which would maximize producers' returns for all producers as a group.

In addition to that familiar "vertical" externality between AS and GPT inventors, there is an additional externality in a cluster with more than one AS. This might be called an "indirect" as well as a horizontal externality. The easiest way to see it is to view the problem as a game among AS inventors, pushing the G inventors into the background. To do that, let $\dot{T}_G(\dot{T}_A)$ be the behavior of the G innovator as a function of all of the inventive activity of all the AS. Then each AS has the private incentive to pick $\dot{T}_a$ to maximize $\int \lambda_a \dot{V}_a(\dot{T}_G(\dot{T}_A), \dot{T}_a, X)\mathrm{e}^{-rt}\mathrm{d}t$. But that leads to a lower choice of $\dot{T}_a$ for all a than if each AS instead maximized $\sum_a \int \lambda_a \dot{V}_a(\dot{T}_G(\dot{T}_A), \dot{T}_a, X)\mathrm{e}^{-rt}\mathrm{d}t$.

The basic GPT structure implies three general results, SIRS in economy-wide invention and the two external effects. The social increasing returns stem from the superadditivity between inventive effort in the GPT and inventive effort in the applications sectors.[13] There will be a high social rate of return to success at coordinating technical progress in the GPT with technical progress in a large number of applications sectors. As is true of any model with social increasing returns, however, there are external effects. Given the particular structure of basic GPT model, there are two external effects. There is a "horizontal" external effect across applications sectors (each application sector would like other applications sectors to invent more than is in their independent interest) and a "vertical" external effect (increases in the economic return to GPT invention at the margin imply either social waste or decreases in the return to AS invention.)

As the original motivation discussion suggests, the analysis of GPT has a number of linked purposes in economic history, growth economics, and the economics of technical change. While the original idea linked GPTs to an age: GPTs would be "steam" and the era in which the GPT creates value to be the "age of steam," this is in many ways too limiting a treatment, as we can see from writing down the model above. One obvious point to make is that the external effects and the SIRS associated with a GPT do not turn on it having economy-wide scope. If there are a substantial number of different applications sectors, but not all the economy, there still can be considerable SIRS from sharing a common input. Similarly, the problems associated with successfully achieving coordination (incentives or information for technical forecasting) would be the same even if the scope were less than economy wide. This led to one of the forks in the road for GPT analysis: there is a "microeconomic" branch which emphasizes the structure of payoffs and a "macroeconomic" branch which looks more at economic aggregates.

In micro approaches, the obvious core of the opportunity/problem contrast is achieving coordination. Absent coordination among a GPT and a number of AS, the private return to innovation in either area fall short of the social returns because of the two external effects. Better coordination leads to a positive feedback loop in which innovations in either AS or GPT raise the private incentives to innovate elsewhere in the system. These are, of course, classical features of anything which is modeled as a coordination game. The implications, embodied in either a static or dynamic context, drive the results that a GPT positive feedback loop may be slow or difficult to start but valuable once it begins.

There are a number of other purposes in the GPT literature which I leave out for purposes of space. There is a considerable literature linking GPTs to changes in factor demand, particularly to skill-biased or to antiskill-biased technical change, for example.[14] And a management literature discusses the difficulties of commercializing a GPT, for example, see Thoma (2009).

It is worth pointing out that the idea that GPTs are important for growth is distinct from the idea that advances in GPTs are themselves important inventions. It is the joint invention in GPT and many AS which creates economic value.[15]

[13] A large and very successful literature has analyzed games with the structure where increases in the activity of one agent raise the return to increases in the activity of other agents, that is, supermodular games. See Milgrom and Roberts (1990). Innovational complementarity means that basic structure of a GPT implies that most formalizations will lead to a supermodular game. Another literature which has much in common with our undertaking is the literature on standards and compatibility, which has a great deal of emphasis, as do we here, on sharing across agents. A review of this literature can be found in Farrell and Klemperer (2006).

[14] There is a large literature on the topic of GPTs and skill-biased technical change. One could start a bibliography from Acemoglu (2002), Aghion and Howitt (2002), Bresnahan (1999), and Bresnahan et al. (2002).

[15] Patent studies have created empirical definitions of an important patented invention, for example, an invention which has been cited many times. Economic historians have also brought forward empirical definitions of important inventions, for example "pivotal" inventions without which social gains would be smaller.

Finally, just as it would be a mistake to say that steam power came out of nowhere to create the age of steam, so too it would be a mistake not to note the many ideas closely related to GPTs, some much older. Clearly there is a relationship between the idea of a GPT and the idea of a technoeconomic paradigm (Dosi, 1982). Similarly, there is a relationship to the idea of a macro invention (Mokyr, 2002) and to a strategic invention (Usher, 1954). Finally, many industries have the concept of an enabling technology, by which they mean a GPT.

Another limit on the ability of GPTs to overcome diminishing returns arises because of the inherent dual nature of using sharing to achieve scale economies across diverse applications.[16] There is an element of compromise inherent in sharing if heterogeneous sectors and subprocesses do not have exactly the same ideal direction of technical progress. Diversity arises if each heterogeneous application would be best off if the shared input were exactly to its specifications (even taking into account the application's ability to undertake costly co-invention.) A single shared input can gain scale economies, but the requirement to fit a wide variety of diverse applications creates tension.

To return to the basic diagnosis/accounting example, an MRI machine and an accounting system have a common input, computing. They also have different optimal specifications for a computer in their derived demand for this input—even after we take into account the co-invention effort. Scientific and engineering applications of computers, like the MRI machines, typically have optimal specifications involving an inexpensive computer which can perform numerical calculations effectively. Business data processing applications, in contrast, have optimal specifications that put significantly more weight on reliability, large-scale data input/output operations, fail-safe maintenance of databases, and so on.

This diversity has been met in the computer industry by market segmentation, with minicomputers and mainframe computers each somewhat more optimized for scientific/engineering or business data processing applications, respectively. This market segmentation involves a limitation on sharing technical inputs. It also creates partially separate, partially overlapping, positive feedback loops. This point, which is a familiar one in industry analysis, also has implications for the analysis of invention. I shall return to these below. For now the point is that the model in which there is "a" GPT and it is seamlessly used in all the sectors and subprocesses of an economy needs to be thought through carefully.

There are several elements left out of this basic structure and exploring them has been the backbone of the literature. One essential feature for the study of long run growth and technical change is an element of time. Indeed, the literature has been very rich in adding dynamic elements to the basic structure, both empirically and historically. I review this element of the literature, which is rich and varied, in section 4, below.

2. Empirical and historical studies of past GPTs

While the basic microeconomics and macroeconomics of a GPT cluster are clear, understanding their empirical application involves resolving some definitional questions and deciding how to deal with some complexities. Since these are inherently matters concerned with application, in this section, I examine them in the context of the historical analysis of steam power and electrification.

[16] Many of these ideas have come out clearly in the standards and compatibility literature. See Gilbert (1992).

2.1. Steam

Steam is the prototypical GPT and the history of steam suggests a number of important complexities for us to consider.[17] The first point is that steam power took a very long while to diffuse across applications sectors. Starting from the eighteenth century, steam was first important in mining and then in textile manufacturing. Yet even a century and a half after the first production use of steam, steam still did not provide the majority of power used in textile manufacturing.[18] Steam power diffused to other manufacturing industries yet more slowly. Steam was used in transportation (especially in ships and in railroads). As in manufacturing, the within-sector diffusion in transportation was a long, slow process. Steam-powered ships replaced some sailing ships quickly, notably in uses where wind was unreliable or reliability was extremely valuable, but sailing ships persisted in other uses into the twentieth century. As in the economics of technology generally, the speed of diffusion is central in the study of GPTs. As we shall see, the literature has taken up the question of slow diffusion of GPTs in detail.

One reason for the slow diffusion of steam was supply constraints. Early steam power had profound limitations. It could not provide continuous rotary motion, for example. This thwarted applications where mechanization was central. After at least two major improvements in the steam engine (from Newcomen and Watt), rotary power was at last possible. A second limitation of early steam power was the problem of process control. Until steam engines could provide power that could be predicted, stable, and steady, that is controlled, it would be unsuitable for applications that could not use jerky or otherwise unreliable power. The invention of the Corliss steam engine and its ongoing improvements provided much more controllable steam engines, long after steam power came into productive use. The basic definition of a GPT emphasizes this capability for ongoing improvement, of course.

Improvement in steam power involved a wide range of different "technologies" in the engineering sense, that is, a wide range of different bodies of knowledge. These included but were not limited to the science-based idea of steam power itself (the discovery that water and steam were the same substance, the working out of the relationship among temperature, pressure, and volume, etc.). Improvements in materials (stronger boilers), improvements in mechanical understanding, improvements in fuels, and so on, all played a part. The complementary nature of these different technical inputs within the GPT itself (steam in this case) occurs over and above the complementarity between improvements in steam power and improvements in using industries which take the form of innovations to take advantage of steam power. In the case of steam, the economic incentive to supply and improve the complementary inputs to the steam engine itself was at work for a long period before the steam engine had suitable features to enable a wide range of innovations in complementary activities.[19]

[17] My account of steam power draws heavily on (*inter alia*) von Tunzelmann (1978), Lipsey et al. (2005), Landes (1969), Crafts (2004), Crafts and Mills (2004), and Rosenberg and Trajtenberg (2004).

