The S&T Innovation Conundrum

Timothy Coffey, Jill Dahlburg, and Eli Zimet

Center for Technology and National Security Policy

National Defense University

August 2005

The views expressed in this article are those of the authors and do not reflect the official policy or position of the National Defense University, the U.S. Naval Research Laboratory, the Department of Defense, or the U.S. Government. All information and sources for this paper were drawn from unclassified materials.

Timothy Coffey holds the Edison Chair for Technology in the Center for Technology and National Security Policy, The National Defense University. He is also a Senior Research Scientist at the University of Maryland. He was the Director of Research at the U.S. Naval Research Laboratory (NRL) from 1982 to 2001. He has published extensively in the scientific literature and has served on numerous advisory panels. Dr. Coffey holds a BS (EE) from the Massachusetts Institute of Technology and a PhD (Physics) from the University of Michigan.

Jill Dahlburg is the Senior Scientist for Science Applications at NRL. Her technical collaborations have included scientists in both the national and international physics and engineering communities. She has served on a number of review, editorial and advisory boards, and is 2005 Chair of the American Physical Society (APS) Division of Plasma Physics. Dr. Dahlburg, a Fellow of the APS, holds a BA in liberal arts from St. John's College in Annapolis and a PhD in plasma physics from the College of William & Mary in Virginia.

Elihu Zimet is a Distinguished Research Fellow at the Center for Technology and National Security Policy, The National Defense University. He is also a Senior Research Fellow at the Potomac Institute for Policy Studies. Prior to this he headed the Expeditionary Warfare Science and Technology Department at the Office of Naval Research. In this position he directed science and technology (S&T) programs in missiles, directed energy, aircraft, and stealth as well as S&T support to the Marine Corps. Dr. Zimet holds a BS (ME) from the Polytechnic Institute of Brooklyn and a PhD from Yale University.

Defense & Technology Papers are published by the National Defense University Center for Technology and National Security Policy, Fort Lesley J. McNair, Washington, DC. CTNSP publications are available online at http://www. ndu.edu/ctnsp/publications.html .

Contents

List of Figures

List of Tables

Executive Summary

This study is motivated by the observation that the state of health of the United States S&T enterprise seems to be simultaneously characterized by opposite assessments. On the one hand the enterprise is described as being especially vibrant, showing remarkable progress, a high level of innovation and confronted with great opportunities. At the same time the enterprise is described as showing disturbing trends in its workforce, rate of knowledge generation, rate of innovation and international standing.

The purpose of the study is to shed light on how this conundrum has come about, and from this perspective to evaluate potential impacts of the underlying drivers of the conundrum on the technological positioning and ultimate national security of the United States. The process employed is to examine various aspects of the apparent innovation paradox by reviewing historical data regarding scientific and technical progress, and by analyzing how S&T innovations occur. In support of this, the concept of research and development (R&D) innovation space is introduced, and a few elementary models are presented for illustration.

The study suggests that we have lost sight of some key realities. We have become so mesmerized by our enormously successful exploitation of past S&T breakthroughs that we have forgotten how they happened. Since society is primarily interested in the creation of functional capability (e.g., computing power), this memory lapse becomes problematic with respect to maintaining a pipeline of future breakthroughs.

For example, the rapid advances in electronics and computer products over the past 50 years have created a general impression of continuous scientific breakthroughs. In reality, the breakthrough S&T innovations for electronics and computers took place in the 1940's and 1950's. The subsequent rapid advances in functional capability were the result of a brilliant and enormously successful program to exploit those early breakthroughs. An unfortunate byproduct of this success was the impression that these rapid advances in functional capability also represented the time scale for S&T breakthroughs.

An examination of the histories of a number of major S&T innovations covering the past 100 years indicates that today a breakthrough innovation takes 15-20 years to progress through the early phases, just as it did 100 years ago. There is often no functional capability produced during this early phase. Our fixation with exploitation and near term profits is incompatible with the realities of these time scales. This is problematic due to its effect on investment strategies.

In reality there are two distinct phases in S&T innovation. For a successful innovation, once the underlying S&T is in hand as a result of the early phase work, a second phase can be undertaken where rapid technical progress resulting in significant new functional capability is possible with the application of adequate financial and human capital. The time scales for progress are much faster than in the early phase and are based on the potential of the technology and the resources applied.

The characteristics of these two separate phases are captured by the descriptors "Prospecting" and "Mining" respectively. There is a dynamical relationship between the prospecting phase and the mining phase. The long-term health of one depends on the health of the other. Both phases

also involve conducting basic and applied research as well as exploratory development (but with a different mix). However, the two phases are fundamentally different and require different governance. Unfortunately we seem to have forgotten this, thereby contributing to the S&T innovation conundrum.

An essential aspect of governance is the allocation of resources (e.g., people and funding). Economists, understanding the important role that technology plays in economic growth, have begun developing theories regarding the impact on economic growth when R&D investments are determined so as to maximize profits. This is referred to as an endogenous investment strategy. These theories are helpful in discussing how economic conditions, combined with an endogenous investment strategy for R&D, affect the scientific and technical talent pool and the generation of knowledge as well as economic growth. They also shed light on how knowledge affects long-term economic growth. It seems clear that a solely endogenous approach to determining R&D investments results in too little long-term research being funded. Talent and resources gravitate to the mining phase at the expense of the prospecting phase and at the expense of the knowledge generation needed to sustain economic growth in the long term. In the short term, however, the private sector, using profit maximization techniques, is extremely effective at introducing innovations that exploit science and technology developments thereby maintaining the U.S. competitive advantage. Part of the S&T conundrum is related to the balancing of these conflicting attributes. Proper balancing of these competing outcomes should be a byproduct of the separate governances required for the prospecting and mining phases of innovation. This suggests a national imperative for an exogenous (i.e., not determined by near term economic and profit maximizing considerations) determination of the R&D investment for the prospecting phase.

An examination of R&D funding data since the 1950's suggests that the United States is comfortable with a steady state investment in R&D of about 2.5 % of the Gross Domestic Product (GDP). The total investment seems to have oscillated about this value for 50 years. It seems reasonable that a nation's need and ability to support R&D should be proportional to the economy that the R&D is intended to support. The observed steady relationship to GDP, therefore, is not surprising.

The federal share of R&D (measured relative to the U.S. GDP) has mostly declined since 1965, while the industry share has mostly increased over the same period. Since the industry investment in R&D is—and should be—mostly endogenous, this has raised concern about the long-term investment in the prospecting phase of R&D. This concern is responsible for part of the conundrum. The severe competitive environment created by globalization has left the federal government as one of the few entities in a position to take responsibility for the long-term knowledge necessary for long-term economic growth. Unfortunately the federal government has also moved towards an endogenous approach for the determination of its R&D investment. This seems to be driven by the reasonable objective of justifying the government's R&D investment. However, it is interesting to note that many of the S&T innovations upon which today's society is based would never have been funded in their early phase had they been subject to a purely endogenously determined R&D investment strategy. It may be that the very attempt to measure output on too short a time frame is producing a research program that operates on too short a time frame hence eliminating its true value.

In the long term, knowledge is a fundamental pillar of economic growth and military power. Therefore, understanding how knowledge grows is important to proper governance. Here again a reality check is needed. It seems that we have concluded that, in today's world, knowledge grows at an ever-increasing rate and actually feeds on itself. This can only be possible if individual scientists and engineers are themselves producing knowledge at an ever-increasing rate. However, an examination of patent and publication data for the past 50 years indicates that the rate of new knowledge production at the level of the individual scientist and engineer has not increased and may have actually declined. This reality should place significant constraints on models that purport to predict the rate of growth of scientific and technical knowledge. Failure to do so results in unrealistic expectations and further contributes to the conundrum. Knowledge cannot grow far more rapidly than does the scientific and technical workforce without any increase in the rate of knowledge generation at the level of the individual scientist and engineer.

This part of the conundrum may be related to the fact that the enormous advance in automation of routine functions such as data collection and data analysis has greatly improved productivity with respect to these routine tasks. However, such improvements should not be confused with increases in the rate of knowledge generation.

A related contributor may be that new functional capabilities such as computers, advanced fabrication technology and increasingly sophisticated analytical instruments have enabled scientists and engineers to work in new regimes (such as at the nano scale). The ability to now work in these new regimes is not equivalent to an increase in the rate of knowledge production. It is human beings who produce new knowledge. New knowledge production for these regimes is still paced by the rate-limiting step of human cognition and understanding that has evolved slowly over many millennia. This rate-limiting step is analogous to those that occur in many dynamical systems such as chemically reactive systems. Speeding up the flow of information and data does not automatically translate into new knowledge production. Indeed it can divert the attention of scientists and engineers, resulting in the opposite effect.

Part of the conundrum is also related to the differing nature of technologies. There are certain technologies, called general purpose technologies (GPTs), that are characterized by broad applicability over many segments of society and act as enablers for societal development. Examples of such technologies are electrification technologies and information technologies. The infrastructure that accompanies the diffusion of GPTs throughout societies create opportunities for further technology development and innovation, and focus financial and human capital around these opportunities. This has the positive effect of accelerating certain types of innovation (mostly in the mining phase) and the long-term potentially negative effect of reducing other types of innovation (mostly in the prospecting phase), especially for those technologies that are not closely related to the GPT.

Another major contributor to the conundrum relates to global trends that have emerged over the past decade. Recent economic theory suggests that long-term economic growth is a result of the scientific and technical human capital involved in knowledge generation. In this regard, trends over the past decade suggest that by 2030 Asia will have a scientific and technical workforce that may be as much as five times the size of the U.S. scientific and technical workforce. It will be very difficult for the United States to take steps that will allow it to match the current Asia rates for the production of scientists and engineers. The global trends also suggest that Asia could surpass the United States in GDP by 2030.

It is reasonable to assume that Asia will use its new wealth and scientific and technical human capital to create a world class R&D infrastructure probably modeled after the highly successful U.S. R&D infrastructure. This could position Asia to become the global center for knowledge creation by the middle of the 21^{st} Century. If, in the long term, knowledge production is truly the foundation of economic growth and military power, then these global trends confront the United States with profound challenges. The U.S. needs to move quickly and correctly to meet these challenges.

