SIEPR Discussion Paper No. 09-022

Reallocating innovative resources around growth bottlenecks

By
Timothy Bresnahan
Pai-Ling Yin
Stanford University

July 2010

Reallocating innovative resources around growth bottlenecks

Timothy Bresnahan

Pai-Ling Yin*

Abstract

Economy-wide increasing returns to scale embodied in a general purpose technology (GPT) and its applications are often a key source of long-run growth. Yet the successful exploitation of increasing returns calls for coordination on a particular technological direction, reducing flexibility and choice *ex post* and potentially creating a growth bottleneck. When the bottleneck is controlled by a single firm, the resulting entry barriers reduce the ability of demanders to choose superior, alternative technologies. We examine how such a growth bottleneck can eventually be overcome under certain key conditions. Demand must be fundamentally diverse so that the original GPT does not serve all demanders. Firms barred from entry into the primary GPT market can then reallocate innovative resources to create new markets to meet the unserved demand. The demand in these new markets must be valuable enough (even if not as valuable as in the primary GPT market) to generate a positive feedback cycle that results in considerable technical advance in the alternative GPT. This ultimately can lead to indirect entry by the alternative GPT into the original GPT market if and when it becomes strong enough to compete with the original GPT.

This sequence of (1) increasing returns to scale around a GPT, (2) reallocation of innovative resources around growth bottlenecks, and (3) indirect entry has growth implications. A large contribution to growth follows the exploitation of increasing returns to scale in the original GPT. Far later, another large contribution to growth follows when demand is finally met by an alternative, competitive GPT. Between these two periods falls a period of lesser contributions to growth due to the dominant firm bottleneck. The market-based resolution of the bottleneck is not merely a theoretical possibility. We illustrate the role of this sequence in the two most important technologies for automating white-collar work of the last 50 years.

* The authors are Professor of Economics at Stanford University and Assistant Professor of Technology, Innovation, Entrepreneurship and Strategic Management at the MIT Sloan School of Management, respectively. We thank the Ewing Marion Kauffman foundation for support. We also thank Franco Malerba and two referees for very helpful comments.

1. **Introduction**

This paper is about the relationship of general purpose technologies to demand and growth.[1] General purpose technologies (GPTs) can address the central growth needs of an entire economy, as evidenced by familiar phrases like the "information age" or the "age of steam." For a technology to be general purpose in this sense, it must meet a widespread demand need.[2] The long run importance of a GPT for growth arises because its widespread use enables its demanders to relax a growth constraint. For example, a growth constraint arose from the limitations of using wind, water, and muscle as a power source. The "age of steam" refers to a period of growth due to the widespread use of steam power to relax that constraint. Exploitation of a broad technological opportunity to answer the demand needs defined by such constraints can foster the growth of an entire economy.[3]

For the rich countries, the last half century has been labeled the "information age." That phrase successfully captures the supply of GPTs that comprise information technology. It does not capture the central growth constraint underlying the demand for information technology. A label such as the "white-collar automation (WCA) age" would better reflect the demand needs and growth constraint being addressed by information technology which will foster economic growth.[4] Much of modern employment is white-collar work, and much of that in bureaucracies that buy, sell, or decide. A number of GPTs based in information technology have emerged to enable WCA, each serving demand needs in a wide range of industries and functions. Satisfying the demand for WCA technologies has been critical to recent growth.

We examine two three-part sequences in the information age, one associated with enterprise computing and automation of white-collar bureaucracies, the other with individual productivity computing and the automation of an individual's white-collar work. Enterprise computing (EC) has gone through a three-part sequence in fulfilling the demand for WCA. The most important

---

[1] The macroeconomic consequences of general purpose technologies are discussed in Aghion and Howitt (1998) and in Mokyr (1990), and Mokyr and Scherer (1990).

[2] See Bresnahan and Trajtenberg (1995) for links between "micro" and "macro" approaches to GPTs. Helpman and Trajtenberg (1998) are an early example of an explicitly macroeconomic growth model with GPTs.

[3] There is a long literature on the role of demand inducement in technological progress. Ruttan (2001) has links to this literature. We share a focus on demand needs with this literature but assume that the market process by which demanders have influence on the direction of technical progress is neither simple nor automatic.

[4] Similarly, the "age of steam" can be more closely linked to growth economics by thinking of it as the "industrial revolution." While that demand-side label is used in the past, in the present we tend to think of "information technology" or "computers" rather than the white collar work they are used to automate.

GPT for EC for many years was the IBM mainframe. IBM and its enterprise customers succeeded in exploiting social increasing returns to scale by building a positive feedback loop centered on IBM mainframes. A period of slower growth followed when the position enjoyed by the dominant firm, IBM, became a bottleneck for technical progress. While demanders sought alternatives to IBM during this period, direct entry by any new GPT aimed at EC customers was barred. However – and this is the key to understanding our three-part sequence -- during this period, alternative GPTs, such as minicomputers and workstations, served demand fundamentally different from the highly valuable EC demand served by IBM mainframes. The establishment of these alternative computer GPTs outside of the scope of IBM's dominance led to new positive feedback cycles and ultimately to very effective competition against IBM. This indirect entry ultimately worked to the benefit of IBM's original customers, ushering in an aggregate growth spurt as many different demanders exploited the alternative GPT. We examine this history analytically to see when indirect entry of this sort is possible.

We believe that a very similar sequence has been partially completed in connection with individual productivity computing (IPC). The personal computer (PC) is the GPT at the center of a round of IPC white-collar automation. For a number of years, technical progress in the PC was remarkably responsive to demand needs. After a period of time, however, Microsoft became the dominant firm in selling what is now known as the "Windows PC." Just as a powerful positive feedback loop protected IBM's position earlier, a different positive feedback loop protects Microsoft's position today. Microsoft's position has become a growth bottleneck for WCA focused on the individual worker. Demanders are seeking alternatives to Microsoft technologies, but direct entry by a new technology aimed at IPC customers is barred. New GPTS are being established outside the scope of Microsoft's dominance, however. The most successful of these new GPTs serves a fundamentally different body of demand than white-collar work automation; they are focused on the consumer rather than the worker. We see the emergence of these alternative mass-market computing GPTs as signs of future, if not imminent, indirect entry. That would complete this sequence and potentially permit a new growth spurt. Will these new entrants, today largely providing infrastructure for consumer demand and entertainment, ultimately become the GPTs that reignite technical progress in the automation of white-collar work for the 21st century?

We present a demand-oriented framework to explain these sequences and argue that under key conditions, market solutions have overcome growth bottlenecks and may do so again. Our framework explains the logical connection between successful exploitation of social increasing returns to scale and later emergence of a growth bottleneck. This is the duality of social increasing returns to scale. It leads to the first two parts of our sequence, in which the GPT first makes a large contribution to growth but then slows. We also argue that under certain conditions, and here we especially emphasize demand conditions, renewed entry and competition can overcome the growth bottleneck.

2. Demand-oriented framework for understanding the three-part sequence

We build on a number of familiar economic analyses of innovation as a bridge between technological opportunity and demand needs. We begin by being careful about the definitions of the analytical tools we use. This permits us to state precisely the conditions under which we see a three-part sequence of initial exploitation of a GPT, emergence of a bottleneck, and renewed contributions to growth via indirect entry of alternative GPTs resulting from innovative resources that had been reallocated around the bottleneck. Demand conditions emerge as central to the three-part sequence.

2.1. Definitions

We define a technological opportunity conventionally as a technological area in which the cost and difficulty of invention are sufficiently low to permit firms to develop commercially valuable new technologies. We are particularly interested in the implications of a broad technological opportunity (BTO), which is open to development by different firms and within which different directions of technical progress are possible.[5]

Whether a technological area can generate value depends upon demand needs. We define demand needs conventionally as opportunities for technical progress to create value by lowering costs or creating new goods.[6] In a departure from the macro GPT literature, we acknowledge

---

[5] It is not important to our definition whether a BTO came into being from academic science or from spillovers from a pioneering development by a commercial firm. Our definition differs from Jaffe (1986) who limits "technological opportunity" to "exogenous" sources such as science in order to distinguish it from spillovers.

[6] There is a considerable literature on the commercialization of innovation which focuses on this. See the papers in Part V of Rosenberg et al. (1992), also Gans and Stern (2003). One important contribution emphasizing the organization of innovative effort is Clark (1985).

heterogeneity in the relationship between demand needs and long-run growth. Some, but not all, demand needs stem from the growth constraints on an economy at a particular stage of development.[7] Technical directions that better address those growth-critical demand needs will, not surprisingly, be more valuable to long-run growth. We are particularly interested in heterogeneity that leads to fundamentally diverse demand needs. Start from the population of demanders who could benefit from some direction of technical progress based on a BTO. That population is large if the BTO might yield a GPT of economy-wide importance. Demand needs are fundamentally diverse if any specific GPT direction taken within a BTO does not serve some of these demanders.

Our definition of a GPT is also conventional. Following Bresnahan and Trajtenberg (1995) we define a GPT as a technology that (1) is widely used, (2) is capable of ongoing technical improvement, and (3) enables innovation in application sectors (AS).[8] Also, when a number of distinct technologies (in the engineering sense) are used together as very close general purpose complements, we follow the convention in the literature of calling them a single GPT.[9]

The literature needs a label for the union of a GPT and its AS, for the AS are the key to the "general purpose" aspect of a technology. Heterogeneity of demand needs prevents most technologies from serving a wide enough range of demand needs to be considered general purpose. The complementary innovations in different AS built around shared technical progress in a GPT permit that GPT to meet heterogeneous demand needs. We introduce the term "GPT market cluster" to denote the set of firms which innovate in the GPT itself, the firms which innovate in the various AS, and demanders.[10]

---

[7] This emphasis is central in the literature on induced innovation. See Ruttan (2001). We depart from the induced innovation literature when it characterizes demand needs by the factors of production saved by technical progress. Instead, we focus on the processes or not-yet-invented products which constrain growth. This departure is consistent with theoretical and historical/analytical discussions of inducement (Acemoglu (2002), Rosenberg and Mowery (1979).) The existing macroeconomic GPT models which link GPTs to aggregate growth (for example, Helpman and Trajtenberg (1998), but also see the review in Jovanovich and Rousseau (2005)) assume that a GPT is used in all of the (symmetric) sectors of the economy and thus do not distinguish among demand needs.

[8] This definition is standard, if terse. See Helpman and Trajtenberg (1998) for further details. A far longer statement of this definition and careful thought about boundaries can be found in Lipsey, Carlaw and Bekar (2005). See also Bresnahan (2010) for a review of the impact of alternative GPT definitions on the analysis.

[9] This treatment of very close complements as part of the same GPT is conventional. It follows from the definition of a GPT, where the "general" attribute is considered from the demand side. See Bresnahan (forthcoming) for extensive discussion.

[10] This is an inclusive definition whose main purpose is to provide a label. Within a GPT market cluster the three classes of actors (GPT firm/ AS innovator / demander) need not be distinct. Sometimes AS innovations will

The GPT literature distinguishes between technical progress in the GPT itself and the complementary technical progress in applications sectors, calling the combination of elements (2) and (3) in the definition of a GPT "innovational complementarities". Innovational complementarities mean that more or better innovations in the GPT raise the return to innovations in each AS and vice versa. The GPT literature has pointed out that the presence of IC means that there are social increasing returns to scale (SIRS): coordinated innovation in the AS and GPT and/or additions to the list of innovating AS improve product quality or lower costs in all AS.[11] By our definition, a GPT market cluster contains the scope of its SIRS.[12] Similarly, as is commonly observed in the GPT literature, the presence of innovational complementarities means that there is positive feedback between the GPT's rate of technical progress and each AS' rate of technical progress and therefore among all the AS. The scope of these positive feedback loops is also contained within the GPT market cluster.[13]

2.2. Fundamentally Diverse Demand vs. SIRS

We now turn to the dynamics of our analysis and the possibility of a three-part sequence. We begin with the assumption that a GPT is invented, and a GPT market cluster is founded to exploit overlaps between the BTO and demand needs, leading to an initial contribution to economic growth.

The benefits of SIRS arise from coordination on a GPT in a market cluster. The choice of a specific technical direction may be essential to focus investments by many different firms

---

be conducted by demanders in-house, sometimes by specialized firms. For example, in enterprise computing one sees both in-house computer departments "applications" for a single firm's use and also applications software vendors like SAP supplying a market software product to many corporations.