[18] The century and a half is from Savery's introduction of a steam-powered pump at the end of the seventeenth century to the state of power in textiles as of 1850 (reported by Landes, 1969).

[19] This is the useful distinction between "backward linkages," for example, the incentive to create better complementary components or inputs to the steam engine, and "forward linkages," for example, the incentive to create complementary applications. The observation that historically backward linkages predate forward linkages in steam is of course from von Tunzelmann (1978).

Steam power is in a particularly interesting relationship to coal, and one that illustrates a general point. It is at once true that coal is a fuel for steam engines and that early steam power was particularly important in mining, including in the mining of coal. The dependence of coal on steam and of steam on coal immediately suggests the importance of a general equilibrium analysis in which technical and market advance in coal and in steam engines are jointly determined. Given that the base idea of a GPT bridges from the analysis of technology to the analysis of society's growth needs, one should of course expect general equilibrium effects. The literature on macroeconomic implications of GPTs has, as I pointed out above, emphasized the scale-economies-relevant point that general equilibrium effects can arise through endogeneity of the size of the entire economy. Here we see a related general point that arises on a wide scale if not all of GDP. In the case of coal and steam, general equilibrium effects do not arise through size of the entire economy, but rather through the market size and technical state of the "modern" economy. This general point relates to such other general equilibrium concepts (if not yet as wide as the whole economy) as the size and technical level of the "industrial" economy, of the "postindustrial" economy, and so on.

The nature of the improvements in steam power that loosened the supply constraint on diffusion brings us to another general point about technical progress and thus about GPTs. These improvements in steam power did not merely take the form of lower costs for an existing set of product characteristics. Instead, the available range of steam engine product characteristics widened. Over a long period of time, there were a number of changes in important product characteristics, above and beyond the two very important ones (rotation, control) just mentioned. Improvements in product characteristics generally are an important source of value creation in technical progress (Trajtenberg, 1990b). This is especially the case when the improvements in product characteristics create either the opportunity for or the incentive for complementary innovations (Bresnahan and Gordon, 1997). The important distinction between improvements in cost and improvements in those product characteristics which are not like cost applies to GPTs as well. In steam power, there was a particularly stark version of this general principle. Improvements in characteristics of the GPT permitted a wider range of applications through enabling complementary innovation.

Another set of general points that arise in the case of steam concern the relationship of a GPT to earlier technologies. Steam replaced water, wind, and muscle power, but not immediately and not everywhere. Each of those earlier power sources had had centuries of technical progress. Wind and water power are unreliable, an important source of market opportunity for steam. More generally, the weaknesses of the preexisting technologies relative to the new GPT are an important determinant of the GPT's early scope of application, as are the degree to which existing complements are specific to old technologies. Second, a preexisting technology may not be static, and improvements in it may delay the diffusion of a new GPT. Those improvements can be accelerated by competition or learning by the old technology from the new GPT.

A second general point revealed by the case of steam is that the preexisting technologies may have complementary inventions in some or all of the potential applications sectors for a new GPT. These may be sunk investments, or the old AS invention may be portable to the new GPT. In the case of steam, portability of some preexisting complementary investments showed an element of dynamic complementarity between an old technology and a new GPT. We can see an illustrative example in the use of water power in manufacturing applications which later came to use steam. The complementary inventions needed to take advantage of a mechanical power source (water) in some cases could be adapted to a new one (steam.) The new power source could be used where and when it was cheaper or more valuable, as steam could be when the water was not flowing (reliably).[20] More generally, if the complementary innovations used with an earlier technology are not (prohibitively) specific to it, they can also become complements of the new GPT. This creates a kind of dynamic complementarity between a new GPT and the older technologies it replaces. The costs and delays associated with inventing (some) complementary applications for a GPT will be lessened to the degree that those applications were already developed for an earlier version.

This dynamic complementarity should not come as a surprise. Much of technical progress is recombinant. Recombination of existing complementary innovations with a new GPT is just one version of that.[21] More importantly, recombination tells us to avoid thinking too much about the "importance" of any particular technology. That is not to say that steam or other GPTs are unimportant. Instead, it is to say that what generates real economic progress is the cluster of a GPT and applications innovations, which together bridge technical opportunity and market opportunity. From one perspective, steam is just another power source. From another perspective, the partial transition from wind, water, and muscle to steam opened up a new avenue of ongoing improvement, one which had a significantly better future in some ways than any of the preexisting technologies.

There is an irony in the observation that the transition from wind, water, and muscle to steam, and later to electricity, created the opportunity for sustainable growth. Many observers today believe that long run growth of the economy can only be sustainable if it involves a transition away from the carbon fuel sources used for steam and electricity. What this reflects is a deep economic observation: what sustains a growth path depends on the underlying economic conditions. When a fundamental constraint facing the economy was the limits of muscle power and the limits, given the technology of the day, of steam and water power, successful exploitation of fossil fuels was central to sustaining long-run growth. Generations later, we have discovered a cost of further growth we did not know we had, the climate costs of carbon dioxide in the air. Given that technical knowledge is at a much more advanced state today, particularly technical knowledge about the manipulation of physical production processes, I would anticipate that the most difficult problems associated with carbon in the atmosphere are policy formation rather than technical change and that the growth limitations associated with carbon in the atmosphere will, after some twenty-first century technical progress, not appear all that important.

While steam had a long slow start as a GPT, it then had a takeoff. Nathan Rosenberg and Manuel Trajtenberg have argued that supply constraints which had been holding back the diffusion of steam to application sectors were relaxed with the invention of the Corliss steam engine in the mid-nineteenth century, triggering the Victorian "age of steam" (Rosenberg and Trajtenberg, 2004). Corliss' inventions, especially a better system of valves for the steam engine, led to a price reduction by improving steam engine's energy efficiency and to a quality improvement by permitting control of the engine to achieve a continuous uniform power flow. Rosenberg and Trajtenberg argue that these changes, especially the quality improvement, made a much wider range of manufacturing applications possible.

Rosenberg and Trajtenberg provide a careful econometric examination of the diffusion and use of the Corliss engine in the United States in the late nineteenth century. It is important to point out at the outset

[20] An extreme version of this complementarity can be seen in the use of steam power to lift water from a storage pond to the intake of a water wheel.

[21] See Weitzman (1998) for a discussion.

that Rosenberg and Trajtenberg—like many of the research papers we will see below—do not have data which directly measure co-invention. Their investigation of co-invention at the leading manufacturing adopters of steam (which they list) is limited to discussion of particular instances. Historical methods, in their work and earlier work, linked the large-scale stationary steam engine to the successful exploitation of scale economies in manufacturing plants.[22] They do not have systematic data on the use of technologies in manufacturing complementary to steam. Instead, they use relocation of manufacturing to places where water power is limited as the observable indicator of co-invention.

This data limitation is one we shall see over and over. "Technical" GPTs create technological excitement, so there are good data about steam. There are less good data about complements to steam, such as the reorganization of manufacturing plants to use new manufacturing methods and a new power source. Rosenberg and Trajtenberg are compelled to use the observable relocation of economic activity as a proxy for its unobservable reorganization. They thus rely on data relating to the economic growth consequences of successful co-invention and on an economic model of the joint determination of adoption of steam power and of co-invention in manufacturing.

Examination of economic relationships involving complementarities is always difficult. Rosenberg and Trajtenberg have a sound attack on the problem that use of a new technology and economic development are jointly determined by actions in the application sectors and in the technology itself. They estimate a two equation system where the fundamental unit of observation is a place. One equation predicts the use of power source, steam, or water power, based on a model of the comparative advantage of those two power sources. (Water power has great advantages where it is present, but demand and/or transportation to demand may not be colocated with water power.) This gives them a reduced-form prediction for adoption of steam. They also estimate a local economic growth equation in which steam use is one predictor, but is treated as endogenous.

Rosenberg and Trajtenberg interpret the role of use of Corliss engines in local growth after 1870 in the United States as an indicator of the economic value of relaxing the locational constraint of water power. In this interpretation, they are seeing the portion of rapid economic growth resulted from moving manufacturing activity to where it was valuable. The further possibility that the co-invention of new steam using manufacturing plants could itself be valuable technical progress cannot be separately identified using their methods.

The example of the Corliss steam engine shows also one of the important features of a GPT, which is that ongoing improvements in it permit a renewal in complementary invention in complements. This is a central element of theories in which GPTs are part of sustained technical progress or growth.