Market forces and competition within the United States have been very effective at focusing scientific and engineering work so as to exploit scientific and technical discoveries, inventions and innovations. This has been responsible for much of the U.S. economic growth over the past several decades. However, this focus on exploitation has led to concerns regarding the generation of knowledge required for continued economic growth in the longer term, when the current crop of technologies have run their courses.

These concerns are exacerbated by the realization that globalization and demographics will likely "flatten" the world over the next 25 years in terms of economic competitiveness. A conceivable consequence is that the United States will move from the position of "Chairman of the Board" for the global economy to a partner in the global economy.

The study suggests that the United States revisit its governance for R&D investment and construct a new governance that recognizes the mutual dependences but distinctly different natures of the prospecting phase of R&D innovation and the mining phase of R&D innovation. Different governance is required for each of these phases.

The private sector is very effective at optimizing the short term R&D investment in the mining phase in situations where market forces dominate. The private sector should therefore be responsible for this governance. However, the very forces that make the private sector so effective at the governance of mining phase R&D also make it ineffective for the governance of the long-term R&D associated with the prospecting phase where profitability is very uncertain and can only be measured in hindsight after many years of sustained investment.

The proper role for the government in R&D is to ensure the health of the prospecting phase R&D (basic and applied research and exploratory development) that is crucial for long-term economic growth and military power but is not going to get done by the private sector. This role is so important to the long term economic and military health of the nation that the government must be staffed with the world class scientists and engineers needed to carry out this responsibility. This responsibility cannot be carried out by functionaries or administrators whose jobs are simply to send public moneys to non-governmental entities. We have chosen the term "governance" quite deliberately in this regard. Exercising the government's responsibilities requires that these responsibilities be carried out by government employees who are card-carrying members in the scientific and technical communities deemed to be of long-term importance for the nation's future. They must have the respect from these communities that is only earned by peers. The communities must accept the government as scientific and technical peers in order for the required long-term planning and steadfast direction to occur and so that the required advocacy is in place both within and outside of the government. At one time, the federal government was staffed to carry out this function. It is not clear that this is true today, especially in the DoD sector. This deficiency must be remedied. Excuses for not dealing with this matter—such as asserting that the government cannot hire or retain the required talent—should not be tolerated. If that is a problem then it should be fixed. It is not an overstatement to say that the Nation's long-term fate may be at stake.

A special situation exists for Defense R&D, where the beneficial effects of the free market do not apply due to the small market size and the specialized nature of warfare. In this case the United States Constitution assigns implicitly governance for the full spectrum (prospecting phase and mining phase) of R&D to the federal government. The same principles articulated above apply here.

It seems clear that the rate of new scientific and technical knowledge generation is related to the number of scientists and engineers who are working on new knowledge production. Global demographics suggest that this could result in the shifting of the center for new scientific and technical knowledge generation from the West to Asia over the next 25 years. If, in the long term, knowledge determines economic growth and military power then such a shift has profound implications for U.S strategy. In this regard, it would be in the long term interests of the United States to advocate a position that knowledge generation should be viewed as a public good available to all countries so that all countries, including the United States, can benefit from the global production of scientific and technical knowledge. This must be done within reasonable constraints associated with intellectual property and national security considerations. The knowledge exploitation phase is where controls will be most effective rather than the knowledge generation phase. Furthermore, the United States should maximize its production of scientists and engineers so that it can more easily interact with the global S&T community. This will require that the U.S. S&T communities continue to be viewed as leading players in science and technology, otherwise they will not have entry into the larger community. Finally, the United States should move to attract as many foreign scientists and engineers as possible to become U.S. citizens or permanent residents so as to maximize the U.S. ownership of the global scientific and technical community and, thereby, of its knowledge generation. Emigrant scientists and engineers have historically made huge contributions to U.S. economic and military strength.

I. Introduction

In some ways, it is the best of times and the worst of times for science and technology (S&T) in the United States. New technologies appear to be arriving at an ever more rapid pace. Advances in electronics, computers, communications, medicine and biology have transformed the way we live, play, conduct business and conduct military operations. New opportunities in many areas such as nanotechnology, autonomous systems, medicine and agriculture hold great promise for future progress. At the same time, however, there is a foreboding that all is not well. For example, in a February 2005 report titled "The Knowledge Economy: is the United States losing its competitive edge?" the Task Force on the Future of American Innovation

> "developed a set of benchmarks to assess the international standing of the U.S. in science and technology. These benchmarks in education, the science and engineering workforce, scientific knowledge, innovation, investment and high-tech economic output reveal troubling trends across the research and development spectrum. The U.S. still leads the world in research and discovery, but our advantage is rapidly eroding, and our global competitors may soon overtake us" {Task Force on the Future of American Innovation, 2005}.

This dichotomy—technologies are arriving at an ever more rapid rate (all is rosy), but they may be doing so at the expense of future potential (all is gloomy)—is what we here term the S&T innovation conundrum. The conundrum has given rise to a situation where one can hardly make it through a day without encountering several individuals, articles, or speeches invoking innovation as the key to resolving the Nation's ills. A simple web search on the word "innovation" results in fourteen million hits. In April 2004, the President announced a program for a "New Generation of American Innovation." In December 2004 the Council on Competitiveness convened the National Innovation Summit in Washington, DC. May 2005 saw the convening of the National Academy of Engineering symposium: "Stimulating Invention and Innovation." Numerous volumes are written on the subject; conferences are regularly held on the subject; and, a number of University Centers for Innovation have been founded. There has emerged an entire industry that makes its living promising to teach the non-innovative how to become productive innovators. This focus on innovation is rooted in such concerns as: the nation may be becoming less competitive in the international commercial marketplace; manufacturing jobs are being lost; high-technology jobs are being outsourced to overseas suppliers; and, the United States has moved from the world's largest creditor nation to the largest debtor nation in all areas including high-tech. These syndromes are often captured under the rubric of "globalization." As a Nation, we apparently are going to resolve these concerns through innovation. The apparent hope is that we will out-innovate the rest of the world by unleashing the innovative might of our free market economy.

This study attempts to move beyond the rhetoric associated with innovation in order to place the current situation within the context of the history that led to it. In the study we interpret the word innovation in the broad sense as defined by {Merriam-Webster, 1999}, i.e., "the introduction of something new." Innovation thus occurs in all phases of discovery, invention, experimentation, and development, although its nature may be different in these various phases. Current literature (e.g., {Branscomb, 2004}) suggests that innovation occurs when a product successfully makes it into the market place. This may be a suitable economic definition for a successful innovation. However, innovative work is done well before, and after, a product is in the marketplace. For example, the invention of the atomic clock was a scientific innovation that changed traditional timekeeping. The miniaturization of and introduction of the atomic clock into a space network to create the global positioning system (GPS) was also an innovation, but of a different nature. Each accomplishment was an important innovation, and for clarity of evaluation, it is important that each be referred to as such.

Society is primarily interested in the creation of functional capability as compared with the development of any specific innovation (for example, awareness of location rather than specifics of GPS). Increased functional capability is often a direct product of a technological accomplishment and often has profound economic implications for society. Hence, while the focus of this study is on S&T, economic considerations cannot be avoided. For the purpose of this study, S&T is considered to be a mix of basic research, applied research, and exploratory development.

The study begins with a brief examination of the origins of several important innovations of the past 100 years that are considered to be breakthroughs. The objective here is to examine the events and timescales that preceded the introduction of each of these innovations into military or product development programs. The examination identified certain enduring characteristics that seem to typify this early phase of scientific and technical innovation. These enduring characteristics suggest a simple conceptual framework within which to discuss the changing nature of the S&T associated with an emerging innovation. Of especial importance is the identification of two distinctly different and generally sequential phases of the S&T associated with breakthrough innovations. The concept of R&D innovation space is introduced and a few simple models are discussed as a means to illustrate the differences and also to point out the mutual dependencies of the two distinct phases. The study then proceeds to discuss certain insights that have become available from recent economic research regarding the role of technology in generating economic growth. The implication of these insights for the S&T innovation conundrum is discussed.

The study suggests that the current course the United States has embraced, combined with global developments and global demographics, does not bode well for the long term. A serious reassessment is needed with respect to the governance of R&D investment, the development of the S&T human workforce, and global partnerships for the generation of new knowledge. Recommendations that derive from this study thus fall into two categories: governance of the Nation’s R&D investment, and knowledge generation and human capital. These recommendations are provided in Sec. VI of this report.

II. Major Phases of S&T Innovation

A methodology is needed in order to examine the conflicting U.S. S&T innovation perceptions—"all is rosy" vice "all is gloomy"—that are described above. The approach we employ is that of hindsight. When a technological innovation is complete, one can look back with full knowledge of what key scientific and technical developments had to come together to make the innovation possible, prior to which the innovation would not have been feasible. Numerous studies have been done over the years employing retrospective examinations for the purpose of identifying the role of science and engineering in the origins of significant innovations. Two of the most notable and controversial were DoD's Project Hindsight {Hindsight, 1969} and NSF's TRACES study {TRACES, 1968}. This paper's objective is much less ambitious, seeking only to gain some understanding of how the timeline for significant scientific and technical innovations may have changed over the years. The paper does not attempt to distinguish if a contribution should be classified as basic research, applied research, exploratory development, engineering development, etc. Indeed, in most cases examined there appeared to be a mix of all of these categories. It would seem that innovations can come from almost anywhere. By considering details of the early development of radar for the United States Navy, we illustrate our methodology, in Sec II*i*. We have chosen this example because: early radar ushered in the age of modern warfare; it provides a benchmark for subsequent 20th Century innovations; and, this development is well documented {Allison, 1981}. In Sec II*ii* following, we employ the methodology to examine how a number of other representative innovations occurred. All of these successful innovations share a range of significant attributes, among them, a chronological progression through two distinct, major phases. These phases, termed "prospecting" and "mining," are described in Sec II*iii*.

i. Methodology Illustration: Development of Early Radar

Radar required scientific and technical developments in the following areas: understanding transmitted and reflected electromagnetic signals in the atmosphere; power supplies; transmitters; antennas; receivers; synchronizers; and, displays. Table 2.1 provides a timeline for related S&T developments.

Table 2.1 Selected Radar S&T Developments Timeline.