[11] Farrell and Klemperer (2006) note that, strictly speaking, the definition of social increasing returns to scale implies a cooperative game theory perspective; network effects are just economies of the per-buyer agent surplus available to a coalition that increases with the size of the coalition. We follow the literature by using both this cooperative perspective and, below, the distinct perspective in which agents act independently.

The link to innovational complementarities is tight: Complementarities or spillovers are always present if there are SIRS.

[12] Many if not all of the macroeconomic models cited by Jovanovic and Rousseau (2005) assume that the scope of a GPT cluster is economy-wide, and that there is only one GPT at a time in the economy. All AS are symmetric and all are symmetrically supported by one GPT at a time. This is not an analytical assumption, however, but merely the habit of macroeconomic modeling to treat all sectors of the economy symmetrically (or in this case all but one sector, the GPT). These assumptions are for analytical convenience rather than empirical accuracy.

[13] As Farrell and Klemperer (2006) point out, positive feedback is inherently a non-cooperative game theory concept. In this context, positive feedback is related to the market or equilibrium implications of innovational complementarities.

within a nascent GPT market cluster in order to get positive feedback going. The early selection of a technical direction may be efficient; indeed, there may be a competitive race among technological directions to establish a GPT. As a result, demanders play an influential role in the direction of technical progress. Our demand-oriented framework highlights the degree to which demanders have an influence on the direction of technical progress. We are particularly interested in the case in which the GPT may support applications serving to the most important demand needs of an economy, those which currently constrain growth. We refer to the initially chosen GPT and associated market cluster as the “original” GPT and market cluster.

Classical GPT analysis identifies a well-known tradeoff between SIRS and fundamentally diverse demand.[14] Heterogeneity in demand means that any particular direction of technical progress in a GPT will fit some demanders’ goals better than others’.[15] If there is only one GPT market cluster related to a BTO and demand needs are fundamentally diverse, some needs will be unserved. The distinction between a BTO and a GPT permits us to identify which demand needs have been served by some particular GPT and which have not. Note that the unserved demanders will not participate in the positive feedback cycle of that GPT and its applications, and they will be therefore excluded from that GPT’s market cluster.

The distinction between a BTO and a GPT also emphasizes that a GPT involves specific technical choices which, if made differently, might have led to different directions for technical progress. We are particularly interested in the case where an alternative GPT from the BTO would have led to unserved demanders being served. That alternative might be cheaper (and less powerful) or more powerful (and more expensive) than the original GPT, or it might involve a different set of choices across multiple product quality dimensions.[16] This situation is illustrated in Figure 1. In that figure, the tradeoff between exploiting SIRS and serving fundamentally

---

[14] Excellent coverage of the GPT literature can be found in Helpman (1998). Farrell and Saloner (1986b) discuss the tradeoff between exploitation of SIRS and variety in the context of standard setting. Lucid discussions of the general relationship between sharing to exploit SIRS and meeting diverse demand needs can be found in Gilbert (1992) and the handbook chapter reviewing these ideas, Farrell and Klemperer (2006).

[15] Of course, one job of the complementary innovation in the AS is to adapt the general to specific circumstances. This is limited, however, by the underlying fact that all AS share the technical progress in the GPT.

[16] For example, the horse and the railroad each had particular product quality dimensions when they were successful transportation GPTs. The automobile, which had the flexibility of the horse and non-muscle power sources like the railroad, was more appropriate than either for what turned out to be a very large body of unserved demand.

diverse demand needs is solved efficiently by having two GPTs with distinct market clusters drawing on the same BTO.

2.3. The Duality of SIRS: Growth Bottleneck

The technical choices made in an early period to create the original GPT market cluster can constrain choices made later. These later constraints can alter the ability of the original GPT market cluster to support growth that might be possible under an alternative GPT. We call this situation a growth bottleneck.

Supply of key components of a GPT by a dominant vendor is a common industry structure. A dominant vendor is a potential coordinating actor for movement to a new technical direction. A dominant vendor with strong marketing relationships may be uniquely positioned to discover demanders' technological needs and change the technical direction of GPT development to suit those needs. However, a dominant vendor has a powerful interest in continued focus of investment by all market cluster participants around its own products and only limited incentive for risky change. An existing dominant firm may also have built an organization that is primarily capable of advancing the existing direction.[17] A dominant vendor makes the direction of technical change in a GPT cluster more purposive: sometimes the dominant firm's strategic purposes can reinforce the growth bottleneck rather than steer around it. When the dominant vendor is part of the bottleneck rather than the agent of change and renewal, we use a shorthand notation and say that the dominant position is the bottleneck.

While our model of bottleneck centers on market forces, the more common model of a growth bottleneck is conceptual.[18] We emphasize the market sources of a bottleneck, not because we believe that conceptual bottlenecks are impossible, but because the positive feedback

---

[17] Will the new technological directions be the ones which best reflect overlaps between technological opportunity and demanders' needs? Yes, if two conditions hold. First, the dominant firm's marketing connections must put it in an excellent position to identify and implement technologies embodying those overlaps; i.e., there must be little gain from having multiple firms rather than one identifying and implementing. Second, market dominance must create incentives to bring new technologies to market in demanders' best interest; i.e., protecting market dominance never involves staying with old technologies in favor of better, new ones. Both conditions can and sometimes do fail in real markets, at which point a growth bottleneck arises.

[18] In emphasizing the market elements, we recognize we are departing from the most common approach. A number of theories exist as to how and why there may be a growth bottleneck. Most prominent are theories of knowledge related to the attractive power of existing ideas and the difficulty of breaking out to new ideas, such as Giovanni Dosi's technological trajectories (Dosi 1982). The core of the Dosi trajectories idea is that paradigms lead firms to be attracted to a particular set of ideas (thereby gaining dynamic efficiencies) and exclude other ideas. Similarly, there are many attractive theories of why a dominant firm in a GPT market cluster might have conceptual weaknesses when faced with the opportunity for a valuable new technological direction.

loop in a GPT itself creates conditions which naturally lead to a growth bottleneck. Existing demanders, existing AS developers, and existing GPT firms in a GPT market cluster all participate in a positive feedback cycle linked to an existing technical direction. As the literature emphasizes, this can be a powerful inertial force.[19] Changing the direction of technical progress in such an inertial system may be very difficult, even if there is new knowledge about demand needs (perhaps learned from the grand experiment associated with construction of the original GPT market cluster).

Whatever the reason for choice of an original technical direction (and possibly an original dominant vendor) we are concerned with the economic problem which arises when demand needs change and there is value to exploring an alternative direction to the original GPT. The inertia means that the direction of the original GPT may not change. The inertial forces also may mean that an attractive new direction is not enough motivation to move a large coalition of demanders away from the original GPT. By the same token, SIRS make the establishment of a new GPT cluster in competition for demand in an existing one very difficult unless the new GPT can somehow convert a large number of customers from the original GPT.

This is the fundamental duality of SIRS. Demanders of a GPT gain the benefits of sharing a common technological input but are also subject to the possibility of inertia that comes from positive feedback. The growth bottleneck will reveal itself as a monoculture: only products consistent with the original GPT will be available to demanders as practical technological choices. Demanders in this GPT market cluster will perceive directly competitive technologies as fringe products; the firms sponsoring alternatives to the existing GPT dominant firm will appear to demanders as fringe firms. The growth during this period of bottleneck will be slower than potentially possible under an alternative GPT.

2.4. Unserved Demand and the Reallocation of Innovative Resources

How, then, can a growth bottleneck be overcome? We locate the conditions for a market solution in the outcome of the original tradeoff between SIRS and demand: unserved demand. Fundamentally diverse demand means that the original GPT did not serve some bodies of demand because their needs were too distant from those of its main customers. In the presence

[19] Farrell-Klemperer (2006) has a careful statement of the conditions under which inertia will and will not be observed.

of a growth bottleneck, fundamentally diverse demand is an opportunity rather than a problem. A new GPT market cluster can, sometimes, take root with the previously unserved demand as its main customers. This sets the stage for the third part of the sequence.

The economic analysis of technical change has long emphasized one part of this idea. New technological directions are often opened up because of the existence of heterogeneous demand pools. Some historical examples include very important GPTs such as steam power.[20] The literature on market segmentation in computing and the creation of new demand segments, upon which we build in our two case studies, has pointed out that entrant technologies and entrant firms can find a market niche to enter even if direct competition with an existing firm or technology is unprofitable. This idea has been picked up in a number of distinct literatures. One is the systematic analysis of new goods introduction. Another literature has a number of recent papers that have examined the strategic implications of multiple demand pools for new technology strategy from a business strategy and an economics perspective.[21]

What we assume in common with all these literatures is that entrants can take advantage of other technical directions in the BTO and create an alternative GPT much closer to the needs of unserved demanders. This creation involves elements of both technological push and demand pull. The technological push arises from entry barriers around the original GPT which prevent new firms from successfully deploying new technologies in direct competition with the dominant firm. Those new firms then seek other market opportunities, possibly less valuable than the original one. The demand pull arises from an unserved body of demand attracting these firms by providing a new market for the entrants. Because of the fundamental demand diversity, the alternative GPT will not initially compete directly with the original GPT; it will serve demand outside the original GPT market cluster. Since the entry barriers in the original GPT do not extend to these new markets, a new GPT SIRS and positive feedback cycle can begin, where interaction between demanders and suppliers leads to innovative co-invention in an alternative GPT market cluster. Note that this possibility does not hinge on the alternative demands being

[20] See Landes (1969) for a variety of historical examples and for their relationship to economic growth, and Rosenberg (1976) on the introduction of steam power into transportation. Further cites on this topic can be found in Bresnahan (2010).

[21] Adner and Zemsky (2005), Christensen (1997), Adner and Levinthal (2001).

highly valuable to economic growth. The key point is that they are outside of competition with the existing technology and large enough to sustain technical progress.

We call the result of this technology push and demand pull away from the potentially most valuable demand needs in the original market cluster a "reallocation of innovative resources" to contrast it with a simple demand-inducement story (see footnote 3, above).

2.5. Related Demand Permits Indirect Entry from the Alternative GPT

Our demand-oriented framework allows us to go beyond the usual partial equilibrium model of industry structure. Innovative resources barred from entry into the primary market did not simply disappear, nor were they randomly dispersed to other markets. Instead, these resources were reallocated to particular types of markets, pulled by the demand in those alternative markets for a technology similar to the original GPT but which had not been met by the dominant firm's products. These alternative markets are related to the primary market through the BTO. Therefore, they are close enough to the original GPT such that technological advances in the alternative GPTs resulting from innovative co-invention permit re-entry into the original GPT market cluster and competition with the original GPT in the long run.

Once again, we recognize that our market approach is less familiar (and more complicated) than a story in which a disruptive technology appears exogenously. Our motivation in emphasizing the forces determining the appearance of a new, effectively competitive, alternative GPT is not that it makes a more beautiful theory. Instead, we base this analysis on the practical observation that a new GPT, to be successful, needs a "killer application." Raw technical ideas are not, natively, disruptive in a GPT context. The creation of a positive feedback loop around a new, alternative GPT can convert a good technical idea into a candidate to compete, on a cluster level, with an existing, entrenched GPT.

The conditions for indirect entry require fundamentally diverse demand needs that have a BTO in common. The most important demand needs which reflect growth constraints will eventually pull alternative GPT to serve it as soon as the alternative GPT is superior enough to the original GPT to overcome entry barriers. In order to become competitive enough against the original GPT, the GPT must progress technically so that it can eventually serve demand in the original market cluster better than the original GPT. That technical progress requires a sufficiently large market cluster to finance and promote innovational complementarities around the alternative GPT. The key to the ability of the alternative GPT to become general enough to serve both its own market cluster and enter the original market cluster span lies in its AS. The AS must be well developed enough to permit such a wide range of use. The path of indirect entry may start with some particularly attractive application to a one part of the demand in the original market cluster. It may also start with an overwhelming observation of faster AS development and technical progress in the alternative GPT relative to the bottlenecked original GPT that causes demand in the original market cluster to become dissatisfied with the status quo. Our demand-oriented framework highlights the importance of the GPT market cluster in advancing technical progress, in contrast to the assumption that technical progress from exogenous scientific advancements will be sufficient to induce indirect entry.