There are a number of broad economic growth trends of the late nineteenth and early twentieth century closely linked to the shift in power sources which so far have not been and probably cannot be studied in a careful econometric analysis. The change in power source is also a complement for the growth of manufacturing's share in final output (new consumer and investment goods, or better or especially cheaper variants of existing consumer and investment goods) for organizational innovations such as the factory system, interchangeable parts, mass production, and so on. The larger scale factory in which the steam "prime mover" was efficient was complementary to the creation of an effective transportation system and of mass marketing. Steam power is thus part of—but one would be unwise

[22] Rosenberg and Trajtenberg also examine the relationship between plant size and steam power within a few important industries systematically, confirming the results of earlier case studies.

to say a main cause of—a transformation of the economy toward "scale-based industrialization and urbanization." In short, while the importance of steam power as a part of a cluster of related nineteenth century innovations is high, it seems unwise to attempt to parse out separately the contribution of each of those innovations.

Did steam power make an important contribution to economic growth? Only if steam power enabled AS innovation, i.e. important inventions in factory mechanization. Nick Crafts has undertaken a calculation which demonstrates this clearly, I think. Crafts' cost saving calculation compares actual expenditures *ex post* the deployment of steam to the costs of undertaking those same applications if steam power were as expensive as it had been a generation earlier. This is a fairly aggressive cost-index calculation, since it uses the quantity of steam power determined after steam has become cheaper and more capable as the baseline.[23] Perhaps that is Crafts' point, as he shows relatively modest gains in the cost-index calculation. Crafts also agrees with Rosenberg and Trajtenberg about the timing of noncost gains from steam power, linking them to the later period after the Corliss variant of the steam engine had begun to diffuse.

Another important area that is revealed by examining the history of steam closely is that there are indeed links between technical advance in the sense we usually understand it and the analysis of a GPT cluster. One point is related to the problem of mechanization. This created considerable technical progress in machines of all kinds to be used as producers' durable goods. That was complementary to new power sources; one might want to be careful, however, before assigning causation to power rather than mechanics or *vice versa*. Another point—and I would apply these both to mechanical improvements and to steam power, is the one that Lipsey et al. have called a "general purpose principle." They have done a valuable service to clarity by drawing a distinction between a "general purpose principle" and a "general purpose technology." The Corliss steam engine was a GPT; it drew on the idea the steam engine as a GPP. There are three useful points to be made in this connection. First, it is a mistake from an economic perspective to focus on a specific and narrow technology in the analysis of GPTs. Instead, the focus should be on a general purpose principle as one of the attributes of a GPT. The switch from water power to steam or from steam to electricity represents a lowering of the costs of an important intermediate input, valuable in mechanization.[24] Another general purpose principle, digitization, similarly achieved a considerable cost lowering when vacuum tubes were replaced by transistors and then by integrated circuits.

2.2. Electricity

In dealing with electricity, we face some of the same questions in deciding what to consider the GPT itself as we did in steam.[25] It includes deep scientific knowledge originally of no practical implication, and also includes the engineering knowledge needed to generate and transmit power, together with a great deal of generally useful engineering knowledge about electric devices such as design knowledge

[23] If one were to make such a calculation for computers today, for example, the "cost savings" to computers would substantially exceed GDP. As I pointed out in an earlier paper (Bresnahan, 1986), if one were to make such a calculation for mainframe computers circa 1972, the "cost savings" to computers through that "after" date would substantially exceed 1972 GPT.

[24] This point has been made powerfully with regard to the history of light in Nordhaus (1997).

[25] This section draws heavily on David (1990), David and Wright (2003), Duboff (1979), Hughes (1998), Lipsey et al. (2005), and Nye (1998).

for motors or light bulbs. The GPT cluster surrounding electricity included, of course, a wide range of kinds of knowledge in all the different AS. From a growth perspective, the most important group of these was associated with application of electricity in manufacturing. Automation and mechanization of manufacturing were a major growth pole during the period of electricity's diffusion in the first half of the twentieth century.

Some important applications of electricity were, in contrast to steam, organized and distributed as a supply network with coordinated invention of a number of different technologies. The telegraph, drawing on earlier inventions, was itself invented as a communications system. So too was the telephone, later on. The city electric light company, with generation system, distribution system, lights at the end of the wires, and other technologies, was supplied as a system. The city public transportation system build around the electric rail/trolley was another system. These industry and organizational structure differences between steam and electricity suggest a close examination of the role of economic organization of a GPT, but I will not attempt such an examination here.

At a minimum, these historical examples illustrate the varied industry structures and information structures which can accompany the founding of a GPT. Some electrical applications (lighting systems, communications system) were invented in coordinated ways as a system. This stands in contrast to steam power, where different technologies were invented far apart. Models of the information structure of invention to help us understand these differences are, to date, lacking.

The electrification of manufacturing plants and processes had invention divided between applications sectors and GPT inventors. Richard Duboff (1979) made an analysis of the diffusion of electric power into American manufacturing and of the co-invention of applications in manufacturing. His analysis, while resolutely specific to the time and technologies he was writing about, and without any claim of generality, nonetheless has a number of important observations which can help us understand diffusion of a GPT.[26] As a central idea, Duboff writes about the advantages of using electric motors in manufacturing plants with a fab-assembly production process.[27]

To be sure, at the beginning electric motors replaced steam engines in many manufacturing plants without much new co-invention. The complex system of belts and drives used to move hard-to-divide steam power could also transmit the power of a large electric motor. Where (typically because of fuel prices or transportation difficulties) electricity was cheaper, it could modularly replace steam power in existing plants. We thus see, once again, the dynamic complementarity between an old technology and a new GPT.[28] However, if electricity were no more than a cheaper power source, its importance would be less than a GPT.

[26] Duboff (1979) refers to electric power as a "strategic invention" in the sense of Usher (1954), but it is clear that his implicit definition is clearly quite close to the modern definition of a GPT. The availability of motors, lights, and heat equipment after about 1890 was ready to permit "radically different procedures" in production, that is, co-invention. Without benefit of modern theories of network effects and externalities, Duboff nonetheless wrote about how a "chain reaction" arose among adopters.

[27] Electrical power other than electric motors would be used in other parts of manufacturing, for example, as a source of heat. This illustrates one of the scope-of-coverage difficulties of GPT analysis. Electrical power? The electric motor? A similar problem arises in modern technologies. Integrated circuits? The computer?

[28] As in the case of steam replacing water power, there was also an ironic interim stage in which steam power at an individual manufacturing plant would be deployed to make electricity, which would be distributed to electric motors within the plant formerly powered by steam.

It was the ability of electrical power to enable co-invention, however, which played a role in a productivity growth boom. Electric motors could be made much smaller than steam engines. This enabled the distribution of the power source to specific locations within the manufacturing plant. This distribution came to be called "unit drive," that is, one power source for each use. The advantages of unit drive were that it permitted industrial engineers to redesign plants to follow the logic of the manufacturing process. Beforehand, they had needed to trade off organization for process against organization to deal with the difficulties of mechanically distributing steam power. Relaxing this constraint enabled invention of new manufacturing processes. Over a period of time, that innovation took root, first in new plants, then in older plants, and involved much larger cost savings than the narrow cost advantage of electricity over steam.[29]

Duboff's analysis of the "stages" of cost reduction through the use of electric motors in manufacturing captures the idea that costs and delays in co-invention can slow diffusion even where technical advance in the GPT itself is speedy. Duboff examines the path by which electric power led to productivity gains in manufacturing, pointing out that invention in the AS (individual manufacturing industries) lagged behind invention of the GPT itself, as the effective exploitation of electric power in each distinct manufacturing industry called for complementary invention in the production process.

The noncost advantage of electrical motors, particularly at small scale, implied another need to share and another public good problem. If small plants were to be electrified, they could share generating equipment (which had much larger efficient scale) only with an electrical distribution system in place. Thus, even in those areas of application of electricity where applications sectors succeeded in inventing, the horizontal externality continued to be problematic for a time. By mechanisms like this, there was repeated positive feedback between invention in the GPT, electricity, and invention in manufacturing AS.

As David and Wright (2003) point out, these forces, together with important changes in the labor market, explain much of the timing of the productivity boom associated with electricity in manufacturing. And as they, like David (1990), point out, a number of very interesting ideas emerge from the historical studies which can illuminate the modern economic situation. I would summarize the lessons of the steam and electricity eras as linked (1) to a number of areas whose examination makes the idea of a GPT cluster more precise and applicable (and which generally represent the hard work of actual application) and (2) to a number of other areas which needed further development conceptually.

A central idea in these histories is that a GPT may have attributes other than cost, and that the transition from one GPT to another or the creation of an important new version of a GPT may involve changes in attributes other than cost. For example, the transition from water power to steam power freed manufacturing activity from the constraint of needing to locate next to water. Similarly, the later transition from steam to electricity relaxed a minimum scale constraint, permitting "fractional power" to be distributed closer to where it was needed on the factory floor. Similar remarks could be made about many other transitions. The broad general point is that expanding the range of application of a GPT via noncost improvements can trigger valuable new co-invention.