Year	S&T Development	Areas Impacted
1873	Maxwell publishes treatise on electricity and magnetism	All
1887	Hertz proves experimentally the existence of EM waves	All
1900	Fessenden conceives amplitude modulation, heterodyne principal	Transmitters, Receivers
1904	Fleming invents vacuum diode	All
1906	DeForest invents vacuum tube	All
1918	Armstrong invents superhetrodyne radio circuit	Transmitters, Receivers
1919	Schottsky invents tetrode vacuum tube	Receivers
1926	Hull describes HF amplification with tetrodes	Receivers
1927	Busch introduces electron optics	Displays
1934	Mesny publishes paper on time constants in multistage amplifiers	Receivers
1935	RCA produces the "Acorn" vacuum tube	Receivers

There were many others, but this selection is adequate to this discussion since practical radar would not have happened when it did without the developments listed in Table 2.1. Consider the availability of each of the sciences and technologies at key event points when radar might have been suggested. We start in 1873, when Maxwell published his treatise on electromagnetic theory {Maxwell, 1892}.

1873: Maxwell's 1873 work was the first that could be considered to reduce electromagnetic wave propagation to engineering terms, although he had actually developed the theory about ten years earlier. At the time of this publication one could have conceived of a radar-like system. However, as Table 2.1 shows, none of the required technologies were available. Therefore, although radar might have been conceived it could not have been built.

1887: The next significant event occurred in 1887, when Hertz demonstrated experimentally the existence of the waves predicted by Maxwell; his results were first published in 1889 {Hertz, 1889}. It is evident that Hertz demonstrated reflected electromagnetic (EM) waves in his experiments, and there is some indication that he envisioned the application of his experimental discovery to a radar-like situation {Allison p. 53, 1981}. However, none of the technologies required to actually build a radar device were in place at this time. Hertz's work resulted in an international effort that led to the development of radio.

1901: In 1901, Marconi demonstrated long-distance radio wave propagation, a demonstration that launched a large industrial-based undertaking that resulted in radio becoming a viable commercial enterprise. Much of the technology that would make radar achievable ultimately was developed by this industry. In 1901, however, none of the technologies needed to make a viable radar system were in place. However, radar could have been fully postulated at that time.

1904: Indeed, in 1904 Huelsmeyer {Hollman, 2001} patented and demonstrated a radar-like device for the detection of ships at sea and in harbors. This device was based on Hertz's experiments. Huelsmeyer intended his invention for the shipping industry. The device successfully detected ships but was not picked up by the commercial industry because it did not have the directionality and sensitivity needed to add value to the shipping business. In retrospect it is quite clear why this was the case, as can be seen in Table 2.1; the required technologies were simply not available in 1904.

1922: The next milestone in the development of early radar occurred in 1922, when Taylor (who has been given the appellation "Father of Radar" {Howeth, 1963}) and Young observed channel-fading in a communication experiment that they were conducting across the Potomac River {Allison p. 39, 1981}. The fading was observed when a wooden ship passed through the communication channel. This is the point where the development of U.S. Navy radar began. Shortly after this observation, Taylor sent a correspondence {Allison p. 40, 1981} to the Navy's Bureau of Engineering that stated: "with suitable parabolic reflectors at transmitter and receiver, using a concentrated instead of a diffused beam, the passage of vessels, particularly of steel vessels (warships) could be noted at much greater distances." He recommended that the Bureau support an effort to investigate this phenomenon. The Bureau never responded to the memorandum, but the Naval Research Laboratory (NRL), to which Taylor's organization was transferred in 1923, allowed a low-level investigation to continue. As can be seen from Table 2.1, at the time of this 1922 observation the technologies developed in support of the radio program were such that transmitters, receivers, synchronizers and displays had not reached the point where a viable radar could have been produced.

1930: Another major milestone occurred when Young, while conducting communication experiments at NRL in 1930, observed echo signals from aircraft landing at neighboring Bolling Field {Allison p. 61, 1981}. The Bureau of Engineering was informed of the observation and once again informed of the merit of initiating an intensive investigation to exploit the phenomena. There is no record that the Bureau responded to this memorandum. In 1931, the Commanding Officer of NRL then corresponded directly with the Secretary of the Navy. His memorandum included the statement: "In the detection of airplanes and probably ships by radio, although this was found feasible over a year ago, it has been impossible to secure Bureau support for the development of this vitally important problem by reason of the fact that its military value will find more ready understanding and appreciation from higher command afloat or from a broad conception of national defense than in a crowded bureau schedule where available funds for development and equipment are already over obligated and primary bureau emphasis is placed on radio as a means of communication" {Allison pp. 67-68, 1981}. The CO's memorandum resulted in the Bureau of Engineering authorizing a low priority radar project.

1934: In 1934, Robert Page demonstrated experimentally that a pulsed echo system is possible {Allison pp. 78-83, 1981}. By this milestone, the required technologies had reached a level of development—with the exception of the receivers—where a viable radar could be built. Shortly thereafter the remaining pieces fell into place. In 1935, RCA produced the "Acorn" pentode, which Page obtained promptly. That year Page also read a recently published article by Rene Mesny {Mesny, 1934}, results from which allowed him to resolve the remaining receiver sensitivity problem.

1936: The next milestone occurred in 1936, when Page demonstrated a functional pulsed-radar to the Bureau of Engineering {Allison pp. 85-97, 1981}. This was also the year that ADM Harold Bowen became the Commander of the Bureau of Engineering; he took a personal interest in the radar project, assigning it the highest priority and the highest classification level {Allison p. 96, 1981}.

1939: By 1938, the Bureau of Engineering staff had agreed that radar was important and pushed for rapid prototyping, an emphasis that resulted in a complete ship-based radar for the U.S. Navy in 1939. On May 8, 1939 a conference was held in the Office of the Chief of Naval Operations from which the following statement was issued {Allison p. 110, 1981}: "On a motion concurred with by all representatives, it was agreed to recommend that procurement of from 10 to 20 of the radar devices in their present form, with only minor and readily accomplished changes, be undertaken at once, for installation and service trial on vessels of the fleet; this procurement is not to interfere with concurrent development. Immediate procurement was considered imperative because (a) the device is of great military value in its present form; (b) the experience in the service will permit exploration of its capabilities and limitations, will provide training in its use, and will point the way for further development; (c) the international situation requires that immediate advantage be taken of every device leading to greater military effectiveness; and, (d) there is no positive guarantee that development of the improved device will be successful."

The consequence was that U.S. combatants were equipped with search radars in time for World War II.

Table 2.2 represents an attempt to summarize the readiness of the required technologies at the various times in the development discussed above, as correlated with the researcher most centrally involved. It can be seen from this Table that while the concept of radar could have been put forward as early as 1873, the technologies to make it viable (most of which resulted from the commercial radio program) were not available until 1936. Once they became available, radar moved very rapidly into production. These early radars were meter wavelength systems because that was what the technology would support at the time. They did however demonstrate all the functionality required for radar, and enabled the rapid introduction of microwave radar in the 1940's, when microwave power tube technology became available to the United States.

Table 2.2. Radar Technology Readiness.

 Radar on U.S.S. New York, 1939	MAXWELL	HERTZ	MARCONI	HUELSMEYER	TAYLOR	YOUNG	PAGE	PAGE
DATE	**1873**	**1887**	**1901**	**1904**	**1922**	**1930**	**1934**	**1936**
Electromagnetic Waves (Theory)	YES	YES	YES	YES	YES	YES	YES	YES
Electromagnetic Waves (Experiment)	NO	YES	YES	YES	YES	YES	YES	YES
Reflected Signals	NO	YES	YES	YES	YES	YES	YES	YES
Power Supplies	NO	NO	NO	NO	YES	YES	YES	YES
Transmitters	NO	NO	NO	NO	YES	YES	YES	YES
Antennas	NO	NO	NO	NO	YES	YES	YES	YES
Receivers	NO	NO	NO	NO	NO	NO	NO	YES
Synchronizers	NO	NO	NO	NO	NO	NO	YES	YES
Displays	NO	NO	NO	NO	NO	NO	YES	YES

This brief summary of developments leading up to early radar is instructive regarding the roles played by individuals, organizations, and commercial technology in the realization of a radically new military capability. The early players (e.g., Maxwell and Hertz) sowed the seeds even though their work was in no way motivated by radar. Commercial industry provided much of the technology that would be required, but its motivation had nothing to do with radar. The invention was serendipitous, but its importance was recognized because the inventors (Taylor, Young, Page) understood the underlying technology and the importance of the invention to the customer (the U.S. Navy). There was great difficulty in selling the importance of the invention to the customer that had its resources committed to accepted technology (communication systems). An organization (NRL) provided cover that allowed the inventors to develop their invention. The inventors configured commercial off the shelf technology (vacuum tubes, oscilloscopes, etc) to produce an entirely new technology. A well-placed sponsor (ADM Bowen) ultimately recognized the importance of the invention and saw to it that resources necessary to make a practical device were provided.

The above sequence of events remains typical of the introduction of major technical innovations. This fact has significant implications for DoD and others regarding organization and staffing.

ii. Representative Technologies

The previous section, by example, laid out a methodology for looking at how innovation occurs based on researcher motivation and technology readiness. Examination of the history of radar suggests that there are two distinct major phases. The first, an early searching phase, is characterized by a number of isolated developments, each of which has a huge impact and each of which is generally closely linked with an individual scientist or engineer who often is not associated with the final innovation. This phase culminates with one or a few technical groups pulling together the requisite disparate developments with, as available, commercial off the shelf (COTS) capabilities. A working innovation prototype is produced that has all of the essential attributes of the final capability and is recognized by individuals with the ability to significantly advance the innovation that is ready for production. This phase is then followed by a more predictable latter phase that is characterized by disciplined technical activity with a large number of innovations, each having smaller impact than those in the previous phase but nevertheless with cumulative impact that can be large.