3. A completed sequence from history: Enterprise Computing

At the end of the Second World War, one of the clearest long-term needs in the rich countries was for technologies to automate white-collar work.[22] A long-term program of automating physical labor in agriculture had succeeded, and a long-term program to automate physical labor in manufacturing was succeeding. A new growth constraint was visible on the horizon: the lack of a technological basis for productivity improvements in the services sectors and in the white-collar functions of goods-producing sectors.[23] A more educated workforce fueled an ongoing expansion of white-collar work and its potential for being automated.[24] The stage was set for substantial demand for technologies to be used in white-collar work.

Demand for white-collar automation was first addressed by dramatic progress in enterprise computing (EC). Enterprise computing is the GPT that uses computing power to automate a wide range of white-collar functions at the firm ("enterprise") level across a wide range of industries. Enterprise computing consists of three main parts. The first are the large and powerful computers, originally called mainframes, and related (complementary) hardware, such

[22] The prospect that no such technologies would be found led to the forecast of "Baumol's disease" in which certain sectors' costs rise without limit until they use all of GDP. See, e.g., Baumol (1967).

[23] The study of productivity in the relevant sectors of the economy is rendered all the more difficult by the problem of measuring output there. See the introduction to and the papers in Griliches, Berndt, Bresnahan and Manser (1992) for an overview.

[24] There is a debate on the relationship between computerization and skilled labor demand; with many empirical scholars finding that computerization of organizations and skilled labor are complements. This is consistent with the demand for computerization being accentuated by an increase in the supply of skilled labor, e.g. Acemoglu (2002).

as disk and tape drives, printers and card readers, data communications equipment, terminals, and so on. The second is software: applications software for different kinds of white-collar business functions, general purpose (fundamental) software such as operating systems, and programmer tools. The third and largest part of enterprise computing, as measured by costs of invention, is the specific applications in different enterprises built using those programmer tools.[25]

Enterprise computing exhibited three innovational complementarities which generate SIRS. The first was economies of scale from sharing expensive technical progress. Advances in computer hardware (computers, storage devices, input/output devices) and general-purpose software such as operating systems could be spread over a large number of demanders. For EC, this mainly involved mainframe computers and their peripherals and systems software. The second was creation of applications software. Industry specific ("vertical," in the odd language of computing) applications software made mainframes useful in a wide range of industries. More importantly, function-specific applications software such as accounts receivable, inventory control, and so on made the general purpose components useful in a wide range of white-collar functions at the firm level, at least in large firms. Third, and economically most important, enterprise computing was an invention which enabled a wide range of inventions.[26] The key to enterprise computing's ability to serve demand for WCA is the very important programmer tools such as database management systems and communications controllers which enable organizations to create a wide range of complex applications.

3.1. Demand Influence: Market Selection of Original GPT

The invention of EC occurred in what was initially a very competitive environment.[27] Three very different kinds of firms sought to set the standards which would define EC. Those included existing office equipment firms, of which IBM was originally only one among many, existing electronics firms, including such heavyweights as GE, RCA, Siemens, and Honeywell, and new firms, of which the most promising (e.g., Eckert-Mauchly) had a strong basis in the new

---

[25] See Bresnahan (2002) for a cost estimate of the different elements (hardware, software, customer software, software written by demanders, etc.) of EC.

[26] The allusion to Griliches (1957) on hybrid corn is entirely deliberate.

[27] A number of different authors have examined this era. See Usselman (1993), Bresnahan and Malerba (1998) and Fisher, Mckie et al. (1982).

technology of computing. IBM emerged from this early competitive struggle as the dominant supplier of general purpose components.

From a demand perspective, IBM's creation of a cluster of EC technologies was initially a very beneficial development. Because the winning firm was determined by market selection, demanders' needs played a large role early on. The initial market selection favored a single, dominant firm in part because there was much uncertainty in the early days about the multiple complementary technologies which needed to coordinate to form the EC supply chain. Enterprise use of mainframe computers required a wide number of hardware, software, and service complements. IBM provided service and hardware complementary assets to help customers implement EC. IBM marketed its mainframes through a direct sales force, educating customers about the utility of their products. The sales force developed relationships with IBM customers, thereby acting as a channel through which IBM gathered information on customer usage and demands, in order to develop complementary software and services to better attract future customers. As a result, IBM created a high quality reputational asset among the demanders who were being served by the current GPT. IBM won the competitive EC battle because it best solved the complex problem that faced demanders of a new and uncertain technology.

While the winner was IBM, it is essential to understand that it was the process of competing in the early period that was highly beneficial to demanders. Given SIRS, there were going to be only a limited number of market solutions after the period of initial competition. IBM had the knowledge of both demand and (after investing in it) technology which could manage the EC supply chain coordination, attracting demanders and achieving SIRS. In pushing past competitors, IBM played the coordinating role in pressing forward the compatibility standards which led to successful exploitation of those SIRS, defining the technologies needed for enterprise computing. IBM also succeeded in making those standards proprietary.

Enterprise computing continued to change and improve, both in the underlying technologies embodied in capital and in the programmer tools. Since there was an IBM proprietary standard in enterprise computing, IBM was largely responsible for the general-purpose parts of these improvements. IBM managed technical change in enterprise computing, drawing on its own technical capabilities, on general purpose inventions which came from customers or from other computer and software companies, and on its deep knowledge of customer needs captured through service, support, and field sales. IBM managed not only the steady accumulation of knowledge about enterprise computing, but also managed and introduced radical or disruptive changes in enterprise computing technology and tools itself.[28]

IBM successfully coordinated efforts to locate overlaps between the BTO and as-yet-unexpressed market demand and adapted over time to find new overlaps as technology advanced and demanders learned more about their needs. The growth consequence of their success was an evolving solution to the WCA constraint.

3.2. Dominant Firm Becomes Bottleneck

IBM mainframes had important limitations. As we now know, competition from other, more technologically progressive firms ultimately removed IBM from its position of dominance in enterprise computing. To understand the full EC sequence, we first examine the period in which IBM's customers and competitors were aware of, but unable to solve, the problems arising from IBM's technological lag.

Figure 2 is useful for emphasizing the separation between the BTO in computing and the GPT cluster in EC around IBM mainframes. IBM was the dominant firm in the market for mainframes, but IBM did not dominate knowledge about the BTO. The original GPT (IBM mainframes in EC) was only one of several clusters within the BTO (computing). Computers had always been used for purposes other than enterprise computing. The stored program computer was invented for scientific calculation. From the beginnings of the computer industry, scientific and engineering computing was the second major market, and minicomputers dominated supply in that market cluster.[29] IBM's customers were technologically sophisticated about computing generally and were able to critically asses IBM's technology strategy, particularly in comparison to other market clusters. Significant customer concerns about IBM technology policy began to emerge in the 1970s.

One customer complaint was that IBM was slow to introduce new technologies. IBM's defenders would say that there were two efficiency reasons IBM was slow. The first was compatibility over time. It is easier to introduce some computer with a new feature than it is to

---

[28] Two examples include the multiprocessing computer (which would permit programmers to invent and test new enterprise applications without turning off production applications) and the relational database management system (which would permit efficient invention of more and more complex enterprise applications).

[29] There was, of course, always some overlap in uses, but our characterization of the two main demand implications of these two clusters highlights their differences. See Bresnahan and Malerba (1998) for details.

offer customers a way to use that new computer without radically rewriting their existing applications. The second reason arises from good customer service. Part of IBM's reputation arose from consultation with its current customers about the direction of technical change, and the firm was careful to work with these customers to avoid problematic surprises.

These efficiency arguments may be true. Nonetheless, they are evidence that not all is positive about the successful exploitation of SIRS around a single standard. IBM's reasons for slow technical progress still resulted in slow technical progress relative to the much more rapid introduction of computer and software technologies outside the IBM mainframe market by other firms taking advantage of the BTO in computing. IBM customers could look to products available in the minicomputer segment, for example, and argue strongly that IBM should be bringing products to market that were as technologically advanced as what could be seen there. No single firm (DEC was perhaps the closest) nor even a single architecture (again, DEC and imitators were the closest) dominated the minicomputer GPT. The marketing relationship between a minicomputer vendor and its customers was much weaker than in EC, so minicomputers were cheaper and market structure more fragmented. Enterprise computing customers bought minicomputers, and some scientific calculations were done on mainframes. Nevertheless, the existence of the minicomputer segment did not itself offer a serious alternative to the IBM monoculture in the enterprise.

A second set of customer complaints was that IBM mainframes were too expensive for a wide range of clearly valuable WCA applications. The single standard behind the IBM enterprise computing systems limited it to a narrow range of applications. IBM systems were large, powerful, difficult to use (requiring years of training) and came bundled with a very high level of support and service. This made them expensive and primarily useful in large enterprises and not in small- or medium-size enterprises or in the smaller and relatively independent divisions of large enterprises.

The large enterprises were IBM's core customers, and the scope of applications of IBM mainframes within enterprise computing was largely limited to them. The inertial forces around existing relationships were strong. These included marketing connections between IBM and large enterprises and compatibility connections between IBM technologies and applications developed by demanders. Together, they constrained IBM's ability to advance rapidly.

With such a large performance problem in a very profitable industry and such a wide number of different firms at the frontier of computer technology, one might have expected direct entry to trigger creative destruction. However, efforts to compete directly with IBM in enterprise computing failed.[30] Because of IBM's successful management of a proprietary standard, the same SIRS that formed around IBM so that it could win the enterprise computing market also formed entry barriers against other firms trying to enter the enterprise computing market. While market selection had picked IBM and its technologies at one stage, demand had little influence after that competitive process. Barriers to entry protected IBM long after society began to bear the costs of its dominance. For evidence of the counterfactual to this claim, one only has to look at the way IBM's customers eagerly replaced its products with lower cost and more powerful products (competitive mainframes, superminicomputers, even workstations) fifteen years later. The downside of having firms become standards due to SIRS is that, over time, these firms may use the entry barriers that develop around them to prevent innovative resources from entering the primary market with superior complementary and directly competitive products.

The long-lived productivity gains from automating blue-collar work in the rich countries finally began to achieve their objective and thus ran into diminishing returns. This made WCA into a growth constraint, creating an opportunity for a growth contribution via the exploitation of SIRS around a GPT that could support WCA in a wide range of white-collar functions in a wide range of industries. By the same token, however, it made the particular GPT that was used for WCA, in this case IBM mainframes, into a potential bottleneck.[31] Aggregate technical progress was, to a significant extent, embodied in investments in information technology capital. Demanders could not always instantly invent WCA applications. As a result, the capital stock grew slowly but steadily. In some eras, this had aggregate implications for growth.[32] This is not

---

[30] The most successful efforts were hardware clones from the US (Amdahl) and Japan (Fujitsu). IBM remained unchallenged in key software components and, critically, retained sole control of the mainframe standards. No firms made substantial inroads with their own architecture. For an analytical history of technological competition in this cluster, see Bresnahan and Greenstein (1999).

[31] In most rich countries, IBM was the dominant enterprise computer firm, and a single national champion telephone company was the dominant data transport firm. We have emphasized computing in our historical work, but it is worth pointing out that AT&T and the PTTs of other nations were also monocultures in their sphere in the 1970s. The world was badly set up for extending enterprise computing into automating markets.

[32] Many scholars have argued that the essential problem of this era was lagging technical progress in specific sectors (Baumol 1986) as opposed to in specific functions. This has led to the development of models of unbalanced growth (Kongsamut et al. 2001). Baumol and his colleagues have argued that the use of computers is,

to say that there were no productivity gains resulting from the deployment of computers over the period from the creation of the computer industry through the 1970s. However, an often overlooked phenomenon is the way in which IBM's dominance limited the range of applications of enterprise computing technologies and thus the range of co-invention of WCA technologies in this era.[33]

3.3. Reallocation of Innovative Resources to Unserved Demand

Although the market process led to the IBM monoculture, the market process could also, in the long run, resolve the problem created for demanders by IBM's dominance. Entry barriers created a technology-push away from EC, while markets where demand for a technology like the EC GPT had not been served by IBM's products generated a demand-pull for the innovative resources. This resulted in the reallocation of innovative resources away from the crucial growth constraint on the economy, WCA.