Indeed, new and improved versions of an existing GPT do more than just reduce cost. Two examples which have been studied in detail are a specific new version of the steam engine, the Corliss steam engine, and the fall in entry barriers which permitted a number of new forms of computer to compete

[29] The advantages included such things as making one-story rather than multistory plants. Clearly, full diffusion of such advantages will take a long time.

against IBM mainframes.[30] In each case, the new version of the GPT triggered a new round of innovation in using sectors and expanded the range of uses of the GPT.

There are two welfare implications of a new and improved GPT, whether it is on an entirely new technical basis (electricity vs. steam) or within a technology (Corliss vs. earlier steam engines.) One is that the change in product characteristics adds value above and beyond a decrease in cost. This is a familiar "new goods" analysis. To take an example from transportation, an automobile or a truck has different characteristics than either a horse-drawn conveyance or a railroad. Costs are lower than a horse-drawn alternative, but not uniformly: costs are dramatically lower for middle distances and for frequent use, for example. Costs are higher than for a railroad, but flexibility in routing is far better for the truck or auto. These improved product characteristics mean that the new transportation GPT adds more value than cost lowering—it adds value through the area under the demand curve for autos and trucks.[31] A second welfare implication of a new version of a GPT arises from new innovation which is enabled in applications sectors. Rosenberg–Trajtenberg, Bresnahan–Greenstein, and Bresnahan–Gordon point out the value of this for the three examples just discussed, and David and Wright (2003) point out the value of this incremental innovation for electrification.

Perhaps the most important confusion about the scope of the GPT idea and analysis comes from a focus on the narrow and technical in defining "technologies." This is not unrelated to the problem (more precisely, the error) of interpreting the idea of a GPT as being about technically led or science-led technical progress.

The first point is that, from an economic perspective, technical progress is defined in terms of the location of the production function. If X lets society produce the same output with fewer inputs, or more or better output (new goods) with existing inputs, X is technical progress. Thus improvements in management, whether "technical" or not, are technological progress. Process inventions, product innovations—they all count.

This is why the list of potential GPTs made by Lipsey et al. (2005) contains such organizational technology examples as the factory system, mass production, and the Toyoda system (sometimes known as lean production). Each of those was, at one time, widely used in manufacturing, was a mechanism for inventing applications in a wide number of different industries, and underwent improvement over time.

Indeed, if we look at some of the research papers which have examined the close relationship between a GPT and its applications, it is not at all clear that the focus should not be on management innovations rather than narrow technologies. Certainly the papers on steam and electricity we just saw emphasized reorganization and relocation as much as "engineering" in a narrow sense. As we shall see, this is a central idea of the modern era of the application of information and communication technology as well (see Bresnahan and Greenstein, 1996).

Another set of interesting issues arising from the historical studies concerns the role of groups or related complements and groups of related GPTs. A large number of different complementary "technologies" in the engineering sense are in each of the GPTs just studied. That would be true of modern GPTs in ICT as well. We could think of a long list of different engineering technologies in a modern computer connected to a network, from the semiconductor to software to communications technologies.

[30] For the steam engine example, see Rosenberg and Trajtenberg (2004). For the computing example, see Bresnahan and Greenstein (1996).

[31] See Bresnahan and Gordon (1997), which has a good deal of material on complementarities in innovation.

This raises a number of interesting questions about how to set the scope of analysis. Do we focus on the semiconductor (important in missiles, aircraft, hearing aids, and many more things) or on the invention based on the semiconductor which is based on the integrated circuit (important in computers, communications, automobiles, and many more things) or do we focus on the computer or the communications system or both? If on computers, do we focus on business data processing, on the PC, the minicomputer, or all? These questions are raised by the historical studies. So, too, is the guiding principle of answering them. One cannot resolve these scope questions looking only at technical concerns; they are ultimately questions about markets and economic organization.

There are also a number of points emerging from the history which are more than improvements to the basic definition of a GPT or ideas about how, precisely, to apply it. The largest of these concern time. The delay in the arrival of a GPT in creating value, its slow diffusion, and its ability to deliver sustained advantages over decades or even substantial parts of a century are important topics I shall treat below.

2.3. *Other historical studies (and more ideas)*

Another very interesting example of the microeconomic branch is Nathan Rosenberg's (1998) work on the creation of the chemical engineering discipline. The span of applicability of chemical engineering is comparatively narrow in that the applications sectors are primarily in petroleum and petrochemicals. However, as he points out, the crucial event in the founding of the chemical engineering discipline was the creation of a set of valuable general pieces of knowledge about the design of (chemical processing) plants in those industries. Those plants are highly heterogeneous, making different products, albeit all chemical. The knowledge was neither contained in the preexisting discipline of chemistry nor, in its general form, understood by the existing body of plant designers. Rosenberg makes a convincing case that this generalization was valuable, and that it involved the joint development of a GPT and of important advances in applications sectors. Essential to his argument that the knowledge in the engineering discipline is a GPT, however, is his point that it forms a toolkit for inventing. The people who use chemical engineering knowledge to design plants are making manufacturing process inventions.

One advantage of this microeconomic approach is that it brings knowledge creation to the foreground, rather than assuming that R&D creates knowledge. The key knowledge creation in chemical engineering involved understanding what could be made general (or abstract) and what could remain specific to the design of a particular plant making a particular chemical product. As the engineering discipline gained experience in designing plants of different types, and as plants came to make more complex and different chemicals, the general knowledge continued to improve. It used the concept of "unit operations" to refer to building blocks of process design which were common to all the different applications sectors. Inventions made in one applications sector were candidates to become general, and many did. Through this learning process, there is continued positive feedback between invention in the GPT and the various using industries, creating—on an industry-wide rather than economy-wide basis—SIRS.

Rosenberg makes the argument, by way of a quickly look at other engineering disciplines, that any toolkit for invention is a candidate to be a GPT. As long as (1) some of the specific applications inventions create opportunities for continued learning about what can be general and (2) improved general knowledge enables new specific invention opportunities, this seems to be quite right. He also points out the value of locating the GPT in a university academic discipline. Academic engineers are (in part) rewarded for priority in having ideas that they put into the public domain and (in part) rewarded through consulting contracts drawing on those public ideas, an apt incentive scheme for the learning model Rosenberg delimits.

There is another important lesson in the Rosenberg analysis. Like chemical engineering, many tools for invention arise from the practice of inventors. That is, rather than science-push (or any other general knowledge-push) the creation of tools is often need-pull. For those who tend to think of GPT analysis as a model in which the general leads and the specific follows, this is an important lesson.[32]

The creation of a GPT through invention of toolkits for invention is not limited to the academic domain. An excellent example is the creation and improvement of the database management system in computing. Co-invention in the uses of ICT often takes place at the level of the individual using firm. Considerable effort is expended by the computer and communications departments of firms in this regard. The technical part of that effort—as opposed to the "business model" part of co-invention—has been steadily improved through the creation of improving tools. Database management systems lower the costs of building and maintaining large applications systems in companies, thereby lowering the costs of co-invention.

Another important set of ideas has to do with the role of finding markets for GPTs in the beginning. We saw that steam and electricity differed in this in an important way. A very detailed study based in part on the microeconomic approach is also used by Helpman and Trajtenberg (1998a) (in a paper where the central model is a closed aggregate growth model.) Their theory allows for a number of different effects, including the aggregate demand (willingness to pay times market size) in an AS, the immediate benefit of the GPT relative to the technologies in use in the AS, development costs in the AS per needed additional innovation and the number (scope or complexity) of complementary components which would need to be innovated in the AS. Their empirical section is a historical investigation of early adoption of semiconductor technology in different industries. Their framework is one in which a GPT replaces a preexisting (general purpose) technology, so that at least at the beginning we need not think about the founding of new industries and new markets.

The conclusion reached by Helpman and Trajtenberg is that sectoral pattern of early adoption of semiconductors was driven, not by tradeoffs among the four factors identified in the previous paragraph, but by the existence of a few sectors (such as hearing aids) in which all four factors were very favorable. Perhaps more interesting, the laggard sectors (such as automobiles or telecommunications) they characterize as strongly determined by the number (i.e., scope or complexity) of complementary innovations which would be needed to incorporate the new GPT. This second conclusion is an important general lesson about GPTs. Their rate of diffusion is determined in significant part by the need to invent complementary inputs in the applications sectors.

Finally, while Helpman and Trajtenberg have a formal model with perfect information, in their empirical section they are not afraid to examine the "forecast errors" which led early observers to see certain applications sectors as likely demanders of the GPT. The forecast errors they identify focus on the potential benefits of the GPT in the AS, underestimating the size and complexity of co-invention needed to make a successful adoption. They note as a historical accident that, while AT&T's development of the semiconductor for telecomm uses involved a "forecast error" there was also a "historical accident" that linked Bell to hearing aids. The general analytical point, which has not been deeply

[32] For the general importance of this problem, see Chapter 9 by von Hippel in this volume.

analyzed in GPT models, is that at the beginning there can be very limited information about the applications of a potential GPT among technologists and about the value of using a potential GPT among applications sectors.