With the purpose of illustrating the generality of the above-sketched life cycle of major innovation, in this section we delineate briefly the histories of six specific, diverse innovations: solid-state digital electronics; cellular telephone systems; the global positioning system (GPS); DNA fingerprinting; compact disk (CD) and digital versatile disk (DVD) technology; and, giant magneto-resistance (GMR) computer memory read heads. This representation of innovations extends over the past century, and provides examples from across several disciplines. In each case, generally isolated developments—each with clearly identified individuals responsible—are hallmarks of the early phase. It is noteworthy, however, that in many cases the discoverers of early key technologies did not envision the innovation that ultimately resulted from their contributions. This early phase generally culminates with influential recognition of the importance of the innovation, and is followed by a latter phase of more predictable and planned R&D. Our principal interest in these brief histories is to establish the beginning and end points of the early searching phase, rather than to describe completely the intervening events.

a. Solid-state Digital Electronics

Arguably, the most sustained and pervasive technical innovation in recent history is that of solid-state digital electronics. Every technical researcher alive today has spent the majority of his or her career under the aegis of this progressing innovation. Solid-state digital electronics underpins virtually all of the innovations to which the latter half of this century can lay claim. The following brief summary is based largely on the work of Riordan and Hoddeson {Riordan & Hoddeson, 1997}.

Solid-state digital electronics is one of the many products that resulted from the discovery and development of quantum mechanics. By 1932, the quantum theory of solids had emerged and physicists were educated in the practice of this science. A key event occurred in 1939, when Russell Ohl of AT&T Bell Laboratories made a discovery that would contribute to changing the world. Ohl was fixated on improving the crystal oscillator. He believed that producing ultra-pure silicon would take him where he wanted to go. At that time AT&T management was focused on improving vacuum tubes that were critical to their switching circuits. They wanted Ohl to drop his silicon work and join the quest for improved vacuum tubes. Apparently, at this juncture Ohl's supervisor ran interference for him and allowed him to continue his silicon work. In 1939, Ohl's efforts to purify silicon produced by accident a cracked crystal that showed anomalous properties regarding electronic conduction. Realizing that this might be important, Ohl asked some members of Bell Lab's technical staff to take a look at what he had found. Among those who reviewed the experimental results was Walter Brattain, who believed he understood what was responsible for the anomalous behavior. What Ohl had discovered was the "*pn* junction," a phenomenon that would be key to the development of the transistor. In 1947, Brattain and colleagues at Bell Labs invented the transistor, for which they were subsequently recognized with the Nobel Prize. The discrete transistor provided considerable new capability. However, it was the subsequent investigations that resulted in the concept of the monolithic integrated circuit that led to the revolution in digital solid-state electronics. For the purpose of this study, the discrete transistor is considered to be an event in the early searching phase of digital solid-state electronics. In 1958, Kilby and Noyce each independently invented the integrated circuit. At this point one could argue that the early phase of innovation for solid-state digital electronics was complete.

By 1965, Moore was able to articulate his now famous "law" defining feature size reduction rates {Moore, 1965}. Solid-state digital electronics was well into its second phase, one of disciplined, contiguous activity. This phase has been enormously productive but may be coming to maturity (see, e.g., {Zhirnov et al, 2003}, {Borsuk & Coffey, 2003}).

b. Cellular Telephone Systems

An innovation that follows directly from the broad understanding of signals transmission developed during World War II is today's ubiquitous communications tool, the cell phone. While today's personal hand held cell telephones are made practical by the inventions in solid-state digital electronics, many of the key elements of the central innovation—the cellular telephone system—are in the areas of signal processing, e.g., how to use communication signals to enable large-scale roaming with connectivity of acceptable reliability.

A cell phone system is comprised of a network of small geographical areas called cells. Communication frequencies are simultaneously reused by different non-adjacent cells and are automatically exchanged for (or handed off to) new frequencies, as a mobile user roams from cell to cell. The seminal cell phone system concept—frequency reuse in many small cells, with consideration of the issue of handoff—was described in a 1947 Bell Laboratories Technical Memorandum by D. H. Ring (see {Roessner et al., 1998}). In 1960, an entire cellular system concept was formally described by three Bell Labs researchers, Lewis, Schulte and Cornell {Lewis, 1960; Schulte & Cornell, 1960}. Soon after, the necessary understanding about fading communications was developed by Bello {Bello, 1963}, based upon the 1950's MIT's Lincoln Laboratories work on the topic of stochastic radio astronomy signals with multiple dependencies. In July 1969, Bell operated the first commercial cellular radio system aboard the New York to DC Metroliner. This system featured reuse of six channels in the 450 MHz band in nine zones along 225 miles of track, and was managed by computerized control located in Philadelphia. With this demonstration, the early phase of innovation for cellular telephone systems can be viewed as being complete.

In 1970, Amos E. Joel, Jr. and Bell Telephone Laboratories filed the first patent for a cellular frequency-reuse mobile communication system; it was approved in May 1972 (No. 3,663,762). In 1971, Motorola demonstrated the first hand-held cell phone. Martin Cooper and others at Motorola filed a patent titled "Radio Telephone System" in 1973, which was awarded in September 1975 (No. 3,906,166). The FCC permitted Bell to begin a trial commercial cellular system in the Chicago area, in May 1974 ({privateline}; and *cf* following). In October 1983 the regional Bell Telephone Company operating in the Chicago area began the first U.S. commercial cellular service, soon to be followed by the Motorola Dyna-TAC service in Baltimore, in December 1983. These systems were preceded in commercial operation by a system in Bahrain (in 1978); by the Japanese (in 1979); by the Nordic Mobile Telephone System in Denmark, Finland, Sweden and Norway (in 1981); and, by Canadian commercial cellular service (in February 1983). To enable more bandwidth, digital dual mode (Interim Standard (IS) - 54B) was first implemented in the North America cellular network in March 1990. In 1994, Qaalcom, Inc. proposed a cellular system and standard based on ideas of spread spectrum communications, IS-95, to address demands of ever-increasing capacity.

Deployment of cell phone systems was ultimately made practical by the enabling development of the inexpensive, lightweight, low-power solid-state digital device: the personal hand-held cell telephone. Although the functional development of cellular phone systems approached technical maturity by the mid-1990's, these devices are continuing to undergo rapid and commercially popular evolution with the introduction of additional functionality (e.g., incorporation of: text messaging; video imagery; interactive games; etc.).

c. Global Positioning System

The Global Positioning System is a space-based innovation with enormous impact. This innovation is rooted in the study of navigation and tracking for military applications. Its viability derives from two disparate and seemingly unrelated technologies: satellites, and precision clocks. The following brief summary is based largely on a paper by Parkinson et al. {Parkinson et al., 1995}.

The time difference of arrival approach for navigation purposes appears to have first been used by the British in the GEE system that was deployed in 1940 to guide British bombers as they traveled over Europe. This was a two-dimensional system developed with advances in radar and timing technology. Another event that proved important for GPS was the suggestion by Rabi in 1945 that molecular beam techniques could be used to develop very precise clocks. His interest had nothing to do with navigation but rather was motivated by a desire to examine the validity of Einstein's general theory of relativity. Other seminal work was performed by Townes on MASERS, and by Ramsey on cesium beam clocks {Ramsey, 1989}. Groundwork for DoD involvement began in the 1950's when NRL, in anticipation of the emergence of artificial satellites, developed the tracking system called MINITRACK, and deployed it in 1957. In 1958, Sputnik was launched (and tracked within hours of its launch by the NRL MINITRACK). Sputnik demonstrated that artificial satellites could be placed in Earth orbit. Space technology contributions that were key to the ultimate development of GPS were made during the TRANSIT program that was launched in 1959, and developed by Johns Hopkins University APL. Further key development occurred in 1961, when NRL added a ranging signal to MINITRACK, and the system became the Naval Space Surveillance System, NAVSPASUR. These early programs laid the groundwork that was necessary to begin the development of GPS as it is defined today.

The possibility of operational GPS came into focus with the establishment of the Air Force Project 621 B (in 1962) and the NRL TIMATION program (in 1963). Beginning with the TIMATION I satellite in 1967, NRL launched several spacecraft that were aimed at examining the various technologies necessary for navigation satellites. In 1973, as a consequence of individual NRL and Air Force program successes, the Navstar GPS Program was initiated. This Program was a merger of NRL's research program with that of the Air Force. Its stand-up signaled the beginning of the latter phase of innovation for GPS. The 1974 TIMATION satellite, third of the NRL series, was equipped with the first space-qualified atomic clock, and was designated Navigation Technology Satellite (NTS) 1. NTS 2, launched in 1977, carried the first space cesium clock and tested essentially all GPS functions. It was the first satellite of the Navstar GPS Constellation.

For the development of the functional navigational capability of GPS, serendipitous events such as Rabi's suggestion for the atomic clock proved to be key even though it could never have been predicted to be so at the time it was suggested. The early and continuing presence of individuals such as Roger Easton of NRL and Ivan Getting of Aerospace Corp., who recognized the importance of various developments to the ultimate navigation system, were critical to its success. Actual demonstration of key technologies in the operating environment also proved to be essential. The transition of the program from its progenitors to a program that was focused on the use of the technology rather than on its development was, finally, necessary in order to move the capability beyond that of demonstration prototype.

d. DNA Fingerprinting

DNA fingerprinting is a now widely accepted capability to identify a person exclusively by means of unique patterns in his or her DNA. It was discovered by Sir Alec Jeffreys in a moment of "eureka," on September 10, 1984 (e.g., {Crace, 2004}; {Australia Prize, 1998).

Several important and necessary events preceded Jeffreys' discovery. They begin in the mid-1960's with the breaking of the program of the cell, the genetic code, for which Robert W. Holley, H. Gobind Khorana, and Marshall W. Nirenberg received the Nobel Prize in Physiology or Medicine in 1968 {nobelprize}. Being able to decipher the genetic code, which specifies universal relations between DNA structure and protein structure, provided the basis from which individual differences in DNA structure could then be evaluated. A decade later, Southern {Southern, 1975} described restriction fragment length polymorphism (RFLP), in which restriction enzymes cut DNA at specific sequences. By 1977, Sanger had solved the problem of DNA sequencing {Sanger et al., 1977}. In 1980, Wyman and White {Wyman and White, 1980} reported observation of one of the first polymorphic regions of DNA. Jeffreys used RFLP to investigate polymorphic regions that consist of tandem repeat DNA, or minisatellites, where short sequences are repeated many times in a row {Jeffreys et al., 1985}. In a separate research program in the United States, Kary Mullis was studying ways to amplify DNA patterns. His research resulted in the invention of the polymerase chain reaction (PCR) technique for making multiple copies of DNA in the laboratory {Saiki et al., 1985}, for which he received the Nobel Prize in Chemistry in 1993.