Figure 2 also depicts the relative importance of demand needs which were met by computing from the 1950s through circa the early 1970s. Towards the left of Figure 2, we show the central growth constraint of the last half century, WCA. White-collar automation itself contains a wide number of different areas, including automation of large enterprises, small- and medium-size enterprises (SMEs), and individual white-collar workers. A key point is that not all of WCA was served with a computer technology in this era, despite WCA's role as a demand need. Enterprise computing served a subset of WCA, the automation of large enterprises' firm-level white-collar work, much more effectively than other WCA demand needs.

Figure 3 helps us to clarify the process of reallocating innovative resources away from the original market cluster to new GPT clusters. As the BTO in computing generally advanced, more new computer markets were founded, all outside of the original WCA market cluster. The

overall, a "stagnant sector" because of the need to have (labor-intensive) software (Baumol et al. 1985). While there is no doubt that innovation in the use of computers involves more difficult technical progress than the invention of computers, it is odd to label the use of computers as a sector. This misses (1) the dramatic improvements in programmer tools with resulting programmer gains and (2) the benefits of re-using software across many firms. In any event, the historical pattern since Baumol and his colleagues were writing has been that productivity growth has been more rapid in the sectors of the economy which use computers more intensely.

[33] There is a large literature on computers and productivity and a smaller one on enterprise computing technologies and productivity. There are two main theories of this early period. (a) There was a long period of investment in computing technology that did not yet lead to any gains (and (b) there were gains to early investments in computing but the total amount of investment in new computer and enterprise computing capital was not large enough to boost aggregate productivity. We are convinced by the second argument, which is much more consistent with the technical history of the period.

invention of the microprocessor led to the founding of two alternative GPTs and associated market clusters. One of these was the *engineering workstation.* Serving technical rather than commercial customers, workstations were machines for a single user, such as an engineer designing components or, later, a graphics designer working on motion pictures. Workstations were microprocessor-based, far cheaper than mainframes, and the demand they served was very distant from that of IBM or even DEC. The workstation market cluster included many academic-based startups, such as SGI and Sun Microsystems.

The *personal computer* (PC) market was a second market cluster where demand pulled resources surrounding the microprocessor. At first, this industry was for "hobbyists," i.e., people who wanted to solder together a cheap computer kit. A supply chain emerged to serve these particular demanders. The open-system architecture for the PC encouraged development of a wide variety of software. Originally, the applications were primarily games and programming languages serving hobbyists, and the PC was marketed to the home user. PCs were far less functional than enterprise computing systems but also cost far less. The PC came to be the center of yet another GPT cluster based in the computing BTO.

Related developments were occurring in fundamental software. A small but effective open-systems movement, based largely in military and educational sites, pushed forward development of the UNIX operating system, for example. These developments bridged the BTO and the widening array of previously unserved demand needs.

There are two very different forces at play here. The first is supply and technological push. There was an enormous amount of money in EC and a modest amount in scientific and engineering computing from the 1950s through the end of the 1970s that led to advances in the BTO. IBM's dominant position in mainframes bottlenecked the deployment of the innovative resources away from its most valuable use, computing in WCA. This created a tremendous technology-push which contributed to the creation of alternative GPT innovations in other markets. The second is demand pull. IBM as dominant firm had made a number of specific choices about the rate and direction of technical change in its particular GPT. Those choices limited applicability. They also limited the range of IBM's connections to customers and thus the span of entry barriers. Hobbyists, engineers, scientists, and defense department computer labs working on UNIX did not have the kinds of budgets that EC buyers did. However, unlike EC buyers, these demanders could pull those excluded innovative resources to serve their needs, since IBM's entry barriers did not extend to their markets. These other new markets had not been served, or at least not effectively served, by IBM mainframes, but their demand could be served by some alternative manifestation of the BTO.

The outcome of these forces circa the end of the 1970s was the more complete conversion of a BTO into a group of GPTs serving different demand needs. In parallel to IBM mainframes serving a range of different applications sectors within enterprise computing, we see in Figure 3 a number of other GPTs, all "computers," serving clusters of other markets. The highest-value for applications of the BTO in computing in the 1970s would have been to further advance WCA, yet that was not the direction taken by technical progress in the creation of new GPT clusters. Fundamental growth constraints are shown to the left in Figure 3, but innovative resources were reallocated more to the right, serving demands less relevant to the main growth needs of the economy than WCA.

While any technological progress is valuable, the most important growth implications of this expansion in the range of demands served by computing would not be realized until later. While serving scientific and engineering computation was valuable for growth (if not a key constraint), none of the alternative GPTs was initially involved in WCA. A critical point about GPTs, however, is that innovative co-invention by demanders can give them new areas of application. This process of co-invention in the market would allow these new, alternative GPTs to eventually enter the original market and be used in WCA. The next section describes this process of indirect entry.

3.4. Indirect Entry

The creation of all those new segments of computing, while not competitive with the IBM mainframe in the short run, would become so in the future. Where direct entry had failed, indirect entry (Bresnahan & Greenstein 1999) ultimately provided enterprise computing with a wide range of alternatives to the IBM monoculture. An interim step was expansion of the range of the new clusters just described into WCA if not all the way into EC. This interim step is shown in Figure 4.

Because of their very high costs and services, IBM mainframes were far more suitable for WCA in large corporations than for small and medium size enterprises (SME) or for independent divisions of large corporations. This led minicomputer manufacturers to introduce the "superminicomputer" or "commercial minicomputer" designed to serve WCA in SME's rather than large firms. The first of these was DEC's VAX family. Superminis came with far less service and support than IBM systems but at a far lower total cost of ownership. They had considerable success with customers who demanded WCA but fell below the effective minimum size for IBM systems. The GPT cluster that initially emerged from the alternative minicomputer GPT served primarily SME and divisional computing.

Rather than being organized around a single, vertically integrated firm like traditional EC supply, where computer, storage devices, operating system and programmer tools would come from a single firms, the new market serving smaller WCA customers was organized on open-systems lines. There was separate competition in different "layers." Most importantly, there was separation between key software, such as programmer tools, and key hardware, such as computers themselves. Thus competition broke out between minicomputer producers (e.g. DEC) and workstation producers (e.g. SUN) to sell the hardware for these smaller, WCA projects. Separately, competition broke out in key software layers, such as the database management system, with vendors such as Oracle and Ingres not tightly linked to any hardware vendor.

The open systems approach was far less technically backward-looking than the IBM monoculture. Critical advances such as the relational database management system were available outside the IBM world (first from Oracle) before they were inside the IBM world (with the later introduction of DB2). The possibility of component-by-component competition led to rapid advances in the best available systems.   Over the course of the late 1970s and especially the 1980s, these systems were successful in WCA outside of EC and also became more and more of a competitive threat to IBM's position in EC.

IBM noted the looming competitive threat from superminicomputers, workstations, and the open-systems approach of its new competitors. IBM's efforts to respond competitively did not succeed in blunting the long-run threat. As IBM entered the superminicomputer market, it faced the very difficult choice of either offering something that was competitive in price and features or (very differently) offering something that was compatible with IBM mainframes technologically and had IBM-like service and support. IBM's efforts to compromise between these goals in the late 1970s and early 1980s were unimpressive and met little success in the marketplace. IBM faced severe internal conflicts serving SME's due to its organization around serving large enterprises and its reputation. The potential for indirect entry by alternative GPTs was beginning to overcome the problems posed by the monoculture: a multitude of companies now offered a wider range of products than even the leading company could offer.

Still, there was little effective direct competition for IBM in enterprise computing in this era, and the firm continued as the world's largest hardware company, world's largest software company, world's largest networking equipment company, world's largest database management systems company and so on. In short, IBM continued as the most profitable corporation in the world through most of the 1980s by serving EC. The competitors, although closing in, were in other parts of WCA as shown in Figure 4. The invention of the spreadsheet and the word processor brought PCs into white-collar work at the end of the 1970s. The PC became a GPT serving a part of WCA. The PC solved a different piece of white-collar work automation than did EC or the supermini. Rather than automating white-collar processes at the enterprise or divisional level, the PC automated an individual white-collar worker's work. While this was responsive to the demand need in WCA, it was not yet a competitor for IBM in EC, and it was not yet a solution to the major white-collar productivity problem, which is bureaucratic rather than individual.

The PC's expansion into individual-worker WCA addressed unserved demand, so it did not represent competitive entry against IBM. IBM accurately saw the PC as a potential future complement for its existing EC products rather than a competitive threat, and so it entered the PC business. The open-systems organization of the PC permitted competition for the market in the PC, and for a period of time IBM was the leading open-systems architecture firm. IBM's success as the PC standard-setter in the 1980s was short lived because of the PC's open-systems organization and because of (once again) internal conflicts with serving both mainframe customers and a different type of corporate customer (or a different customer within the corporation).[34]

The potential future complementarity between the PC and other WCA computing technologies was an important long-run development. It opened up the distant future prospect of making existing EC systems or divisional computer systems much easier to use, thus dramatically increasing their effectiveness by employing the PC as a "client" for the EC system. Enterprise customers and IBM both saw the potential complementarity of the PC with

---

[34] See Bresnahan, Greenstein and Henderson (2008) for analysis of scope diseconomies within IBM's organization between the mainframe and PC markets.

mainframes. The co-existence of bureaucratic (enterprise or divisional) WCA systems and individual white-collar worker automation created a demand need for full integration of these systems. We shall return to that demand need when we examine our second example in the Section 4, for it remains one of the great unfulfilled demand needs of today.

As the 1980s came toward a close, a number of alternative GPTs served some part of WCA. The indirect route taken by reallocated innovative resources through hobby and entertainment applications (the PC) and scientific and engineering calculations (the minicomputer and workstations) had created alternative GPTs. Those technologies were finding some applications in WCA, pulled by the important demand force stemming from white-collar work automation that had not been served by IBM's version of EC. The growth performance of the rich economies reflected the progress being made to bypass the monoculture bottleneck. The best, however, was still to come.

3.5. Ultimate Lagged Impacts of Indirect Entry (1990s)

Over the course of the 1980s, there were three different economically important WCA markets, each with a GPT cluster. One was organized around the IBM mainframe and was a closed system. The other two, the PC and the superminicomputer plus its competitors, were organized around open systems lines. While IBM was serving the most valuable customers, the industrial organization of the other two markets was permitting much more rapid technical and market exploration. This coexistence was not sustainable.

The supermini caught up to IBM. Within that market, various sellers began making computers (and related fundamental software) which could only be reasonably characterized as mainframes (e.g. DEC). These computers came with rapidly improving software from competitive markets, including fundamental programmer tools such as the relational database management systems The competition in EC was intensified by the entry of new applications categories which could run on a wide variety of computers, whether from IBM or not. These included enterprise resource planning and customer relationship management applications, new important areas for WCA. IBM's entry barriers were at last overcome by the long, slow, indirect entry route. Over the course of the 1990s, workstations also became important in the supply of EC. The monoculture around the mainframe now became a competitive open-systems industry supplying "servers." Server computers could be mainframes, minicomputers or workstations, and enterprise server software came from a wide variety of suppliers not linked to any particular seller of hardware. Enterprise computing buyers could choose from a wide variety of tools, applications, and computers.

This change in industry structure overcame the IBM growth bottleneck and produced a dramatic burst in economy-wide white-collar productivity growth in the 1990s. Enterprise computing spread to a wider variety of uses. It grew dramatically cheaper, as well. The process of capital broadening enabled advances on an economy wide basis front, and the rich countries' economies had a dramatically good decade. Our demand-oriented framework shows that it was not technology alone, but the effective build-up of a full GPT cluster around minicomputers and workstations, which ultimately led to successful entry and competition in EC. More importantly, our demand-oriented framework reveals that the burst of aggregate productivity growth in the 1990s and, following a recession, in the 2000s, was <u>not</u> a result of the Internet in WCA.[35] Instead, growth arose from the recovered rate of progress in EC and progress in the use of PCs in WCA following the elimination of the IBM monoculture.

The success of EC markets today is the result of reallocation of innovative resources to other markets and ultimately re-entry into the primary market as firms and demanders found a way to work around very powerful barriers to entry. The market process to overcome those entry barriers was indirect and lengthy. Once that process allowed entry and created competition from alternative technological directions – GPTs which were alternatives to the IBM mainframe – the automation of white-collar work at the firm level accelerated and expanded beyond large firms. To a great extent, the near decade-long growth spurt of the 1990s can be attributed to the benefits of competition in the EC cluster. Our demand-oriented framework reveals that overcoming that bottleneck, and not the mass-market exploitation of Internet technologies, was the source for the economic growth in the 1990's.