3. Econometric and (Further) historical investigations

A number of different studies have attempted econometric investigations related to GPTs. The difficulty of obtaining systematic data on co-invention in the applications sectors is the core problem for these studies.

3.1. Using patent data

One bottleneck in the systematic study of technology is appropriate data. In looking for data for studies of GPTs, as in many other areas, scholars have turned to the use of patent data. Patent data offer a potential path out of the problem that most data source focus on the technically exciting GPT and less on the mundane AS.

The chief advantage of patent data (beyond their existence) is that patent citations can be used to tell us of technological links between different inventions. This is the core of the empirical strategy used by Hall and Trajtenberg in their effort to identify, in patent data, modern GPTs.[33] As they acknowledge, there are a number of drawbacks to the use of patents in identifying GPTs as well. Some of these are familiar from the use of patents in other contexts. Not all innovations are patented, and many patents are associated with very little innovation. Hall and Trajtenberg (like patent-using scholars generally) do a solid job of dealing with the problem that many patents are associated with little innovation, and acknowledge the limitation that important innovations may not be captured by patents.

Here as in other contexts, the patent data are surprisingly informative. Hall and Trajtenberg identify a list of patents which are cited across a wide variety of patent classes (are general,) are widely cited and are cited themselves by widely cited patents (are fecund) and that are in patent classes which grow rapidly (are in a generally important area.) Extrema on those characteristics[34] and on certain features of the patent class provide a working definition of a patent which may be advances in GPTs. Their method points to patents in computer programming methods (particularly in object oriented programming) and to patents related to electronic commerce on the Internet. Those findings suggest that the method is stronger for identifying patents which may have GPT features within the list of objects being patented in a given era (and given the incentives for strategic patenting in that era) than for identifying GPTs generally. Yet those general, fecund, and growing patents are interesting technical advances.

Another study based on patent data illustrates the same principle. Moser and Nicholas (2004) undertake a study of electricity patents in the 1920s. They find that, compared to other patents of that era, electricity patents were broader in scope and more original than patents in other classes. (Their definition of "originality," based on multidecade-long citation lags, is an interesting idea in patent

[33] See Hall and Trajtenberg (2004) on this specific use of citations and Jaffe (1986) and the substantial literature based on his work for the idea that citations can help identify technology spillovers more generally.

[34] . . . and on other characteristics, such as having a long tail of citations, which are also examined in the paper.

analysis.) They face a severe data challenge, however. As we saw above, the important co-invention for electricity in this era were typically industrial engineering advances. These typically are not patentable. Thus Moser and Nicholas cannot observe the inventions which are related to electrification.

There is another challenge in using patents to study GPTs which is specific to this context. Patent citations measure (in theory) knowledge spillovers, not necessarily the spillovers which follow from IC. The spillover from a GPT to an AS may involve two patented inventions and yet there may be no citation linking them. Consider a (patentable) improvement in the tools used to make integrated circuits (steppers and the like) which permits successful manufacture of improved microprocessors and computer memory chips (also patented); suppose the existence of computers based on those faster chips enables new (patented) software innovations. It would be odd for the software patent to cite either the semiconductor or the semiconductor-manufacturing-machines patents. Those patents are not antecedents of the software invention in a technical-knowledge sense. Instead, their IC arises through their complementarity in the marketplace.

However, a technology can be general in the sense that it enables complementary innovations in a wide variety of sectors without a patent on that technology being cited by patented inventions in a wide variety of sectors. This distinction is particularly important for the eras which are studied by Hall and Trajtenberg (1967–1999) and by Moser and Nicholas (electricity in the 1920s). In each era, historical or econometric methods have brought forward a specific hypothesis about the innovations which are complementary to the GPT. For the modern era, computers in business data processing are an important GPT, and many of the important AS innovations associated with them are hard-to-patent changes in management practices, organizational structures, or marketing practices. For electricity in the 1920s, the important complementary innovations were in hard-to-patent plant-floor layout improvements and related industrial engineering inventions.

One area in which the discussion of GPTs has become considerably richer with further research is in understanding how and to what kind of phenomena models built around the basic structure can be applied.

3.2. *Efforts to create data in the modern era*

Studying the modern era of the application of information and communication technology in a wide number of white-collar work automation contexts has posed real data challenges. One issue is already familiar from discussions above. Data sets for counting computers, telephone lines, data switches, and other "technical" innovations at the firm or plant level are quite good, based primarily on business surveys undertaken to link salespeople to clients.[35] Data sets with information about applications are much more difficult to obtain. Some data sets have "applications software" but this often means programmer tools such as database management systems or spreadsheets. That information does not tell us what economic applications are running on the database or the spreadsheet.

Historical and case-study work reveals that the most valuable applications of ICT in large organizations over the first 50 years of computing had a number of nontechnical features. Typically these applications involved reorganization of (white-collar) work and involved new products. A credit card from a bank, for example, involves a complex new organization to market credit cards ("database

[35] Most work in economics follows Bresnahan and Greenstein (1996) in using what is now known as the Harte Hanks data. This is a large scale survey of (approximately) establishments focusing on their ICT use. Erik Brynjolffsson has done the economics profession a great favor by creating a firm-level database using these data.

marketing") to underwrite the extension of credit, avoid fraud, and to collect debt. These organizational and product features fall outside the scope of the data sets available to researchers.

The literature has attempted to overcome this limitation by three kinds of strategies. The first is "data are the plural of anecdotes." This approach uses case studies to determine the kinds of applications are being undertaken in a number of sites, and then uses observables in the data set as proxy measures for applications categories. Shane Greenstein and I used this approach to categorize applications of computing in large organizations according to the complexity of the organizational changes associated with co-invention. Another approach has been to draw on, or attempt to create, measures of organization at the firm level. Organizational measures include labor-relations practices in most of these studies.[36] More general organizational measures linked more broadly to "good management," to centralization of authority, and so on are another area of active research. Since existing data sets do not measure these variables, scholars have launched survey instruments and linked the results to the existing data sets at the firm or plant level.[37]

One finding from this literature is that the combination of information and communications technology with organizational changes is profitable and productive at the firm level. The complementarity between invention in ICT and co-inventions of new organizational structures is, of course, an important part of a GPT cluster. Much of the research effort has been focused on the question of whether investment in ICT is productive, surely one of the oddest questions ever to distract economists.[38] A smaller literature has focused on the more useful and important economic question of complementarities and co-inventions.[39] This appears to have established, as a broad general proposition, that the reorganization of work is complementary to changing the basis of white-collar work to use ICT. In calling for reorganization to make large gains, the computerization of white-collar work appears to be like the steam-based or electricity-based co-invention associated with those earlier GPTs. What kind of reorganization is optimal in which industries remains less than fully understood, however.[40] As yet, we have no story as compelling as "unit drive" for computers and communications.

A smaller literature attempts to understand the timing of adoption of ICT and of co-invention.[41] For the part of computing used in large organizations in white-collar work, it is clear that much of the delay

[36] For example, Bresnahan et al. (2002) use measures of labor-relations management practices gathered in a survey by Kochen. Ichniowski et al. (1997) undertake a detailed study of those practices and of other organizational variables in the plants they study.

[37] Bloom and Van Reenen (2006), for example, report on an international survey of manufacturing practices, one of a number of large scale surveys they and coauthors have taken to deal with the data dearth in this area.

[38] While many organizational information systems fail, the vast bulk of investment in these systems occurs *ex post* learning about whether they are going to succeed or fail, that is, consists of capacity enhancements for systems that are heavily used, maintenance of systems which are in steady use, and so on. We have in the research literature the spectacle of economists being prepared to assume that tens of thousands of firms were making repeated mistakes over decades on decisions typically involving millions or tens of millions of dollars and, in many industries, constituting the most basic choices of productive technique.

[39] Brynjolfsson and Hitt (2000) have a review of the productivity literature and the co-invention literature as well with a wide variety of citations.

[40] A long literature examines the role of centralization of managerial authority in computerized white collar work. Bloom et al. (2009) have a recent study showing that both centralization and decentralization may be complements to ICT in different circumstances. Athey and Stern (2002) take up the very interesting direction of looking at a very specific and narrow ICT use and studying not productivity outcomes but the more theoretical salient product quality improvements.

[41] See Bresnahan and Greenstein (1996) for this analysis of computer use in organizations. As Goolsbee and Klenow (2002) show, the diffusion of personal computers is driven by very different considerations, notably (in the modern period they study) the network effects associated with using the computer as a communications tool.

over the last 60 years in the diffusion of ICT has been driven by the costs of co-invention. Computer and communications equipment has very rapid price falls driven by technical progress in the GPT itself. That technical progress exploits powerful engineering and physical science opportunities, and, as science and engineering advances have grown more expensive over time, larger and larger markets for the equipment have financed more and more impressive advances. In contrast to that rapid technical progress, however, installed computer systems undertaking a productive task involve co-invention that can be much slower. Co-invention in marketing and in organizational change, for example, are vastly slower invention processes than invention of computer and communications hardware. Accordingly, it is the co-invention, not the invention, which is the bottleneck for diffusion.