Meanwhile, Jeffreys' investigation of minisatellite regions was bearing fruit. By examining a large number of minisatellites, he and his colleagues identified a common short "core" sequence—a piece of DNA that is the same in many minisatellites—in each repeat unit. Upon comparing DNA core sequences from a number of people, Jeffreys realized that he was looking at readily identifiable patterns. He also realized, in a moment of epiphany, that the patterns are different for each person, *i.e.* that each individual must possess a unique such pattern or DNA fingerprint. This discovery marked the culmination of the early phase of innovation for DNA fingerprinting.

Within months of Jeffrey's 1984 observation, DNA fingerprinting was formally used to establish the family relationship of a child in an immigration dispute, and to identify an unsuspected murderer. At this time, the method for determining a suspect's DNA fingerprint was possible, but was far from clinically routine. When Mullis' PCR technique became sufficiently robust {Saiki et al., 1988}, it was immediately applied to the procedure for DNA fingerprinting. PCR, which enables amplification of specific regions of DNA, significantly facilitates DNA fingerprint pattern identification in a production environment. By the mid 1990's, the DNA fingerprinting technique had gained wide technical and operational acceptance as a functional capability, leading to the 1994 U.S. DNA Identification Act, and to the 1998 FBI Laboratory establishment of the National DNA Index System that enables the sharing of DNA profiles across jurisdictions for purposes of forensic identification.

e. Compact Disc and Digital Versatile Disc Technology

Memory storage discs (CDs and DVDs) are a recent innovation in the functional capability area of data storage media; CDs were commercially released in 1982, and DVDs, a closely related sequential development, were available by 1997. In January 2005, memory storage discs were chosen by a panel for the Lemelson-MIT program to be one of the top twenty-five innovations of the past quarter century {Makofske, 2005}. This widely used innovation relies on a mix of hardware and theory for its existence. The following brief summary is based largely on an account by Pohlmann {Pohlmann, 1992}.

Physically, a CD is a 12 cm diameter injection-molded circular polycarbonate substrate over-coated with a reflective metallized layer (usually gold) that is spin-coated with acrylic lacquer for protection. Commercially available CDs can store 700 MB of data, and current DVDs can hold about 7x that amount, or 4.7 GB. CDs and DVDs are read/write accessible by means of red laser light.

The ability to store the quantities of data that make CDs a viable technology derives essentially from the multiple error-correction code that I. S. Reed and G. Solomon published in 1960 {Reed & Solomon, 1960}. CD hardware was first described in 1969 by researchers Klass Compaan, Piet Kramer, and co-workers, at Philips {Philips}; a glass prototype design was completed in 1970. In 1978, Polygram, a division of Phillips, determined polycarbonate to be a good substrate material for the CD. Phillips and Sony agreed to work together on this technology in 1979, a collaboration that lasted through 1981. This collaboration ushered in the beginning of the latter phases of innovation for memory storage disks.

In 1980, Phillips and Sony jointly proposed a CD standard, with Reed-Solomon code as the encoding format. The two corporations soon had products ready for the commercial market, for an autumn 1982 release in Europe and Japan, and spring 1983 release in the United States. The DVD (originally termed “Digital Video Disc”), a follow-on technology to the CD, was introduced in 1996 and released to consumers in 1997.

f. Giant Magneto-Resistance Computer Memory Read Heads

Our final example of innovation is that of giant magneto-resistance. The effect of magneto-resistance, first reported by Lord Kelvin in 1857 {computinghistorymuseum}, is the alteration of electrical properties of ferromagnetic material such as iron when placed in a magnetic field. Giant magneto-resistance, a quantum mechanical magneto-resistance effect that is observed in structures composed of very thin layers of alternating ferromagnetic and non-magnetic metals, was simultaneously discovered by Albert Fert, then at the Laboratorie de Physique des Solides at Orsay {Baibich et al., 1988}, reporting a resistance change of 50%, and by Peter Grunberg in Julich, Germany {Binasch et al., 1989}, reporting a resistance change of 6%. The effect was quickly incorporated into commercial sector products. By 1994, GMR magnetic field sensors became commercially available, and by the end of 1997, IBM had ready for release the first GMR read heads for magnetic hard disk drives. With their new high sensitivity GMR heads, IBM achieved a much-publicized, self-imposed goal, the storage of 1 to 2 Gbits of data per square inch of magnetic disk space. GMR heads are now the data-storage standard of the computer industry. IBM estimates that further developments with GMR technology will enable the storage of data at densities exceeding 100 Gbits per square inch of magnetic storage platter space {computinghistorymusuem}.

The history of GMR read heads follows two tracks, that of the GMR effect, and that of the technology of computer memory read heads. Commenting on the remarkably rapid incorporation of the fundamental GMR effect into a digital storage product that is used world-wide on a daily basis {cnrs, 2003}, Fert noted in 2003 that "the first teaching which I take away from this adventure is that technological advances generally stem from fundamental research performed long ago. Giant magneto-resistance and spin electronics did not arise spontaneously in 1988."

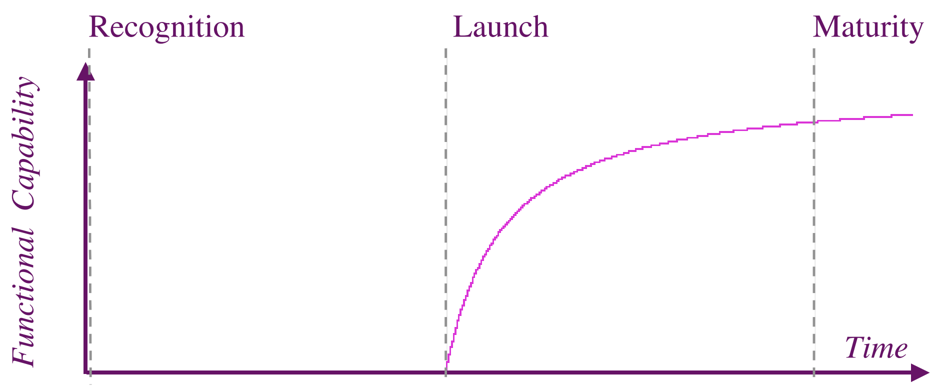
Operationally, the path to the 1988 discovery of GMR by Fert and Grunberg began in 1975, with the discovery of tunneling magneto-resistance (TMR) by Julliere {Julliere, 1975}. The TMR effect, which occurs from the same sort of physics as the GMR effect, is observed with a thin-layer configuration that is very similar to that required for GMR except that the spacer layer between the two ferromagnetic layers is an insulator, not a non-magnetic metal as is needed for GMR. Prinz, in 1981, first applied molecular beam epitaxy (MBE) to the area of thin metallic film research, work that Fert noted "inspired .. [his decision to] work on magnetically layered structures" using the Prinz methodology, and to then discover the GMR effect {Fert, 2000}. In the spring of 1988, Stuart Parkin attended a scientific meeting in France, where he heard Fert report about his observation of unexpected large resistance changes in his magnetic multi-layer structures. When Parkin returned home he set himself to reproducing the result using a mass-production technique, sputtering, instead of the time and labor intensive MBE {Economist, 2001}. With no believable theoretical predictions to go on, he and his group at IBM empirically tried 30,000 multilayer combinations of elements and non-magnetic spacer layer thicknesses. Eventually they achieved success, developing a multi-layered structure that produced significant changes in resistance in response to relatively small magnetic fields, and that operated at room temperatures; they had developed the "spin valve," which was patented in 1992. With this invention, the early phase of GMR effect innovation was complete.

Meanwhile, on a separate trajectory of inquiry, IBM and other companies had been researching the area of magnetic memory read and write head technologies, to develop ways to pack more information on disks (e.g., with the IBM development of thin film induction read heads in 1979 {Grunberg, 2001}) while at the same time avoiding the Super Paramagnetic Effect (SPE) barrier. In 1992, IBM introduced read-heads based on the anisotropic magneto-resistance effect (MR), which were successful and contributed to an annual storage capacity growth rate of about 60%. The subsequent rapidity of the commercial transition to GMR heads in 1997 was made possible by these previous and long-ongoing thin-film and MR head developments: very little in the surrounding disk drive components needed to be re-engineered for the practical incorporation of the eclipsing new GMR read head technology.

iii. Phases of S&T Innovation

While each of the above-described innovations is technologically unique, essential similarities can be found among all of the case histories regarding how the innovations came about. In this section we consider the time scales, and associated major phases, for each of the innovations discussed in Secs. II *i* and *ii* above.

Figure 2.1 displays a map of the time histories we attribute to the early searching and later more predictable phases for each case. Times spent in each phase are also indicated.



Early Radar	1922	14 years	1936	14 years	1950
S-S D Electronics	1939	19 years	1958	57 years	2015†
Cell Phone System	1945	24 years	1969	25 years	1994
GPS	1957	16 years	1973	16 years	1989
DNA Fingerprinting	1966	19 years	1985	...	?
CD/ DVD	1960-1969	19 - 10 years	1979	20 years	1999
GMR read heads	1975	17 years	1992	18 years	2010‡

† *estimated; see, e.g., {Borsuk & Coffey, 2003}*; ‡ *estimated; see, e.g., {Prinz, 1995}*

Figure 2.1. Functional capability innovation time scales.

There is always some uncertainty in deciding a start date or an end date of an innovation. For example, in the case of early radar, one could go back to Maxwell's work in 1873. It is clear, however, that most of the work that immediately followed Maxwell should not be attributed to radar but rather to communications. It is also clear that it would be difficult to select a start date for radar that is later than the 1922 observation of Taylor and Young. Similarly it would be difficult to make the launch date for the mining phase of Navy radar much different from 1936 when Page fully demonstrated the concept experimentally. These two events bound the 14-year period between the entry to the early, searching phase and the entry to the later, production-focused phase. Early radar was eclipsed by microwave radar during the 1940's. As a result we arbitrarily assigned the date of 1950 for the maturation of the early U.S. Navy radar. That does not mean that innovation with meter wave radars stopped in 1950, but only that the particular technology we term "early radar" was depleted by 1950.

For solid-state digital electronics, it is difficult to assign the start date of the early searching phase to be any later than 1939 when Ohl discovered the *pn* junction. It would also be difficult to assign the start date for the latter phase to be much different from 1958, when the integrated circuit was invented. Based upon our current understanding of the situation with the silicon MOSFET, we are expecting a maturation date for solid-state digital electronics to be sometime around 2015 (*cf* { Borsuk & Coffey, 2003}; {ASCR 2005}). That results in a 57-year mining phase, which is quite extraordinary.