[35] Many observers attribute the spectacular productivity gains of the late 1990s to the widespread use of the Internet and the resulting revolution in the automation of markets. These observers are correct that automation of markets will become an important element of automation of white-collar work, which involves a lot of buying and selling activities. But the productivity interpretation trips over the same problem of capital broadening we discussed earlier. By the late 1990s, for all the discussion of the dot com boom (and associated bubble), there is simply not enough installed capital in productive work associated with automating markets to drive the observed aggregate gains. Those gains came from the EC explosion documented above.

4. A current (incomplete) sequence: Individual Productivity Computing

We now turn to another sequence from even more recent history that exhibits the duality of SIRS and the centrality of demand: the important WCA area of IPC. The history of IPC shows the first two of the three stages of a sequence we saw in EC. Some of this history is familiar to scholars.[36] The most important GPT for IPC has long been the personal computer (PC). The PC went through a two-decade period of rapid improvement with strong demand influences on the direction of technical progress through market selection of leading technologies and supplier firms. It then transitioned to its current state, a proprietary standard, the Windows PC, with a single dominant firm, Microsoft. Today, entry barriers block efforts by other firms with new technologies to achieve widespread usage serving IPC. Demand's influence on the rate and direction of technical change in the PC is correspondingly small, creating a bottleneck. This bottleneck is particularly important in the era following the conversion of the Internet into a mass-market technology.

The next stage of the growth sequence is less familiar, but our demand-oriented perspective clarifies the likely route for indirect entry. IPC at work is not the only use of computer systems by individuals; there is also a great deal of non-work computing undertaken by consumers. Indeed, following on the conversion of the Internet into a mass-market technology, many PCs are used at home for a wide variety of consumption applications. A number of new GPTs have emerged to serve this large and potentially lucrative body of demand. These new GPTs support a large number of mass-market consumption (as opposed to white-collar production) applications. We have in mind a range of new server-side technologies, including search engines (Google), trading platforms (eBay), social networking sites (Facebook, MySpace, etc.), and a wide array of individual-user oriented social communication technologies like instant messaging, VoIP (Skype), blogs, tweets and so on.[37] Following on the diffusion of wireless telephony from business use to consumer use, a number of new "client" devices which are not PCs have emerged (e.g., "smart" cell phones and e-book readers).

The modern Windows PC and many of these new GPTs draw on the same BTO: the possibility of connecting individuals (reasonably) cheaply to networks. The Windows PC at

[36] See Bresnahan (2007) for a summary.

[37] There are, of course, some uses of these technologies which are valuable at work.

work is more and more involved in supporting the integration of individual workers into networked systems, connecting IPC to EC. The new GPTs have emerged by drawing on that BTO in a very different way than the Windows PC at work. Some (search, EBay, social networking sites) avoid competition with the existing GPT because they are complements to a PC rather than substitutes for it. Others (iPhone) avoid competition by being, in the present, distant substitutes. Although these new GPTs are technically different from the Windows PC (e.g. search is centered on the server side of the Internet rather than on the PC itself, and the iPhone has a different user interface than a PC), the most important factor in the potential for indirect entry by these new GPTs in WCA is that they serve a fundamentally diverse body of demand, centered in consumption.

We consider the possibility that these new GPTs, while not presently competitors for the Windows PC in IPC, might represent the early stages of indirect entry. To understand the present and attempt to forecast the future it is essential to introduce a demand perspective. These new GPTs do not (much) serve the critical growth need of WCA today. Already we see that innovative resources associated with a BTO have been reallocated away from investment supporting WCA toward consumption. In contrast to our discussion of EC in the last section, we do not have the luxury of writing about the future WCA applications of the new GPTs, for the process of indirect entry has not been completed. However, given the value to economic growth of future innovations that link individual white-collar workers into systems which automate bureaucracies and markets, it is worth attempting to forecast how indirect entry by these new GPTs could overcome the growth bottleneck.

4.1. Demand Influence: Market Selection of Original GPT

From the founding of the PC industry until the early 1990s, demanders frequently had the opportunity to influence the direction of technical change. While SIRS meant that there was only limited competition <u>in</u> the market, competition <u>for</u> the market was a powerful force. In contrast to the case in EC, competition for the market in IPC was not limited to a single initial phase. Until the early 1990s, regularly recurring opportunities for entrants to undertake competition for the market meant that no seller could afford to ignore demanders' concerns for too long.[38]

[38] See Bresnahan (2007) for an analysis of the sources of competition for the market in PC software.

The PC industry employed open compatibility standards as the mechanism for exploiting SIRS. Indeed, the standardization of the personal computer architecture as an open system was one cause of its widespread proliferation, delivering the beneficial side of SIRS. Meanwhile, vertical disintegration between key component markets created divided technical leadership. This meant that the industry was not dominated by any single firm, even if specific markets, such as operating systems, spreadsheets, or microprocessors, had dominant firms. A number of the important firms in the PC industry lost leading positions in particular component markets through this competition for the market. These include IBM (the PC itself), WordPerfect (word processor), WordStar (word processor), Lotus (spreadsheet), VisiCorp (spreadsheet), and DRI (operating system). While each of these firms was once dominant in a particular component of the PC, none were able to prevent entry and new competition, because they never had undivided technical leadership of the entire PC industry.[39]

New rounds of competition for the market occurred a number of times in the PC industry, and a number of existing dominant firms were swept away. That led to a rapid rate of technical progress. It also signified repeated market selection of leading firms and technologies. As a result, demanders' needs played a strong role in influencing the direction of technical progress in the PC GPT through the early 1990s.

4.2. Dominant Firm Becomes Bottleneck

Beginning in the 1990s and continuing until today, the market situation in the PC cluster is quite different. Demanders have not been able to use market selection as a tool to influence the direction of PC technical progress. A single firm, Microsoft, has made the Windows PC into a proprietary standard. Microsoft manages all technical progress in the GPT itself (the PC and close, universally-used complements, including programmer tools).[40] Supply of Windows, Office, the dominant browser and the main programmer tools has created a bottleneck around the PC GPT and the ability of the PC to be an invention enabling other inventions.

---

[39] The industry life cycle of the PC thus had several rounds of market selection, not a single one as is more common. For their history, and the relationship between divided technical leadership and their outbreak, see Bresnahan (2007). The most compelling models of industry life cycles treat only a single cycle of entry and exit e.g., Klepper (1996). The Klepper model also emphasizes uncertainty as the source of the entry and exit waves; the PC industry had renewed cycles of uncertainty because of changing demand and thus a series of "life cycles."

[40] There is ongoing open entry in new applications categories, as long as those categories are unlikely to contain a product used widely enough to become a GPT. That is because such a product is the main source of a threat to Microsoft's unified technical leadership of the PC GPT.

If one (incorrectly) takes a firm view, there is no question that the PC industry today is less vertically integrated than was the supply of the enterprise computing GPT in the era of IBM's dominance. However, this simply illustrates that the firm view is insufficient to determine whether a GPT market cluster is dominated by a single firm. For example, the supply of PC hardware by a competitive industry does not reduce the extent to which the PC industry is dominated by a single firm; neither does vertical disintegration of many applications or of the microprocessor. An important trigger for competition for the market in the PC industry's long period of demand responsiveness was divided technical leadership among suppliers of the PC's general purpose components. This is not a plausible source of renewed competition today. Intel, designer of the dominant architecture for PC microprocessors, once shared technical leadership of the "Wintel" standard, but today the "Windows" standard defines the most widely used PC. Disputes about GPT direction are resolved in favor of Microsoft.[41] Another effort to create divided technical leadership was Netscape's browser. The outcome of the browser war removed this threat. The prospect for a dominant browser supplied by anyone other than Microsoft has passed.[42] Today, there are no strategically important universally used PC components not supplied by Microsoft.

Given the degree of demanders' disaffection with this situation, it is not surprising that there have been a large number of entry efforts into the PC operating system market, such as Linux, the Macintosh OS, and Google's new PC operating system, Chrome. These serve primarily as illustrations of the technical feasibility of alternatives to the original GPT, not as real market threats.[43] The power of SIRS to preserve established incumbent positions keeps the market shares of these competitors low.

---

[41] Speaking of the resolution of a dispute with Microsoft in a joint interview with Bill Gates, Andy Grove, then CEO of Intel said, "We didn't have much of a choice. We basically caved." Schlender (1996). Microsoft can work with an alternative to Intel (AMD) while Intel has no real alternative to Microsoft for purposes of selling its products to the largest market of PC users.

[42] This point does not rely on an assertion that Microsoft's conduct in the browser wars violated antitrust laws. Both Netscape and Microsoft could foresee a non-Microsoft browser as a new source of divided technical leadership if it were dominant. From time to time browsers from other firms or open source browsers achieve moderate market share, but none since Netscape has gotten close to achieving dominance.

[43] Apple continues in second place to Microsoft Windows in operating systems, and many people are very excited about new entrants in PC operating systems. None has achieved real influence through becoming a real choice for demanders. The same situation holds for entrants in the word processing, spreadsheet, and other personal productivity software categories.

Another possibility would be a new outbreak of divided technical leadership in the PC. However, innovations that are very close complements for Windows will face the problem of being appropriated by a "strong second" strategy on the part of Microsoft and becoming part of, rather than an alternative to, Windows or Office. For example, netbooks have quickly become an alternative form of Windows PC, not an alternative GPT. Absent a dramatic and rapid change in the patterns of PC usage (such as the one brought about by the browser), entry barriers will defeat these initiatives.

The remaining possibility for indirect entry is through innovations not dependent on the Windows PC.

For many years, mainframe computers in EC were a monoculture, and the PC was open to a wide range of new ideas. Today, the PC GPT is a monoculture, closed to new ideas and participation, and server computers in enterprise computing are organized as an open system.[44] This reversal has relocated the growth bottleneck to the PC.

The widespread use of PCs to access the Internet following the invention of the browser and the commercialization of the Internet created a new GPT based in an important BTO. That BTO links individuals using computers to powerful networks. It is a broad technological opportunity because there could be any of a wide range of programs, data, or media on the network and because the "client" device used by an individual could be a PC or a very different kind of device; this technological variety is broad enough to support an extremely wide range of applications, both at work and in consumption.

Demanders in corporations eagerly sought a recombinant GPT from this BTO to fully integrate servers already in use for EC and PCs already in use for individual WCA. Any such new GPT would lead to further gains in WCA than were possible with either the IPC or with EC alone. In particular, any such new GPT would enable creation of applications that extend across firm boundaries and automate markets as well as firms. Today, the GPT that has emerged to fully integrate EC and IPC in WCA is the Windows PC connected via Internet Explorer. However, because of its control by a dominant firm, this particular GPT is not fully responsive to demand. As a result, unaddressed domains for WCA within the firm and the lack of extensions

---

[44] Leading suppliers of server components in different layers are attempting to end this open systems situation with a goal of creating a new dominant closed architecture. This is an important driver of the recent merger wave in EC.

of WCA beyond corporations and into markets remain the growth constraints of the early 21st century.

Another important GPT uses Windows PCs as a “client” to connect to EC applications through a browser. This has been an important area of the development of server-side functionality. The individual user connects to a system, such as an enterprise resource management (ERM), customer resource management (CRM) or related system. This has permitted more technical progress in traditional EC. [45]

These new GPTs, like others using PCs in the BTO of networked computing, are a complement to, rather than a substitute for, the PC. There is no serious mass-market alternative to the Windows PC as the “client” technology for new WCA applications in IPC; developers of EC applications for the PC overwhelmingly tend to write applications that connect to Windows, to Office, or to the dominant browser, Internet Explorer. Open, highly functional interconnect standards would permit application development by any of the wide range of competitive firms now participating in EC. Closed interconnect standards limit the participation to those approved by the proprietary standard holder, in this case, Microsoft. Less functional interconnect standards permit only less functional applications. Since Microsoft dominates the PC, any new GPT which is a complement to PCs would have to depend on Microsoft programming tools and access to Microsoft-controlled interconnect standards.[46]

Full integration also calls for cooperation of server-side firms in EC. However, given the competitive (current) structure of the server side, EC firms have neither the opportunity nor the incentive to attempt to control the direction of a new GPT. They are not the bottleneck.