That applies to only one of the GPT clusters in ICT, the one surrounding business data processing. Here the GPT itself was first mainframe computers and later servers, and the AS were as I have just described. Other uses of computers and communications equipment, such as for scientific and engineering calculations, have had far less of a problem of slow co-invention.

4. Timing and the relation to economic growth

A number of GPT theories have attempted to understand the timing of benefits to GPT and AS coordinated invention. The research goal is ultimately to understand real world phenomena, especially macroeconomic growth phenomena.

4.1. *Delay and diffusion*

One timing issue is simply delay or diffusion.[42] Many technologies are first invented, and only later adapted and adopted widely in industry. The diffusion lag between the date of invention and the date of full realization of economic gains is well established in the empirical literature for technologies generally.[43] The very idea of a GPT draws a distinction between raw technical progress (GPT invention) and the further innovation needed to create value-in-use (AS invention.) Thus it is very natural to examine the diffusion of a GPT as being a source of lags between raw invention and the ultimate productivity or output growth. Also, at least three widely analyzed GPTs, electricity, steam, and ICT (computers), showed slow diffusion at the beginning followed by acceleration.[44] Similarly, Griliches' classic study of at first slow and later accelerating diffusion is for hybrid corn, an invention for making inventions (i.e., a microeconomic GPT.)[45] Griliches shows how different rate of diffusion apply for the

[42] See Helpman and Trajtenberg (1998a). Jovanovic and Rousseau have a survey of this literature and a detailed survey of the various kinds of evidence brought to bear on the electricity and computerization GPT eras.

[43] See Stoneman (1983) for a somewhat dated but very inclusive survey and Hall (2004) for a review of the modern literature.

[44] On electricity, see *inter alia* David (1990) (though David rather oddly starts the computer age with the invention of semiconductor memory and the microprocessor, about 15 years after the earliest substantial use of computers in business data processing, he nonetheless sees slow diffusion.) On steam, see *inter alia* Crafts (2004). On ICT, see Bresnahan and Greenstein (1996) among many others.

[45] See Griliches (1957), which is a study of the diffusion rates of the different strands of hybrid corn but which also has a discussion of the economics of invention of the specific strands.

same technology in different applications depending on market conditions, clearly relevant to the diffusion of GPTs.

There are a wide number of different possible reasons for slow diffusion of technologies at the early stages followed by more rapid diffusion later on. That is, there are many theories of the classic S-shaped diffusion curve. These can involve supply constraints; value may be created in use only after a technology has fallen below a certain cost or achieved certain features. They can involve demand constraints; demanders may be heterogeneous, with lower value users (who are more numerous than higher value users) adopting later,[46] or there can be adjustment costs from, for example, learning in adoption.[47] Networks can communicate information about new technologies at speeds that vary over time.[48] When prices are falling in an anticipated way, the user cost associated with adopting a new technology includes a depreciation term (raising the true economic price of adopting.). A similar dynamic concern arises when there is anticipated technical improvement, this, too, can lead to rational delay. The user-cost theories will predict slow adoption at first followed by acceleration most simply, after a while, it is no longer rational to expect improvements and the depreciation term goes away. All of these forces arise even when there is a single technology (not multiple ones as with a GPT and AS). The forces can be quite difficult to distinguish empirically.[49]

One might have thought, before writing down a model, that the diffusion of a GPT should be doubly slow at the beginning. First, there is the slow diffusion of the GPT across the different AS of an economy. Second, once a complementary innovation has been made within a particular AS, there is the slow diffusion of that innovation across firms. Only after both source of slow diffusion have been overcome will we see acceleration.

All of the causes listed in the last two sections are subject to diminishing returns. Supply constraints, once overcome, are overcome. Once a heterogeneous body of adopters has been served, it is served. Another reason to consider the case of a GPT is that positive feedback between and among GPT and AS could sustain the period of diffusion past the problem of diminishing returns. This is the point of the GPT diffusion model (Helpman and Trajtenberg, 1998a). Helpman and Trajtenberg identify a further effect which is linked to the potential for continued improvement and the positive feedback cycle of GPTs. They build a model in which applications sectors must take time as well as resources to make complementary investments to the GPT, the level of inventive effort of either GPT or an AS can vary, and the timing of GPT innovation (and later improvements) and AS co-inventions (and later improvements, if any) are endogenous. They show that positive feedback between GPT and AS and among the

[46] The simplest theory has exogenous technical change and a distribution across potential adopters in value-in-use. Higher value users adopt earlier. If there is a unimodal distribution of user's values (with the mode toward the center of the distribution) then an S-shaped diffusion curve follows. The rapid phase of adoption occurs when the improving technology leads to adoption by the large number of users with values near the mode.

[47] If the learning is an adjustment cost and if any of what is learned by early adopters spills out to others, both slow initial learning and later acceleration can be equilibrium phenomena.

[48] This can also predict the S-shape for diffusion if information spreads through a network of users who are broadly symmetric. At first, there are few users who know of the innovation, and thus few messages about it over the network. At the end, there are many messages, but few users who have not yet heard them. In the middle, there is a burst of adoption when there are both many senders and many receivers of messages. One would in general expect information to spread slowly then more quickly (depending on the structure of the network, as long as it is near symmetric.)

[49] As these footnotes suggest, there are literally dozens of other theories from a number of disciplines. They are certainly not distinguishable using aggregate data and can be very difficult to distinguish econometrically using microdata.

AS affects the diffusion path. In particular, they show that there must be a "second wave" of improvements; after all the AS have first begun to use the GPT, they have a renewed round of co-invention and a renewed round of growth.

The basic GPT structure adds, in the Helpman–Trajtenberg model, another reason for S-shaped diffusion. There is a feedback mechanism between GPT and AS innovations. The steep part of the S, the period of rapid adoption, can arise because the innovation feedback cycle has taken off. Because of the GPT's innovation incentives, this is a size-of-market phenomenon. The feedback cycle is strong once there are sufficiently many AS adopting (or foreseeably adopting in the near-enough future for discounting not much to reduce incentives). Perhaps the most important thing about that particular theory of delayed rapid adoption is that the spread can be widely across an economy.

The Helpman and Trajtenberg analysis illustrates two important things about the role of GPTs. Since a GPT and all the related AS form a positive feedback loop, a new GPT has an important element of creating a new system of innovation. At the beginning, this will be limited in importance by (among other factors) the low state of development of the GPT technology itself and the overhang of successful solutions connected to an older technology or technologies. At the beginning, therefore, the incentives to innovate are less strong than later on. Endogenously, through the early adoption of GPTs and early co-invention in some or all AS, the incentives to innovate rise later on. Thus, in addition to the ordinary forces leading to slow diffusion of new technologies, Helpman and Trajtenberg identify a second round acceleration caused by a widespread switch to the new system of innovation in the new positive feedback loop.

It is worth pointing out that this effect arises entirely from incentives as long as some aspect of the innovation process takes time. The part which takes time could be time-to-invent as in the formal model, or discovery of new technical opportunity only after old invention, or visibility of applications only after a GPT is invented (or *vice versa*!) or anything else that introduces inevitable delay. It arises even if there is an excellent mechanism for coordination among inventors, even if there is no inherent reason for multiple rounds of innovation like "ladders," and even if there is no problem of incomplete information. The essential mechanism is that the (fully foresighted, fully informed) incentive to innovate within the new positive feedback system rises over time as the GPT takes hold.

In short, there are a number of reasons why a GPT might lag behind identification, at a broad general level, of an overlap between technological opportunity and a growth constraint. These include slow diffusion, perhaps even slower than usual because of the need to co-invent. They include the typical delays associated with waiting for lower costs or improved features in the new technology (or technologies) to cross many demanders' thresholds. They include difficulties to internalizing returns to co-invention because of fragmentation (the horizontal externality) or because of contracting difficulties between GPT and AS (the vertical externality.) Noting that many of the GPTs we have seen historically have important widely used complements, another potential source of slow diffusion is waiting for the weakest link among a long chain of complements. All of these can be analyzed using the ordinary tools of the economics of innovation.

5. Aggregate growth waves

The long run process of economic growth does not appear to be smooth. First, there are eras of higher and of lower productivity and output growth. Since the work of Paul David (1990) we have seen much renewed interest in using the analytics of "productivity slowdowns" of the distant past to understand the "productivity slowdown" of the recent past. The analogy suggested by David links the era of the late twentieth century to the period of slower growth before the First World War. Each of these slowdowns, we now know, was followed by more rapid growth (though of course how long the current boom will last into the twenty-first century is still a matter of forecast.) More broadly, there is renewed interest in long waves in output and productivity growth.