A comparable thought process was applied to the other innovations examined during this study. It is noteworthy that the early searching phase for all of the major innovations considered (covering a 100-year period) appears to be of the order of 15 to 20 years. These include the most recent technologies. This is interesting because one tends to think that today's technologies are arriving at an ever more rapid rate. However, the development time lines, examples from which we have presented in this study, do not support this general perception.

The examples discussed above indicate that there are at least two distinct phases in the history of a major S&T innovation. There is an early, searching phase that is evocative of prospecting. It is characterized by a few discrete but high impact events. There is little functional capability produced during this phase and the individuals contributing to the discrete events, while generally sure that they are involved in profoundly exciting research and technology, often have no idea what the ultimate functional capability will be. Continuing with the analogy, this early phase is followed by a later, more predictable phase that is much like mining. This latter phase seems to be dominated by continuous improvement in functional capability as characterized by a larger number of lower-impact innovations than occurred in the early phase. During the mining phase, the capability produced can usually be related to the funding applied and to the inherent potential of the technology being exploited, resulting in desirable features such as measurable and predictable return on investment.

It is important to understand that the prospecting phase involves a mix of basic research, applied research, and exploratory development. The prospecting phase is not simply a basic research undertaking. For example, the prospecting phase of GPS involved substantial basic research regarding atomic clocks, substantial applied research on such items as power sources, and substantial exploratory development work such as that associated with the NRL technology demonstration spacecraft. Similarly, the mining phase of digital solid-state electronics involved basic and applied research needed to resolve issues associated with staying on "Moore's Law" and also very large development undertakings needed to provide the fabrication facilities.

Figure 2.2 is an attempt to illustrate the thoughts presented in the preceding paragraph. This figure proposes two major phases in the development of functional capability: prospecting; and, mining. The first phase illustrated in Fig. 2.2 is described as the "prospecting" phase. This is followed by the oft referred-to "valley of death" (i.e., the funding gap that must be crossed in order to move from the prospecting phase to the mining phase), which in turn is followed by the second major phase, "mining." Finally, "depletion" indicates that the technology that is being exploited has delivered all that it can. Because the valley of death has been much talked about in the contemporary literature on innovation, we show it separately. However, in the experience of the authors, the valley of death and the prospecting phases overlap. The prospecting phase and the valley of death are both typically characterized by a small number of discrete events that have a very large impact, and are generally closely linked to an individual scientist, engineer, or entrepreneur. A successful prospecting phase culminates with a working innovation prototype that is recognized to be ready for production. Throughout the prospecting phase and the valley of death, it is difficult to find the resources necessary to bring innovations to the point that predictable development programs can be launched.

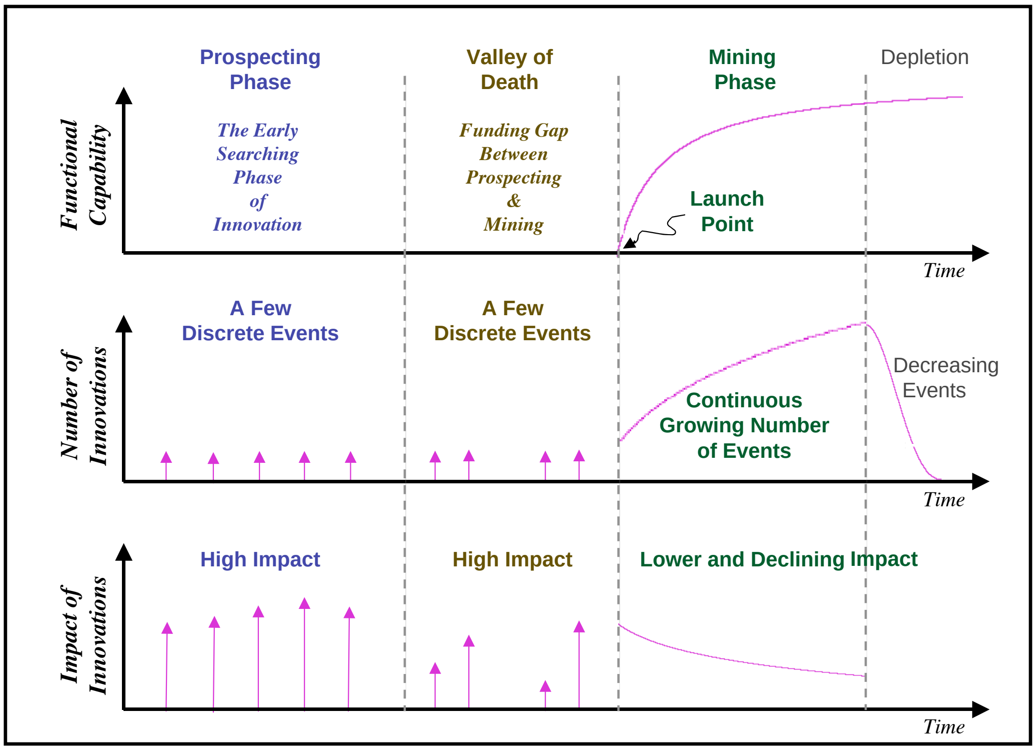


Figure 2.2. The landscape of innovation.

The mining phase, which follows successful prospecting, is a more predictable phase that is characterized by disciplined activity with a large number of innovations, each having smaller impact than those in the prospecting phase but nevertheless with cumulative impact that can be large. This phase is generally ushered in with non-trivial recognition that the innovation is important in a practical sense. Developments in this phase are underwritten with near-term business profit strategies in mind. Despite the generally lower impact of each individual mining phase innovation, however, the integrated effect of innovations during this phase can have a huge impact.

In the next section, we model the phases of innovation and the interplay among these phases, to come to a clearer understanding of the essential features of innovation genesis and exploitation.

III. Investigating S&T Innovation from Genesis through Exploitation

In this section, we will draw some inferences based upon the representative innovations discussed in the previous section, to come to a paradigm understanding of the evolution of innovation. We begin with an analytical model of the mining phase, followed by a possible way to heuristically investigate the prospecting phase.

i. A Simple Analytic Model of the Mining Phase

Once a mining phase is well-launched, the innovation process becomes generally quantifiable. This predictability in some cases can even lead to successful forward-directed "scheduling of innovation" such as Moore's Law, pointing to a profound economic aspect of the mining phase. It is straightforward to envision applying disciplined management techniques during this phase. For example, financial return on investment is directly related to the slope of the curve that describes the temporal evolution of functional capability in a mining phase, which allows one to estimate capability improvement for a given investment and relate that to profitability. The International Technology Roadmap for Semiconductors {itrs, 2004} is a paradigmatic example of management planning during a mining phase. Once the major innovation has been discovered—i.e. once the integrated circuit had been invented and a device (MOSFET) and material system (Si, SiO_2) were then settled on—industry was able to focus its resources on pursuing the program path embodied by Moore's Law. This required a great deal of innovation (for example, in lithography, metrology, surface chemistry, bonding, and so forth) but it was planned innovation to meet an agreed-to need, whereas in the prospecting phase for this innovation (1932 to 1958) it was not clear what the ultimate innovation would be. The innovation required in the mining phase was paced by the desire to stay on Moore's Law and resources were provided to do so. That investment paced the innovation that would be required during the mining phase.

For the purpose of illustration, we here develop an elementary model for the evolution of functional capability during the mining phase. We will consider that there are three essential variables during a mining phase: the desired functional capability C; the potential functional capability P; and, the funding F, with normalization that one unit of funding is required to create one unit of functional capability and in the process consumes one unit of potential.

Let us assume that the interaction among these variables, C, P, and F, can be made analogous to a simple chemical reaction, namely

$$P + F \rightarrow C \qquad (1)$$

implying that

$$dP/dt = dF/dt = f(P,F) = - dC/dt \qquad (2)$$

where f(P,F) is some function that describes the reaction between potential functional capability P and the funding F. It seems reasonable to assume that if there is no potential P then f(0,F) = 0, and if there is no funding F then f(P,0) = 0. A function that accomplishes this is

$$f(P,F) = -R(t)\, PF \qquad (3)$$

where R(t) is a rate coefficient that might, for example, be representative of the manufacturing technology of the time and the accumulated knowledge associated with the technology being exploited. Since

$$F - F_0 = P - P_0 \qquad (4)$$

where F_0 and P_0 are the initial values of the funding F and the potential functional capability P respectively, it is straightforward to show that

$$dC/dt = R(t)(F_0 - C)(P_0 - C). \qquad (5)$$

The initial functional capability P_0 should be viewed as being analogous to the amount of gold in a newly discovered gold mine. The discoverer may not know how much gold is in the mine but the mining process cannot extract more than is there at the time of discovery. The same is true for any specific technology: it has an intrinsic potential for functional capability and cannot produce more than that.

From Eq. (5), the stationary points for C are

$$C = P_0 \quad \text{for } F_0 > P_0 \qquad (6)$$

$$C = F_0 \quad \text{for } F_0 < P_0 \qquad (7)$$

This shows that functional capability will stop increasing when the full potential is reached, or when the money runs out. If the initial funding F_0 required to exploit fully a technology is too large to be supported by the relevant economy, the technology will not be exploited, or if it is exploited it will not deliver its full potential because adequate capital cannot be provided.

a. With Constant Rate Coefficient

For the special case where exactly enough money has been set aside to capture all of the potential functional capability (i.e., $F_0 = P_0$), and the rate coefficient R is a constant R_0, the functional capability becomes

$$C = C_0 + P_0\tau/(1 + \tau) \qquad (8)$$

where C_0 is the initial functional capability (which we will take as being zero in this paper) and $\tau = P_0 R_0 (t - t_0)$, for t_0 the time at which a new mining phase begins. This curve is plotted in Fig. 3.1.