Many observers suggest that a path to a new GPT will emerge as the more competitive server-side is more responsive to demand and adds new products and technologies demanders find particularly important. One key assumption is that the more rapid technical progress in the server side leads to the PC becoming less relevant. For example, “cloud computing,” a widely

[45] The browser also permitted cost-reducing and new ways to do old business within WCA (e.g., software as a service.) Our point is that the possibility of a new web-based platform with full integration has not emerged.

[46] Both the US and EU antitrust cases objected to Microsoft’s efforts to control interface standards between the PC and computer networks (the browser standard in the US and standards for security, authentication, and identity in the EU are important examples). The US also objected to Microsoft’s efforts to control key programmer tools areas (Java). The antitrust authorities objected to these efforts on narrow and specific market competition grounds. Absent the control of interconnection standards, Microsoft would face more competition for the PC operating system market (US case) or the workgroup server operating system market (EU case.) As a result, the claim that the antitrust cases were restrictions on innovation seems spurious.

supported current technology movement, would take advantage of any emerging reduced importance of the PC in IPC by moving many applications and data to the server side. The ideas brought forward by cloud computing advocates may well be technically valid. That will not, however, make the control of interfaces between the PC and servers in IPC go away. The ideas brought forward by critics of IBM mainframes and key complements a generation ago were technically valid. It was not those ideas, together with a frontal assault, that brought about the end of IBM's blockade. It was indirect entry. The same principle holds today. Cloud computing will not succeed as a direct effort to make the PC less important in fully-integrated white-collar applications. Within the WCA clusters related to IPC, the entry barriers mean that incumbent Microsoft is likely to successfully expropriate technological innovations from entrants. Market analysis suggests that we must wait for indirect entry.

The implication of single-firm control of key components such as interconnection standards and programmer tools is broad. It turns a GPT market cluster into a monoculture. In this case, the GPT serves WCA to the extent it involves individual workers. Within that scope, there is single-firm control over applications, and its toolkit for developing new applications. The participation of a wide number of companies in this GPT market cluster would be immediate if not for their need to get the approval of a single dominant firm. This blockaded situation is similar to the one managed by IBM a generation ago.

4.2.1. Monoculture Blocks Change in Technical Direction for Original GPT

Even when an existing GPT supplied by a dominant firm blocks entry, innovation introduced by that firm can take advantage of a BTO. While likely slower than market innovation, this mechanism can get around a growth bottleneck. In the present case, this seems an unlikely path.

Microsoft was initially slow to recognize the value of the browser as a commercial tool and distribution channel to the mass market PC user, so they were slow to respond to the Netscape browser. When they finally did respond, Microsoft used control of distribution to PC demanders to block Netscape's continued diffusion into the market.[47] Microsoft was able to extend the entry barriers protecting its PC position to cover the most important new piece of PC software in

[47] There is a large literature debating whether it was unlawful distribution advantages or product quality advantages that led to the browser war outcome. Citations and econometric evidence can be found in our earlier paper, Bresnahan and Yin (2009).

years, the browser. Today, Microsoft is the dominant firm in the browser market, and, for the browser used in WCA, the unassailable leader.[48]

As we emphasized above, one of the roles of the single dominant firm in a GPT is choosing which directions of technical change are taken up within the GPT. Not long after coming to dominate the browser market, Microsoft made an important decision. Rather than pursuing a technical direction in which the browser and networked applications played a central role, the firm chose a direction in which Windows remained central to the PC platform. This decision followed a long debate within the firm. A pro-Internet faction had proposed that "The Web is the next platform."[49] A more technologically conservative faction instead argued that "Windows is the platform."[50] The firm chose the latter strategy. This meant a firm-level focus on extending the role of the existing Microsoft position of dominance of an existing GPT, not of creating alternative GPTs.[51]

Although Microsoft became the dominant firm in the browser market, its strategies for extending the reach of Windows led the firm to eschew the browser as the key technology for creating an alternative GPT. Evidence of the growth bottleneck that Microsoft imposed on the potential browser market is the lack of Internet security for those connected by a PC. Although a single, dominant firm would be well suited to establish and enforce safeguards, the lack of Internet security has prevented larger dollar-value of transactions from occurring online, curbing the power of SIRS and positive feedback loops to develop around a fully-integrated GPT. Thus a path around the current bottleneck led by the main incumbent dominant firm seems unlikely.

It is worth clarifying here that we are not claiming that Microsoft completely blocked the development of Internet applications; one only has to observe the innovations all around us today. However, we do claim that the scale and rate of progress in WCA applications taking advantage of the Internet or otherwise fully integrating PCs with servers is much smaller and slower than it could have been in the absence of the Microsoft monoculture.

---

[48] For browsers used outside of WCA, the situation is more open. We turn to this below.

[49] This is the title of Slivka (1995), one of the manifesti of the eventually losing side in the internal conflict.

[50] Jim Allchin, eventual winner of an internal battle, wrote "The platform is Windows, isn't it? duh… it would seem obvious," in Allchin (1997).

[51] See Bresnahan, Greenstein and Henderson (2008) for the argument that this decision was driven by scope diseconomies between the existing Windows line of business and the potential new Internet line of business. Bank (2001) has many colorful anecdotes about the internal conflict that ultimately led to the decision to stay with Windows as the sole platform.

4.3. Reallocation of Innovative Resources to Unserved Demand

The BTO associated with mass-market networked computing is available to a wide range of firms. Some are barred from taking advantage of this BTO in WCA; this creates a technological push. At the same time, the browser and related technologies have created a new consumer-use mass-market for networked applications. But not all networked consumption activities are well served by the PC, especially not all entertainment-oriented networked consumption activities. This fundamental diversity in demand has created a demand-pull for new GPTs.

These forces have led to the reallocation of innovative resources away from WCA and to the creation of new GPT clusters primarily serving consumers. One set of these is based in server computers connected to users via the browser. These new GPTs avoid the Windows PC bottleneck by offering only limited functionality on the "client" (PC) side, basing much of their computation on powerful "server" computers. An example is search, which in Google's implementation has powerful server computers, enormous data collection on the WWW, and complex algorithms – all off the PC. The connection to the PC is through a simple browser interface. Each of these new GPTs is close enough to the WCA demand for fully integrated EC servers and PCs that one can envision their indirect entry back into the WCA market cluster.

Many observers explain the emergence of new mass-market consumer computing GPTs on the server as the result of a new, technical comparative advantage of server computers.[52] Our framework provides a better explanation by including demand: server-side technologies are the only way competitors exploiting the BTO can get around the PC bottleneck, and consumer demand is pulling new GPT invention to serve its needs. The browser created a new way to distribute applications to the mass-market computer consumer. Up to this point, almost all computing technology which was distributed to the mass market consumer was distributed through the PC, over which Microsoft had control. The browser could allow firms with applications, and even non-computing commercial businesses with websites, to access the PC user without owning any "real estate" on the PC itself. Access by consumers to server-side

[52] These explanations have some difficulty dealing with the apparent acceleration in the rate of technical progress in PC hardware components such as the microprocessor and memory, a fact which appears to cut against any theory positing a purely technical advantage for large computers.

applications through cell phones and other smaller-than-PC client devices also moves the applications farther away from the bottleneck.

Search thus far has been an advertising-supported GPT which relies largely on consumers searching for entertainment, product information, and so on. A critical mass of demand for search generates co-invention around marketing applications, taking advantage of information revealed by potential customers. Based on a database of this more personal information, the applications here answer the needs of mass-market consumers better than Microsoft but do not yet answer the needs of corporate individuals. The interaction of this information with the Internet as a distribution channel particularly suited to information transmission could lead to a new GPT that automates marketing in general. At the moment, however, the application of mass-market information resembles traditional advertising-supported media rather than a new WCA GPT for markets.

The browser opened opportunities to create new markets based on the mass consumer. Entrants like eBay could take advantage of large numbers of consumers connected via the Internet to form thicker consumer-to-consumer or business-to-consumer markets for exchange. While not nearly as pervasive or as powerful as the online markets forecast during the dot.com boom, these technologies could become a GPT that enters WCA indirectly. When applied to businesses (e.g. Freetrade), these markets offer the possibility of WCA of procurement relationships and one-to-one negotiations into arms-length, market transactions.

The ability to create thick markets on the Internet has also introduced new modes of entertainment (e.g., Facebook, MySpace, Twitter, YouTube), which allow consumers to create content for other consumers. These sites run on a server and are not a close complement for Windows, utilizing the browser as a platform instead. Social networking has been a popular destination for reallocated resources because it plays to the strengths of the server side. The constraints and disadvantages of the PC experience are avoided by tapping into off-line social networks and social networking activities. The core business of these sites is to link people as consumers, so these sites are not a direct threat to Windows or Office.

However, one of the most important demand needs of fully-integrated IPC WCA is collaboration among workers. This has led to repeated claims that the indirect entry process has been completed and that a consumption technology is being used in WCA. Throughout the last decade, we have seen the announcement of the corporate blog, corporate wiki, and/or corporate instant messaging. Each of these supposedly successful WCA technologies draws on consumer mass-computing GPTs that have been tremendously successful, but none has achieved that kind of usage in WCA. Despite this series of overhyped failures, the call in WCA for a collaboration technology is real, and a new, online consumer social GPT may eventually succeed in entering WCA.

The Internet and browser provide a new channel through which information goods can be delivered. As a result, music has been a particularly important source of new, online GPTs. Hollywood's stunning sloth in moving into the Internet era has left a demand opportunity for the new delivery of music. One of the new GPTs, sponsored by Apple, may become an important indirect entrant into computing. Apple began with an online music business closely linked to a client device, the iPod. This has put the firm into the position to act as an indirect entrant into telephony already, via the iPhone. The AS for devices such as the iPhone involves creation of applications by consumers for other consumers, with an open-ish applications development environment. More important for our inquiry, the AS of the iPhone includes professional developers seeking to deliver (notably, among many other smaller categories) marketing information to iPhone-carrying customers. While not nearly as functional as a computer, the smart cell phone (which is growing more competitive as imitators of Apple enter) is an important consumer client device with obvious potential applications in IPC-WCA. The ease of application development for the iPhone means that some small organizations can customize applications for internal use. Although originally marketed to consumers for consumption applications, the iPhone benefitted from SIRS in that mass-market installed base. The iPhone and other smart phones are slowly entering the enterprise market at a very small scale. This indirect entry may be originally in the form of consumption-on-the-job, and at this writing most enterprises view the use of smaller-than-PC devices by employees as a support headache rather than as an applications development opportunity. As the installed base of these devices in the hands of employees grows, they may become an indirect entrant and erode the PC's hold on businesses.

The new entertainment GPTs associated with search, the iPhone, social networking, etc. share a crucial feature of the PC in the (very early) days when it provided primarily entertainment. (In the case of the PC, this was entertainment for techies.) An enormous amount of "small" innovation by somewhat technically knowledgeable users is occurring today. The type of invention here, due to its mass-market nature, is prone to recombination through mash-up and small tool development. Tools for developing applications are inherently a GPT (as Rosenberg (1976) points out). The form of the tools being created for mash-up, etc., may be inappropriate for corporate endeavors, but adding the structure and control needed in that area might be a key step in indirect entry.

Innovative resources have been reallocated away from a crucial growth demand need, WCA. One of the destinations of the reallocation is entertainment, far from the demand served by the existing dominant firm in WCA. The productivity implications of this reallocation echo those of the IBM era: the productivity gains from entertainment are less immediate than if these resources could be allocated to solving the WCA constraint, and the nature of indirect entry suggest a significant lag before the alternative GPT in entertainment translates into WCA productivity gains.

4.4. Indirect Entry?

The demand conditions for the first steps of indirect entry are satisfied: a number of new categories of mass-market computing GPTs have been invented which evade the growth bottleneck of the Windows PC. Applications of these new GPTs tend to emphasize entertainment, at least in the present, rather than WCA. Once again, we see the value of fundamental diversity in demand. . We sketched the potential for several indirect entry paths in the last subsection; will these new GPTs complete the sequence and create new opportunities for WCA with full integration? We cannot know the answer since the sequence stage of indirect entry is still only beginning, but we can consider what conditions must be met in order to complete the indirect entry process.