A good deal of the motivation for renewed interest in "long wave" theories is the modern "productivity paradox" about information and communications technology, encased in a timeless quip by Robert Solow. As in the late nineteenth and early twentieth centuries during the early period of the diffusion of electricity, measured productivity growth in the rich countries was clearly lower during the early period of the diffusion of computers (say 1960–1980) than beforehand, while productivity growth accelerated considerably, and in ways apparently tied to ICT, in a more recent period.

We thus have an explosion of models which attempt to explain various features of Kondratieff's "long waves" using significantly more advanced modeling tools than were available in the 1920s.[50] A number of papers, beginning with Helpman and Trajtenberg (1998b) have investigated the possibility of long macroeconomic cycles associated with the invention and diffusion of GPTs. While these models differ in their details, they all identify conditions in which the invention of a GPT can first lead to a slump as resources are devoted to it, only later creating a boom as the economic return to the invention is revealed. This literature is enormous and well summarized in another review article, so I shall not deal with the different models in detail.[51]

While it is not obvious that these will ever be tested, this has led to a rich set of models. One point is common to other models of aggregate growth, that there is a macroeconomic feedback effect between the size of the economy and the incentive to innovate. The SIRS in GPTs are more valuable in a larger economy; exploiting them may itself create a larger richer economy. Thus once a GPT cycle gets going, it may have a period of general equilibrium positive feedback. Similarly, if technology becomes available to augment the available factors of production, an increase in the size of the input market and the incentive to innovate may have a positive feedback loop. The empirical importance of these general equilibrium effects remains, however, as a challenge for future empirical work rather than a currently established fact.

The S-shaped diffusion of a GPT is one way to predict periods of more rapid technical progress at the aggregate economy level. Aggregate economies will have periods of particularly rapid technical progress whenever (1) there is a growth-macroeconomic coincidence between discovery of a GPT and a multiindustry growth need and (2) the GPT and its AS follow an S-shaped diffusion curve.

The literature has also sought to explain why there might be a *lower* level of aggregate productivity growth than normal in the period before a GPT takes off. There are at least two broad classes of such theories.

In the first set of theories, the capital or R&D investment in a GPT can occur before the payoff to the use of the GPT occurs. Given the long lead times for many GPTs discussed in the previous sections, this

[50] For a modern translation, see Kondratieff (1984). Kondratieff, looking at historical periods before the early twentieth century, saw a link between these "long waves" and new energy sources. He also cautioned against building a long wave theory primarily around "technics."

[51] One can find an impressive review of theoretical contributions in Jovanovic and Rousseau (2005), who also have a very sympathetic review of the empirical evidence of the empirical relevance of specific theories.

is a potentially large effect. The effect will be large if there are substantial early costs in the R&D or capital investment of a GPT. A long lead time for science or for pure research seems an unpromising way to obtain such a theory, as the science and pure research associated with past (and likely future) GPTs are simply not expensive enough in the aggregate. Instead, this seems a more promising line of analysis for "infrastructure" GPTs like railroads or (wire line) telecommunications where there is a substantial capital investment common to many customers. For computers, electrical power, steam, etc., it is very difficult to identify substantial costs of that form. (After electrical power generation began to be shared across smaller firms, the same effect could cut in, but I do not think this is what theories of a preelectrification slump have in mind.) To be sure, those industries all involve substantial capital investment, but the capital investment occurs (largely) at the same time as use. Indeed, as we have seen, the key feature of all of electricity, steam, and computing was that there was not all that much investment in the relevant capital goods in the early phases. This effect, therefore, seems unlikely to be of great general importance.

Another set of theories in which there is a lower productivity growth era before the takeoff of a new GPT works through anticipation that slows the start of a GPT, thus leaving the economy growing slowly, rather than through the timing of expenditures *per se*. These theories are not, in principle, implausible. In general, large scale coordination can be hard to achieve. It could be harder to achieve if there is anticipation that slows it. For example, the competition from an old technology could be particularly effective in slowing diffusion early on, and thus in imposing an external cost on the new GPT.[52] While logically possible, showing this chain of causation seems unlikely for steam or electricity. And it is very unlikely for the early phases of the use of ICT as a technology to automate white-collar work in the late twentieth century. Earlier white-collar automation technologies were unimpressive, for one thing.

Note that to predict periods of more or less rapid technical progress at the aggregate level there is no need to assume that there are periods in which growth is suppressed. Fluctuations can arise if most technical progress most of the time is incremental and slow, with intermittent outbreaks based around a GPT.

6. Concluding remarks

Two potential stories of long run come to mind, and I finish on them as they are related to modern fluctuations in aggregate growth.

The first story might be called "statistical." It hinges on the point that every economy is remarkably diverse in its technologies and in its varied sources of technological progress. Aggregate productivity growth is a mixture of local and incremental advances with general clusters of linked innovations, like GPTs. There is not an aggregate-growth-influencing GPT cluster every decade. When there is such a cluster, aggregate productivity growth is higher. When there is not, aggregate productivity is lower, but not "too low" in the sense of the theories just reviewed. Aggregate productivity growth rates will fluctuate if the local and incremental are reasonably constant over time and the general occurs only intermittently. Similarly, aggregate productivity growth will fluctuate over time if there are a large

[52] See, for example, Atkeson and Kehoe (2007). A review of a number of different models with complex dynamic elements can be found in Jovanovic and Rousseau (2005).

number of micro GPTs at all times at different stages of diffusion and if the GPTs with economy-wide scale arise only intermittently. Again, to explain cycles there is no need for a theory of the periods of lower productivity growth in which something is going wrong in periods of slow growth.

The second story puts more weight on the importance of GPTs but also does not conceptualize the period before a GPT's rapid diffusion as having unnaturally slow growth. Consider the possibility that the long run growth of the aggregate economy involves overcoming a series of constraints through innovation. At a very long time scale, fossil-fuel-based power and mechanization in agriculture and manufacturing helped relax critical constraints associated with gaining most of human-consumed heat and light directly from the sun and gaining most of productive power from wind, water, and muscle. This constraint was worth relaxing. Technical progress relaxed, over a long period of time, that growth constraint. The mechanization and later automation of much work in agriculture relaxed an important constraint, as did the mechanization and later automation of much blue collar work in manufacturing. Periods of success in mechanization and in automation were periods of rapid growth for a long period of time. But automating factory work, for example, did not emerge as a growth constraint until the factory system was in place and not until manufactures were an important piece of output.

To continue the same long time scale, once human labor in the physical sense has been automated, other growth constraints are revealed. The twentieth century saw the lack automating traditionally white-collar functions as a growth constraint. These functions include managing production but perhaps more importantly buying inputs and selling outputs. Only after there was work to manage, and only after there were input and output markets, was this an important constraint. Similarly, the problem of automating (largely white collar) work in the services sectors did not become a crucial growth constraint until services loomed large. Automating white-collar functions, to go back to the current example, did not have a high shadow value as a growth constraint until earlier economic progress had created something to manage.

My point here is not that there is some natural order of technologies. One could make the same point by noting diminishing returns. Even if the large cluster of technologies associated with the first and second industrial revolution had very large increasing returns over a wide range, eventually technologies which automated (to continue to paint in the broadest strokes) some firm functions (the blue collar ones) while not automating other firm and market functions (the white-collar ones) were going to hit diminishing returns.

Instead my point is that if the demand for widely used new technologies, such as GPTs, arises from relaxing growth constraints, it is easy to understand why there are swings in the level of productivity growth over time. As we saw above, there are many good reasons to believe that a GPT that *does* address a growth constraint will diffuse slowly. Further, there is no particular reason for a GPT associated with a growth constraint to emerge shortly after the constraint gains a high shadow value.[53] In general, we should expect a long and variable lag between the emergence of a key growth constraint and its

[53] Daron Acemoglu (2002) disagrees, making the useful remark that because it has both AS and general application, a GPT is more open to demand inducement than other kinds of technology. While I agree that AS innovations are frequently demand-induced, there are still fundamental limitations on what any particular GPT can deliver. Also, this argument does not imply the clearly false proposition that GPTs themselves arise through demand inducement associated with their most valuable applications.

relaxation through diffusion of an apposite GPT. Even under the best of circumstances in terms of general macroeconomic policy, of cultural and business norms, and of pro-innovation and progrowth policy, the lags will be long and variable.

Long swings in productivity growth can arise simply because of those long and variable lags. Do we need a theory in which there was something going wrong to explain the productivity slowdowns of the late twentieth century or of the late nineteenth/early twentieth century? An alternative explanation, arising from my second story, is that in each case relaxation of earlier constraints had begun to slow down in its growth impact and relaxation of new constraints had not yet begun to occur at a high pace. In the late twentieth century slowdown, for example, the gains to automating blue-collar work were slowing and the gains to automating white-collar work associated with computerization had yet to cut in.