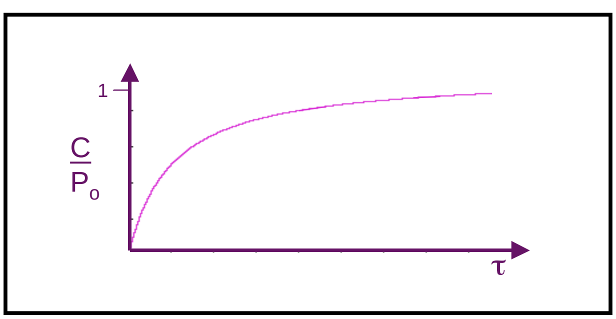


Figure 3.1. Simple mining phase model representation of functional capability as a function of time.

Figure 3.1 indicates that functional capability will increase at some rate determined by the initial potential functional capability P_0 and the rate coefficient R_0, and will asymptote when the initial potential functional capability is depleted. There are important innovations that develop according to the Eq. (5). These innovations go through the mining phase with the same process rate coefficient with which they entered it, with a mining phase that is usually short-lived.

b. With Feedback

In a more general situation, the process rate coefficient R increases as the functional capability increases (for example as a result of improved manufacturing technology, accumulated knowledge regarding the technology being exploited, etc). One can gain some insight into this situation by allowing the rate R to increase as the capability C increases. In this case, to first order in C

$$R(t) = R_0[\ 1 + \varepsilon C(t)/P_0\ ] \qquad (9)$$

The parameter ε quantifies the feedback introduced as a result of improvements in relevant innovation exploitation processes such as manufacturing technology, learning, and so forth. If Eq. (9) is substituted into Eq. (5), the resulting equation can be integrated straightforwardly. This integral is somewhat complicated and will not be presented here. However, when $P_0 = F_0$, $\varepsilon >> 1$, and $C(t) << P_0$, the expression for C(t) reduces to

$$C(t) = (P_0/\varepsilon)(\ \exp(\tau) - 1)\ , \qquad (10)$$

where

$$\tau\ =\ P_0R_0(\ 1 + \varepsilon\ )\ (t - t_0), \quad \text{for} \quad t > t_0. \qquad (11)$$

Equation (10) allows us to estimate the number of characteristic times (*e*-folding periods) that are permitted for a given improvement feedback factor ε. This is done by recalling that Eq. (10) is valid for $C(t) < P_0$, and that this limit will be violated when

$$\exp(\tau) =\ \varepsilon \qquad (12)$$

or

$$\tau_e = \ln(\varepsilon). \qquad (13)$$

The quantity τ_e is the number of allowed *e*-foldings of functional capability. Note that, for instance, 14 *e*-foldings requires an improvement feedback factor $\varepsilon = 10^6$. Equation (13) thus indicates that if a functional capability is to undergo many *e*-foldings, the rate coefficient R must be very strongly coupled to increases in functional capability (e.g., when an increase in functional capability from an innovation that is being mined leads directly to an increase in the size of the skilled labor pool that is available for mining the innovation further, or to application of new manufacturing technology).

Figure 3.2 presents the exact solution for Eq. (5) for several values of ε in the case where $F_0 = P_0$ and where R is specified by Eq. (9). The reader should note that in Fig. 3.2 the normalized time τ is defined by Eq. (11). Therefore, real time scales as $(1 + \varepsilon)^{-1}$ and hence functional capability develops more rapidly as ε increases. It should also be kept in mind that the required rate of expenditure of funds increases as ε increases (nothing is free).

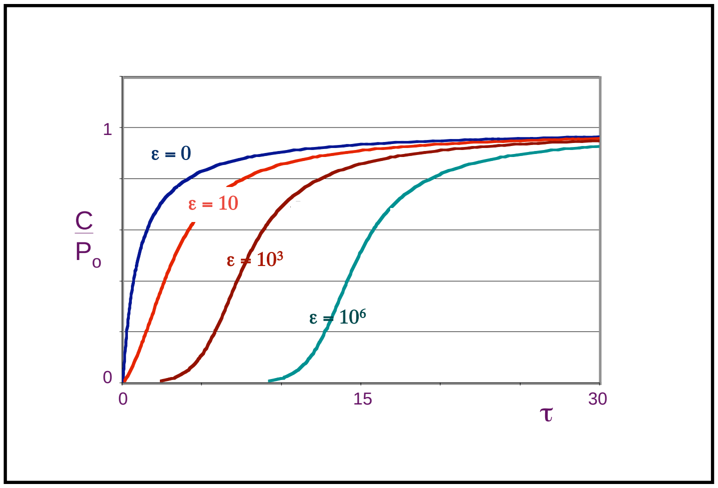


Figure 3.2. Normalized functional capability for a range of improvement feedback factors ε.

Note that not all significant functional capability developments are "large ε"□ processes. GPS, discussed in Sec. II*ii*c above, provided an enormously important functional capability. However, if one defines the functional capability for GPS as the precision with which one can locate a point on the earth, then the innovation of GPS was nearly an ε = 0 process (recall that an ε = 0 process is one where the rate coefficient remains at the value that it had when the mining process was initiated). When GPS left the prospecting phase in the mid 1970's and entered the mining phase, most of the key GPS technologies had been brought to the point where they could provide close to today's accuracy. It is unlikely that one will see many orders of magnitude increase in GPS location accuracy in the coming years. This is because it is unlikely that the technology has the potential to provide such increased precision, and it is also unlikely that orders of magnitude improvement in GPS location accuracy are needed (or therefore would be funded). What society desires of GPS is that it be maintained as a stable, reliable utility; continued investment into GPS reflects that objective.

The area associated with GPS that has seen real growth is that of providing broad access to the utility that was established by the government. This growth has been driven by advances in solid-state electronics that allowed the mass production of small, low-cost GPS receivers. The proliferation of GPS receivers would not have happened if the solid-state electronics industry had not made the investments and progress that it did. The GPS receiver community could never have obtained the funding necessary for the receiver development, if it would have become necessary for that community in isolation to make the investments and progress provided by the electronics industry and the government.

There are many innovations like the commercial proliferation of GPS receivers that are perceived as independent new technology but which are, in fact, largely spin-offs from the solid-state electronics industry and, as such, use extant infrastructure. They have the property of rapid insertion into the economy. These sorts of innovations are intimately related to developing an understanding of the S&T innovation conundrum. We will return to this point later.

We next consider S&T prospecting, to come to a better understanding of the salient characteristics that underpin this first, and necessary, phase of innovation.

ii. R&D Innovation Space and the Prospecting Phase of Innovation

The first phase of innovation, which we here term prospecting, is a very different beast from mining. For purpose of discussion we introduce the concept of R&D innovation space. This space is envisioned to be a large dimensional space composed of independent planes, analogous to the thought matrices of Koestler {Koestler, 1964}, that represent all aspects of innovation. A scientific or technical discipline is probably not a plane but rather is itself a multi dimensional subspace. For example, physics has many distinct planes such as electromagnetic, condensed matter, atomic and molecular physics, optics, gravitation, elementary particles, fluids, etc. Comparable distinctions apply for the other scientific and technical disciplines. There is a military plane or perhaps a larger subspace in which military endeavors intersect with the various other planes. There is also a financial plane or perhaps a larger subspace in which the government-funding agencies, private companies and venture capitalists move about. The mining process for a particular instantiation of a functional capability can be viewed as a special, planned trajectory that cuts through a subset of a larger R&D innovation space, only some planes of which are technical.

Figure 3.3 is an attempt to illustrate individual prospecting planes in this R&D innovation space. The prospecting plane P_1 shown in Fig. 3.3(a) may represent a scientific or technical field such as communications. The trajectory shown in the figure is meant to illustrate progress of an individual or a group as it conducts routine investigations within a subject matter prospecting plane. Figure 3.3(b) illustrates a trajectory in an independent or orthogonal prospecting plane, P_2. This plane might represent a different field, e.g., early radar, that may or may not be active at the time research is underway in plane P_1.

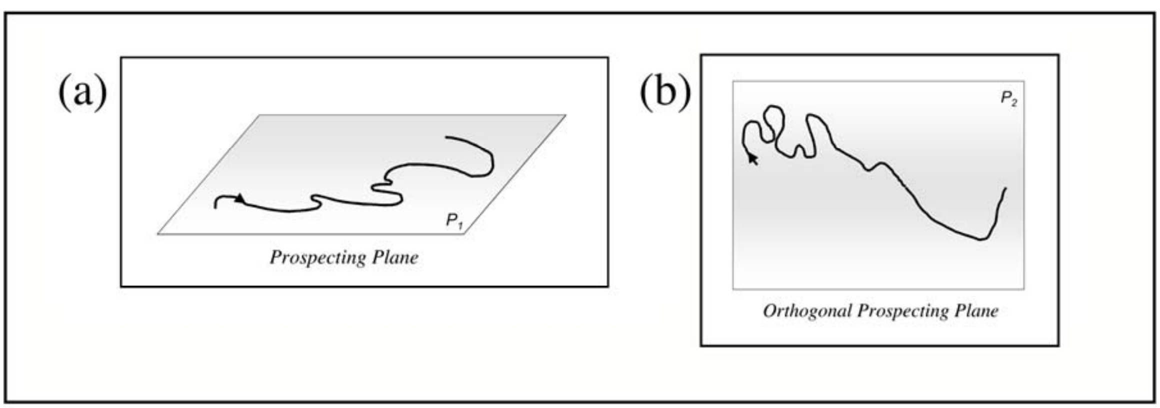


Figure 3.3. Prospecting planes in R&D innovation space. (a) an example plane P_1; and, (b) a prospecting plane P_2 that is orthogonal to plane P_1.

R&D innovation space, like other multi-dimensional spaces, has the characteristic that independent planes intersect along a line. This is illustrated in Fig. 3.4.