For indirect entry to occur, some creation in an alternative market cluster must be of general enough use that it expands beyond the original sources of revenue (in the lower-value area) back toward WCA and full integration. In order to achieve this expansion, the alternative market cluster must be valuable enough to fund technical progress in the alternative GPT and have a large enough body of demand to support AS growth, fueling positive feedback cycles and SIRS. A few of the new GPTs, including Facebook and Twitter, have found valuable enough applications in the entertainment realm to fund development of technical improvements. Also, sponsoring firms have grown up at the center of the new GPT clusters who might lead an indirect entry effort. Google and Apple, for example, have the resources for a general push. Many of these alternative GPTs have created a positive feedback cycle involving a large number of applications and a large number of users.

To complete the indirect entry cycle, the alternative GPT must eventually appeal to the WCA customers of Microsoft. Critically, this includes corporate IT departments at least as much as end users. The most obvious path would come via buyers in corporate computing environments finding non-PC solutions compelling for a wide range of applications involving a large number of workers. At the moment, however, there is no serious prospect for this.

Another path by which indirect entry could be completed would focus on the role of much computer-based WCA, whether IPC or EC, as occurring in service industries or in the service functions of other industries. Service industries are characterized by the customer doing work. So are some of the service functions, such as selling, in all industries. (Similarly, buying often involves the supplier doing work.) Automation of white collar work which shifts the locus of the work across the boundary of the firm, i.e., in part to the customer or supplier, may draw on new consumer-oriented technologies. This is most obvious in consumer goods and services industries. But it could also occur in other industries to the extent the customer or supplier is an individual person whose sense of how to use a computer system is linked to a new, non-PC GPT.

One path which may potentially allow indirect entry to be completed is the ongoing devaluation of the PC by alternative computer-like devices such as smart cell phones. In the current situation, PCs and phones are not particularly close substitutes. Nonetheless, many of the same workers who use a PC also use a smart cell phone for partially overlapping tasks, such as email. As the worker comes to have more and more applications programs and data connected to the cell phone rather than the PC, corporate applications developers may begin to find opportunities to further encroach on the PC's turf. To the extent users store and manipulate data stored on a server somewhere rather than on their PC (i.e., to the extent there is "cloud computing"), such applications will be easier for developers. This still appears to be a long haul over difficult ground and one would be unwise to bet both that the Windows PC is unconnected to the "cloud" and that Microsoft has no prospect of making a proprietary cloud-like technology. The trend, however, is slowly and incrementally for worker's computer connections to be less PC-centric.

Maintaining dominance over Windows, Office, Internet Explorer, the PC industry and fully-integrated WCA as a closed system is a difficult business and technical challenge for

Microsoft that has led to severe problems in a number of areas, such as computer security. As home PC use generates more and more utility through Internet applications relative to work PC use – even if both are Windows PCs – a movement of end users and corporate developers demanding equivalent utility from work PCs and replicating non-WCA methods to evade the constraints of the corporate PC could break out. Given the inertial forces surrounding the Windows PC SIRS, this would need to be a large movement and one that involved a large number of new applications or new users (or both) to succeed.

Finally, one of the valuable GPTs developed on the server side could find important WCA applications. While none are yet visible at large scale, at least two paths seem possible. One is that advertising-supported consumer-oriented entertainment websites will develop new marketing applications. Those could become a basis for electronic commerce in mass markets. That would position a new server-side GPT with a strong demander constituency among marketers (and their customers) as a potential entrant. Such a GPT might be able to compete in WCA more broadly from its base in marketing applications, just as the supermini and the workstation were ultimately able to compete with mainframes from their base in SME/divisional WCA. Similarly, one or more of the technologies under social networking or other consumer-to-consumer communication could become used in organizations as a collaboration technology. In situations characterized by a lack of PC infrastructure (health professionals coordinating crisis relief efforts, operations in developing countries), these platforms combined with more mobile non-PC devices might begin to play a WCA role. If it were being used in WCA in a large-scale way, a technology born in social networking could become a potential entrant as personal use spills over into work use.

Finally, we note that the decreased dependence of consumer mobile computing on PCs and the browser is opening up a range of market experiments. Some of these are more closely linked to telephony, like the development of "lightweight" applications for smart cell phones. This alternative path and these experiments are encouraging for the prospect that an entirely different part of the broad technological opportunity associated with networked computing, perhaps one more telephony-centric than PC-centric, could become part of WCA.

These prospects seem possible if not imminent. An outbreak of competition across multiple full integration GPTs for growth would enable demanders to create a wide range of valuable new applications. This would be the third part of our sequence, and the prospect for the future is once again the elimination of bottleneck allowing the economy to experience a growth spurt.

5. <u>Conclusion</u>

We have built a demand-oriented framework to understand the role of GPTs in growth. We begin with a conventional model of the role of a GPT. We posit a demand need present in many industries that forms a growth constraint which can be answered by a GPT spawned from a BTO. The conventional story of a GPT arising, addressing that growth constraint and driving productivity gains forms the first phase of our dynamic sequence.

We depart from that conventional view in two ways which reflect our focus on demand. First, the broad nature of the technological opportunity lying behind many GPTs leads to diverse bodies of demand: some of these are more valuable in addressing the growth constraint than others. In these circumstances, GPTs will likely manifest as different clusters of technologies that answer different types of demand, since not all demand can be addressed by a particular technology.

Second, a bottleneck can arise with a successful GPT. This bottleneck will be associated with the dominant firm supplying the GPT, if there is one, but the existence of a bottleneck is not primarily a matter of firm behavior. We focus on the original GPT, the one which is most valuable in use because it is closest to addressing the growth constraint. The SIRS associated with a GPT mean that successful exploitation involves selection of some firms/technologies for the original GPT market cluster to the exclusion of others. Rapid productivity gains follow, but if a firm comes to dominate the market in that cluster, it may create a bottleneck and slow productivity gains by implementing entry barriers in that market. This limits demanders' influence on the direction of technical progress.

If there is a bottleneck but also fundamentally diverse demand, innovative resources are pushed into other markets not subject to entry barriers and pulled into markets which remain unserved by the original GPT but which could be met by the BTO. While these alternative markets need not be as valuable as the original market, they may be valuable enough to generate a cycle of co-invention between demanders and suppliers to drive technical progress in an alternative GPT. The alternative GPT may possibly evolve into a technology competitive enough to re-enter the market serving the original demand need. This indirect entry bypasses the bottleneck and returns competition in the more valuable market, accelerating productivity gains by re-introducing better, alternative GPTs to address the growth constraint.

Our demand-oriented framework permits us to recognize the importance of diversity in demand in providing a market solution to a productivity bottleneck. The importance of similarity in demand is also highlighted by our framework, producing a process by which resources allocated away from a bottlenecked market will remain within markets sharing demand for a BTO. This creates the possibility of indirect entry into the original GPT market at a later period. We apply this framework to understand the sequence of large contributions to growth, slowdown, and renewed growth in enterprise computing, and, we hope, IPC. While the IPC sequence is still underway, the reallocation of resources around the growth bottleneck has already occurred, and key market conditions that would permit indirect entry seem to be satisfied.

Bibliography

Acemoglu, D., 2002, Directed Technical Change, *The Review of Economic Studies*, **69**, pp. 781-809.

Adner, Ron, and Levinthal, Daniel, (2001) "Demand Heterogeneity and Technology Evolution: Implications for Product and Process Innovation" *Management Science*, **47**, pp. 611-628

Adner, R. and Zemsky, P., 2005, Disruptive Technologies and the Emergence of Competition, *The RAND Journal of Economics*, 36, pp. 229-54.

Aghion, P. and Howitt, P., 1998, On the Macroeconomic Effects of Major Technological Change, *Annales d'Economie et de Statistique*, pp. 53-75.

Allchin, J., 1997, "Losing a Franchise -- The Microsoft Windows Story" (a new Harvard Case Study). Microsoft Email.

Bank, D., 2001, *Breaking Windows: how Bill Gates fumbled the future of Microsoft*, New York: Free Press.

Baumol, W. J., 1967, Macroeconomics of Unbalanced Growth: The Anatomy of Urban Crisis, *The American Economic Review*, **57**, pp. 415-26.

Bresnahan, T. (2010) "General Purpose Technologies", *Handbook of the Economics of Innovation and Technical Change*, Edited by B. Hall and N. Rosenberg, Elsevier North-Holland.

Bresnahan, T., 2007, Creative Destruction in the PC Industry, In: *Perspectives on Innovation* (eds Brusoni, S. and Malerba, F.). Cambridge University Press, Cambridge.

Bresnahan, T., Greenstein, S. and Henderson, R., 2008, Schumpeterian competition and diseconomies of scope; illustrations from leading historical firms in computing, In: *Working Paper*. Stanford University.

Bresnahan, T. and Malerba, F., 1998, Industrial dynamics and the evolution of firms' and nations' competitive capabilities in the world computer industry, In: *The Sources of Industrial Leadership* (eds Mowery, D. and Nelson, R.), Cambridge.

Bresnahan, T. and Yin, P.-L., 2009, Economic and Technical Drivers of Technology Choice: Browsers, *Annales d'économie et de statistique*.

Bresnahan, T. F., 2002, Prospects for an Information-Technology-Led Productivity Surge, *Innovation Policy and the Economy*, **2**, pp. 135-61.

Bresnahan, T. F. and Greenstein, S., 1999, Technological Competition and the Structure of the Computer Industry, *The Journal of Industrial Economics*, **47**, pp. 1-40.

Bresnahan, T. F. and Trajtenberg, M., 1995, General purpose technologies: Engines of growth'?, *Journal of Econometrics*, **65**, pp. 83.

Christensen, C.M., 1997, *The Innovators Dilemma: When New Technologies Cause Great Firms to Fail* Harvard Business School Press, Boston, MA 1997

Clark, K.B., (1985) "The Interaction of Design Hierarchies and Market Concepts in Technological Evolution," *Research Policy* 14 (1985), pp. 235–251.

Dosi, G., 1982, Technological paradigms and technological trajectories: A suggested interpretation of the determinants and directions of technical change, *Research Policy*, **11**, pp. 147-62.

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

Farrell, Joseph, and Garth Saloner, "Standardization and Variety," *Economics Letters*, 1986, 20, 71-74.

Gans, J. and S. Stern (2003). "The Product Market and the Market for Ideas: Commercialization Strategies for Technology Entrepreneurs," *Research Policy*, Vol.32, No.2, February, 2003, pp.333-350.

Gilbert, R. J., 1992, Symposium on Compatibility: Incentives and Market Structure, *The Journal of Industrial Economics*, **40**, pp. 1-8.

Griliches, Z., 1957, Hybrid Corn: An Exploration in the Economics of Technological Change, *Econometrica*, **25**, pp. 501-22.

Griliches, Z., Berndt, E. R., Bresnahan, T. F. and Manser, M., 1992, *Output measurement in the service sectors*, Chicago: University of Chicago Press.

Helpman, E., 1998, *General purpose technologies and economic growth*, Cambridge, Mass.: MIT Press.

Helpman, E. and Trajtenberg, M., 1998, "A Time to Sow and a Time to Reap; Growth Based on General Purpose Technologies" in E. Helpman, ed., *General Purpose Technologies and Economic Growth*, MIT Press, Cambridge, Mass.

Jaffe, A. B., 1986, Technological Opportunity and Spillovers of R & D: Evidence from Firms' Patents, Profits, and Market Value, *American Economic Review*, 76, pp. 984-1001.

Jovanovic, B. and Rousseau, (2005), "General Purpose Technologies", *Handbook of Economic Growth*, Volume 1B. Edited by Philippe Aghion and Steven N. Durlauf (c) 2005 Elsevier B.V

Klepper, S., 1996, Entry, Exit, Growth, and Innovation over the Product Life Cycle, *American Economic Review*, **86**, pp. 562-83.

Lipsey, R. G., Carlaw, K. and Bekar, C., 2005, *Economic transformations: general purpose technologies and long-term economic growth*, Oxford ; New York: Oxford University Press.

Mokyr, J., 1990a, *The lever of riches : technological creativity and economic progress*, New York ; London: Oxford University Press.

Mokyr, J. and Scherer, F. M., 1990, *Twenty five centuries of technological change : an historical survey*, Chur ; New York: Harwood Academic Publishers.

Rosenberg (1976) *Perspectives on technology*, particularly the chapter entitled "Diffusion and Adaptation of Technology."