Indeed, this appears to be the most attractive theory of the late twentieth century "productivity slowdown" and its later reversal. Many people (foolishly) concluded that there must have been something going wrong with firms' investments in ICT technology during the early phase of the diffusion of computers. We now know that the problem with "we see computers everywhere around us except in the productivity statistics" was not with productivity, but with looking at computers in economics departments rather than in firms. At the time of Solow's remark, the ICT capital stock was far too small to have (yet) created a growth boom, even though the private returns to use of computers were very substantial.[54]

Acknowledgments

I owe a great debt to Manuel Trajtenberg for many wonderful conversations about this topic and this chapter. I appreciate very helpful editorial advice from Bronwyn Hall. I very much thank Emily Warren for all her hard work in support.

References

Acemoglu, D. (2002). "Directed technical change". The Review of Economic Studies 69, 781–809.

Aghion, P., Howitt, P. (2002). "Wage inequality and the new economy". Oxford Review of Economic Policy 18, 306.

Arora, A., Fosfuri, A., Gambardella, A. (2001). Markets for Technology: The Economics of Innovation and Corporate Strategy. MIT Press, Cambridge, MA.

Athey, S., Stern, S. (2002). "The impact of information technology on emergency health care outcomes". The RAND Journal of Economics 33, 399–432.

Atkeson, A., Kehoe, P.J. (2007). "Modeling the transition to a new economy: Lessons from two technological revolutions". American Economic Review 97, 64–88.

Bloom, N., Van Reenen, J. (2006). "Measuring and Explaining Management Practices Across Firms and Countries". NBER Working Paper.

Bloom, N., Garicano, L., Sadun, R., Van Reenen, J. (2009). "The distinct effects of Information Technology and Communication Technology on Firm Organization". NBER Working Paper.

Bresnahan, T.F. (1986). "Measuring the spillovers from technical advance: Mainframe computers in financial services". The American Economic Review 76, 742–755.

[54] On the high private returns to investment in business data processing in banking before 1972, see Bresnahan (1986).

Bresnahan, T.F. (1999). "Computerisation and wage dispersion: An analytical reinterpretation". The Economic Journal 109, F390–F415.

Bresnahan, T.F., Gordon, R.J. (1997). "The economics of new goods: An introduction". In: The Economics of New Goods. University of Chicago Press, Chicago, pp. 1–25.

Bresnahan, T., Greenstein, S. (1996). "Technical progress and co-invention in computing and in the uses of computers". Brookings Papers on Economic Activity Microeconomics 1–83.

Bresnahan, T.F., Trajtenberg, M. (1995). "General purpose technologies: Engines of growth?" Journal of Econometrics 65, 83.

Bresnahan, T.F., Brynjolfsson, E., Hitt, L.M. (2002). "Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence". The Quarterly Journal of Economics 117, 339–376.

Brynjolfsson, E., Hitt, L.M. (2000). "Beyond computation: Information technology, organizational transformation and business performance". The Journal of Economic Perspectives 14, 23–48.

Crafts, N. (2004). "Steam as a general purpose technology: A growth accounting perspective". The Economic Journal 114, 338.

Crafts, N., Mills, T.C. (2004). "Was 19th century British growth steam-powered?" The Climacteric Revisited 41, 156.

David, P.A. (1990). "The dynamo and the computer: An historical perspective on the modern productivity paradox". The American Economic Review 80, 355–361.

David, P., Wright, G. (2003). "General purpose technologies and surges in productivity: Historical reflections on the future of the ICT revolution". In: David, P. A. T. (Ed.), The Economic Future in Historical Perspective. Oxford University Press, Oxford.

Dosi, G. (1982). "Technological paradigms and technological trajectories: A suggested interpretation of the determinants and directions of technical change". Research Policy 11, 147–162.

Duboff, R. (1979). Electric Power in American Manufacturing, 1889–1958. Arno Press, New York.

Farrell, J., Klemperer, P. (2006). "Coordination and Lock-In: Competition with Switching Costs and Network Effects". Competition Policy Center.

Field, A.J. (2008). "Does Economic History Need GPTs?" SSRN eLibrary.

Gilbert, R.J. (1992). "Symposium on compatibility: Incentives and market structure". The Journal of Industrial Economics 40, 1–8.

Goolsbee, A., Klenow, Peter J. (2002). "Evidence on learning and network externalities in the diffusion of home computers". The Journal of Law and Economics 45, 317–343.

Griliches, Z. (1957). "Hybrid corn: An exploration in the economics of technological change". Econometrica 25, 501–522.

Hall, B. (2004). "Innovation and diffusion". In: Fagerberg, J., Mowery, D., Nelson, R.R. (Eds.), Handbook of Innovation. Oxford University Press, Oxford.

Hall, B., Trajtenberg, M. (2004). "Uncovering GPTs Using Patent Data". The Journal of Economic History 64(1), 61–99.

Helpman, E. (1998). General Purpose Technologies and Economic Growth. MIT Press, Cambridge, MA.

Helpman, E., Trajtenberg, M. (1998a). Diffusion of General Purpose Technologies. MIT Press, Cambridge, MA.

Helpman, E., Trajtenberg, M. (1998b). A Time to Sow and a Time to Reap; Growth Based on General Purpose Technologies. MIT Press, Cambridge, MA.

Hughes, T.P. (1998). Rescuing Prometheus (first ed.). Pantheon Books, New York.

Ichniowski, C., Shaw, K., Prennushi, G. (1997). "The effects of human resource management practices on productivity: A study of steel finishing lines". The American Economic Review 87, 291–313.

Jaffe, A.B. (1986). "Technological opportunity and spillovers of R&D: Evidence from firms' patents, profits, and market value". The American Economic Review 76, 984–1001.

Jovanovic, B., Rousseau, (2005). "General purpose technologies". In: Aghion, P., Durlauf, S.N. (Eds.), Handbook of Economic Growth, Vol. 1B. © 2005 Elsevier B.V.

Kondratieff, N. (1984). The Long Wave Cycle. Richardson and Snyder, New York.

Landes, D.S. (1969). The Unbound Prometheus: Technological Change and Industrial Development in Western Europe from 1750 to the Present. Cambridge University Press, London.

Lipsey, R.G., Carlaw, K., Bekar, C. (2005). Economic Transformations: General Purpose Technologies and Long-Term Economic Growth. Oxford University Press, Oxford; New York.

Milgrom, P., Roberts, J. (1990). "Rationalizability, learning, and equilibrium in games with strategic complementarities". Econometrica 58, 1255–1277.

Mokyr, J. (2002). The Gifts of Athena: Historical Origins of the Knowledge Economy. Princeton University Press, Princeton, NJ.

Moser, P., Nicholas, T. (2004). "Was electricity a general purpose technology? Evidence from historical patent citations". The American Economic Review 94, 388.

Nordhaus, W. (1997). “Do real output and real wage measures capture reality? The history of light suggests not”. In: Bresnahan, T., Gordon, R. (Eds.), The Economics of New Goods. University of Chicago Press, Chicago.

Nye, D.E. (1998). Consuming Power: A Social History of American Energies. MIT Press, Cambridge, MA.

Romer, P.M. (1986). “Increasing returns and long-run growth”. The Journal of Political Economy 94, 1002–1037.

Rosenberg, N. (1976). “Technological change in the machine tool industry”. In: Rosenberg, N. (Ed.), Perspectives on Technology. Cambridge University Press, Cambridge, pp. 9–31.

Rosenberg, N. (1982). Inside the Black Box: Technology and Economics. Cambridge University Press, Cambridge [Cambridgeshire]; New York.

Rosenberg, N. (1998). “Chemical engineering as a general purpose technology”. In: Helpman, E. (Ed.), General Purpose Technologies and Economic Growth. MIT Press, Cambridge, MA, pp. 167–192.

Rosenberg, N., Trajtenberg, M. (2004). “A General-Purpose Technology at Work: The Corliss Steam Engine in the Late-Nineteenth-Century United States”, The Journal of Economic History, 64 (1), 61–99.

Stoneman, P. (1983). The Economic Analysis of Technological Change. Oxford University Press, Oxford [Oxfordshire]; New York.

Thoma, G. (2009). “Striving for a large market: Evidence from a general purpose technology in action”. Industrial and Corporate Change 18, 107–138.

Trajtenberg, M. (1990a). Economic Analysis of Product Innovation: The Case of CT Scanners. Harvard University Press, Cambridge, MA.

Trajtenberg, M. (1990b). Economic Analysis of Product Innovation: The Case of CT Scanners. Harvard University Press, Cambridge, MA.

Usher, A.P. (1954). A History of Mechanical Inventions (Rev. edn). Harvard University Press, Cambridge.

von Tunzelmann, G.N. (1978). Steam Power and British Industrialization to 1860. Clarendon Press, Oxford [Eng]; New York.

Weitzman, M.L. (1998). “Recombinant growth”. The Quarterly Journal of Economics 113, 331–360.