Rosenberg, N. and D. Mowery, (1979) "The influence of market demand upon innovation: a critical review of some recent empirical studies," *Research Policy* **8** pp 102-153

Rosenberg, N, R. Landau and D, Mowery, eds, *Technology and the Wealth of Nations*, Stanford University Press 1992

Ruttan, V. W., 2001, *Technology, growth, and development : an induced innovation perspective*, New York: Oxford University Press.

Schlender, B., 1996, A Conversation With the Lords of Wintel, In: *Fortune*.

Slivka, B., 1995, The Web is the Next Platform. Microsoft Email.

Usselman, S., 1993, IBM and its imitators: organizational capabilities and the emergence of the international computer industry, *Business and Economic History*, **22**, pp. 1-35.

Figure 1: Broad Technological Opportunity, Fundamentally Diverse Demand and GPTs

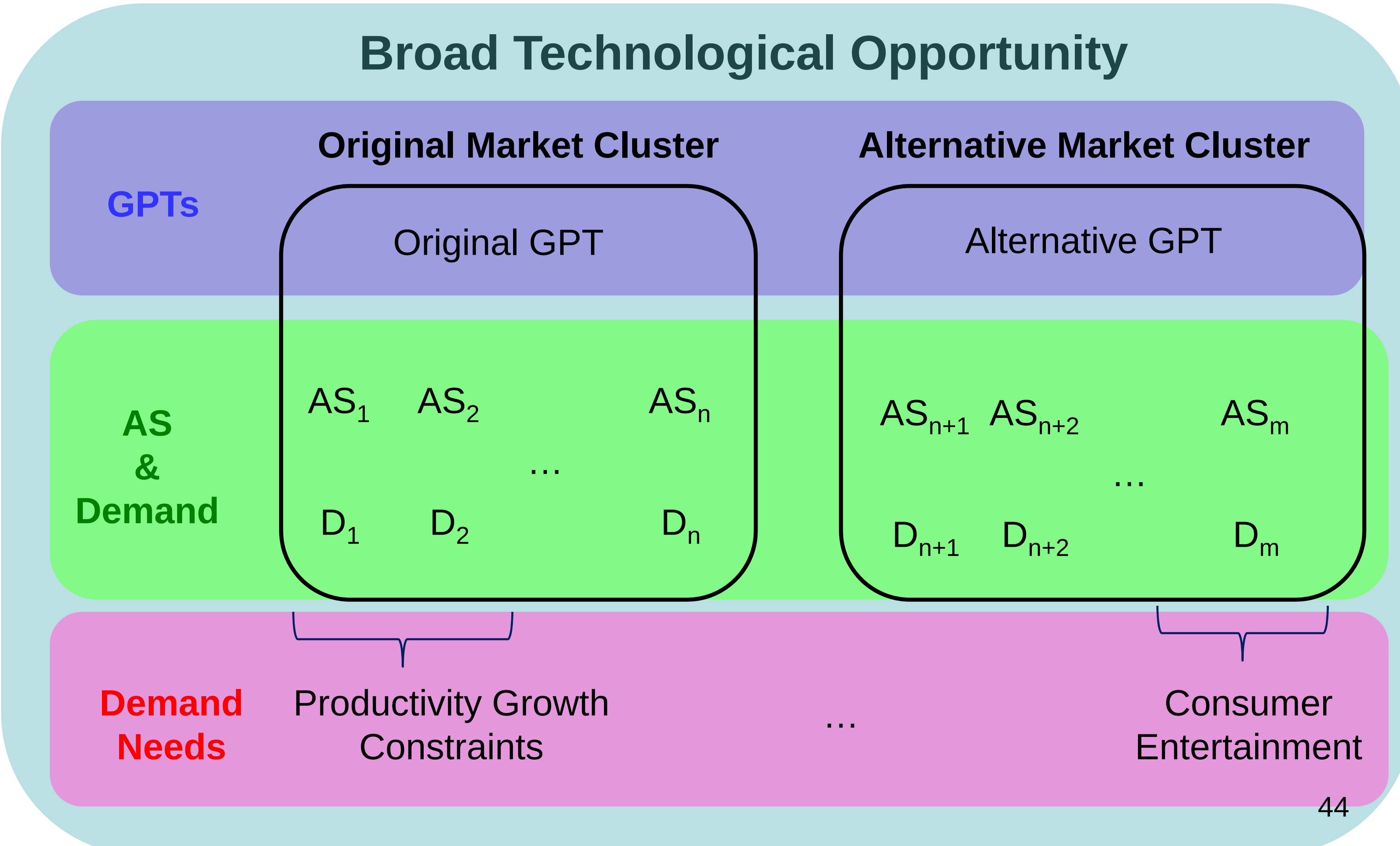

Figure 2: Broad Technological Opportunity, Fundamentally Diverse Demand and IBM

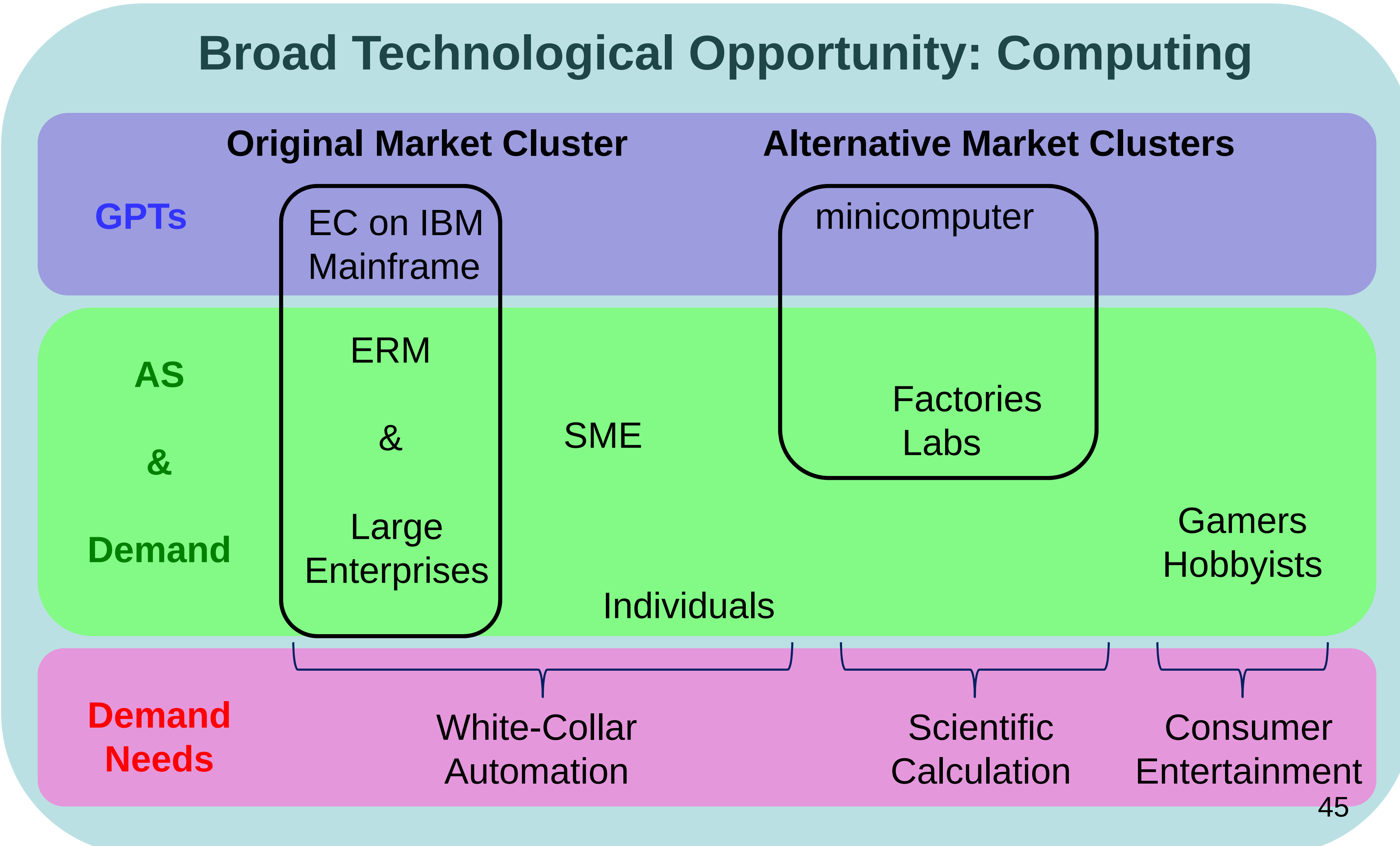

Figure 3: Reallocation to Unserved Demand

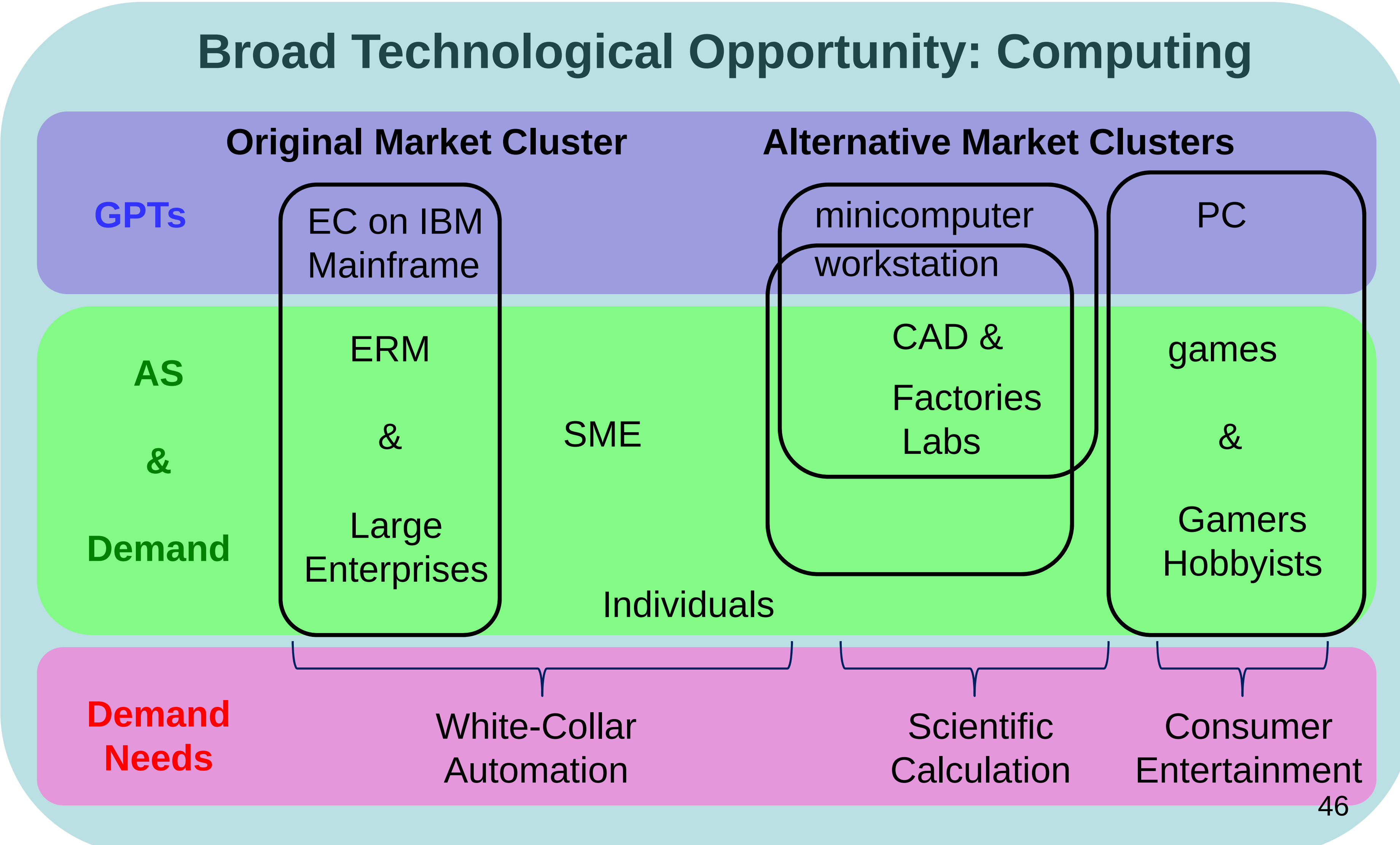

Figure 4: Indirect Entry