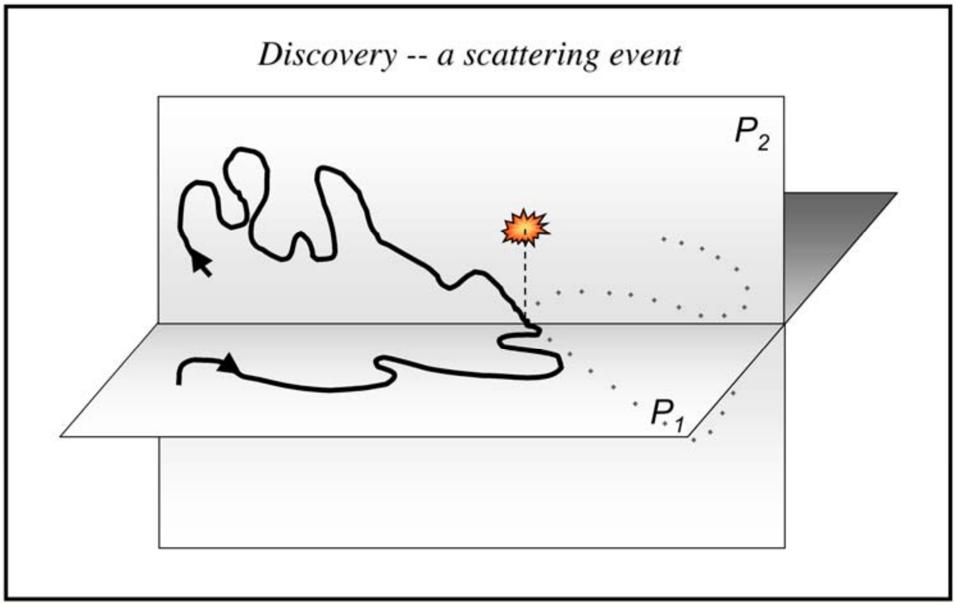


Figure 3.4. Intersection of two prospecting planes in R&D innovation space.

It has been observed that significant discoveries, inventions and innovations sometimes occur when the trajectories from independent disciplinary or functional planes intersect. When an important intersection occurs, and when this intersection is recognized, it often leads to radical modifications of the trajectories. In some cases this results in the establishment of entirely new fields of inquiry or completely new, unexpected innovations. This may be viewed as analogous to a physical scattering event between interacting particles, which results in completely new trajectories for each of the particles. The new trajectory may lie within the intersecting planes or might possibly scatter into an entirely new plane.

In some ways the discovery of radar was such a "scattering" case. Taylor and Young were working on communications problems. They were operating in the communications plane but fortuitously sampled the radar plane with their 1922 experiment. At that point there was no activity ongoing in the radar plane. However, in 1922 the communications trajectory intersected with the radar plane, and Taylor and Young were astute enough to understand the implications of this intersection. The result was the establishment of a trajectory that led to the innovation of early radar for the U.S. Navy, which contributed fundamentally to the increase in situational awareness and battlespace control exhibited by the U.S. Pacific Fleet during the course of World War II.

Trajectories in a prospecting plane are not predictable. They result from discoveries, funding opportunities, chance encounters with other scientists and engineers, and so forth. As a consequence the intersection of prospecting trajectories, as illustrated in Fig. 3.4, is a random process. There are many documented cases where prospecting trajectories crossed but no scattering event occurred. Unlike particle scattering in the physical world, scattering in the innovation world requires recognition and decision. While the R&D innovation space scattering process may be random, we should bear in mind the axiom "*Fortune favors the prepared mind.*" {Pasteur, 1854}. This involves human judgment, and a willingness on the part of researchers in prospecting planes to embrace radically new concepts and directions of research. For success, it also implies the need for individuals with effective authority (and resources, e.g., those operating in the financial planes and military planes) to foster the developing innovation through a time when possible consequences can only be intuited. If potential adversaries or competitors can make better-informed judgments at such junctures, they will prevail. This has important implications regarding organizational principles that are needed to maximize effectiveness in moving through the R&D innovation space.

An analytical description of the progress of prospecting through R&D innovation space is difficult to achieve. Scientists and engineers working in the prospecting planes are generally very systematic in their approaches. Further, they are usually operating according to rigorous rules sets that have been established over time in their disciplines. However, their objectives as "prospectors" are to find things previously unseen or unknown, and therein lies a randomness that makes modeling this phase very difficult.

Some scientific prospectors are simply trying to advance the state of knowledge. This is important since it forms the foundation for future progress and thereby modifies prospecting plane trajectories. Some are engaged in construction of exploratory systems or devices. Still others are in the business of connecting new knowledge with the needs of business, the military or of society in general. Out of such efforts have come most of the revolutionary scientific and technical changes in human history. The world in which we live today is a result of these efforts, as will be the world of the future.

For all practical purposes, while there is an infinite amount of knowledge yet to be found, there are not infinite resources to be applied. This somewhat vexing reality leads to a natural question: is it possible to more effectively search during the prospecting phase? The answer is certainly yes. Indeed, one need only examine the approaches used by modern biology that are employed to search the enormous databases that confront it, to realize this. However, as science and technology advances, the problems that are faced become more complex, thereby placing a limit on the pace of routine progress. The problems that can be solved with today's technologies are solved and those that cannot remain open until the required knowledge and capabilities are achieved (e.g., Fermat's Last Theorem {fermat-s-last-theorem}). In addition, in attempting to speed up discoveries in the prospecting phase one cannot avoid the reality of an even more fundamental rate limiting process, that of human creative thought.

A simple chemical analogy might make this point. Consider the well-known reaction that describes the reaction of nitric oxide and hydrogen to make water and nitrogen:

$$2NO + 2H_2 = N_2 + 2H_2 + 2H_2O \ . \qquad (14)$$

The above reaction is actually a two-step reaction where the intermediate step involves hydrogen peroxide that, once produced, reacts quickly with hydrogen to produce the end products. However, the rate of the composite reaction is essentially independent of the fast hydrogen peroxide rate; it is determined by the much slower nitric oxide-hydrogen rate. The hydrogen peroxide reaction could proceed infinitely fast but it would not speed up the composite reaction. The nitric oxide-hydrogen reaction is called the "rate-limiting step." There is an analogous rate-limiting step involved with discovery and innovation. That step involves the time it takes for human beings (in this case scientists and engineers) to come to grips with the information presented to them and to assess meaning and assign significance. More capable scientific instruments and faster and more precise analytical tools allow scientists and engineers to address ever more complex problems. However, it still takes finite time for the creative human brain to assimilate data, turn data into information, and usefully understand what is occurring. Technology is far from being able to replace the required human creative contribution. The creative human process is the central rate-limiting step of prospecting.

Perhaps the speed with which data can be circulated and processed by modern information technology will fundamentally alter this situation. However, this rapid movement and processing of data may be analogous to the rapidly reacting hydrogen peroxide generation that is hidden in Eq. (14). Consider the early radar innovation that was discussed above. We know that Page was quickly aware of new commercial vacuum tube developments (that were being provided to him from work ongoing in the mining phase of vacuum tube R&D), and that he was current in the relevant scientific literature to within a few months of its publication. It is doubtful that his completion of the radar innovation would have been accelerated by more than a few months (of the 14 years for early radar prospecting) if the tubes and the information had been instantly provided to him upon their availability. The limiting process was that of doing the careful work and the thinking that were necessary to put the puzzle together.

The previous discussion points to a relationship regarding the interaction between the prospecting phase and the mining phase for a given innovation. The next two sections will consider that interplay among innovations, in the evolution of functional capability.

iii. Evolution of Functional Capability

Our discussion so far has focused on the emergence of individual technologies and how they came together to create particular realizations of a functional capability. It is the creation of functionality capability that society is primarily interested in, rather than the development of any specific science or technology. This can be illustrated by examining computing, where functional capability may be defined as the number of calculations per second. In his book "The Age of Spiritual Machines" Kurzweil {Kurzweil, 1999} provided a plot of this functional capability—the number of calculations per second that one could purchase for $1,000—spanning the decades of the 20th Century; a modified version of his plot is shown in Fig. 3.5.

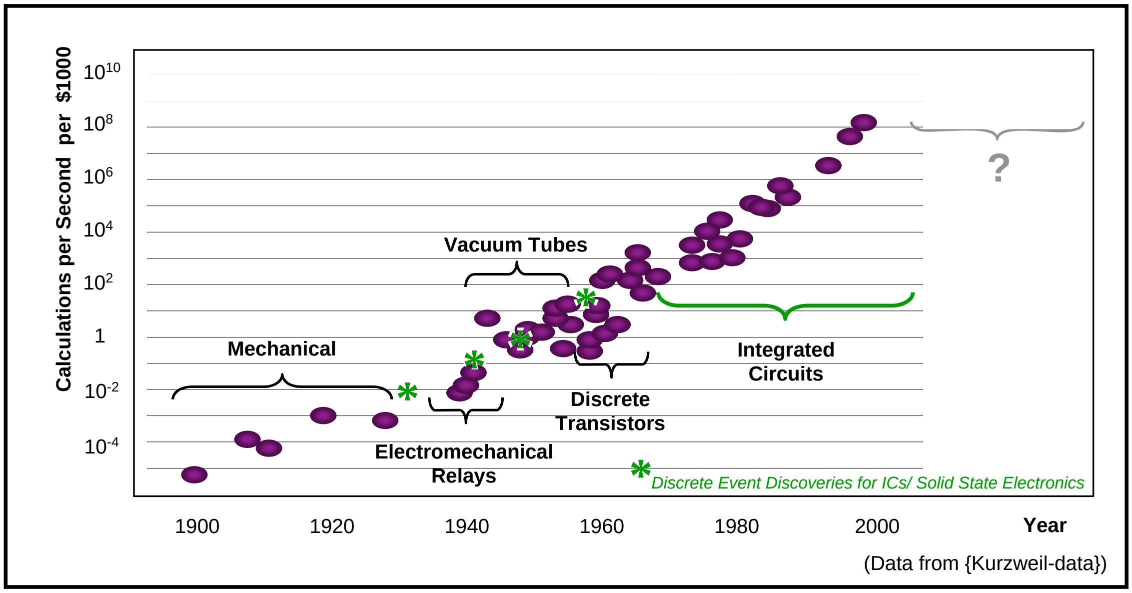


Figure 3.5. Functional capability of computing (calculations/ sec for a fixed cost of $1000) *vs* time (adapted from Kurzweil).

Between the years 1900 and 2000, the functional capability of computing increased by an astounding 13 orders of magnitude. What is especially interesting is that the functional capability over this time was provided by five very distinct particular technologies: tabulating machines; electromechanical relays; vacuum tubes; discrete transistors; and, integrated circuits. Each of these technologies was able to provide some increase in functional capability until it matured and was eclipsed by a different technology.

Since the late 1960's, the six orders of magnitude increase in functional capability has been provided by the integrated circuit, where the increased capability is described by Moore's Law {Moore, 1965}. The discussion of Sec. III*i* (per Eq. (11)) indicates that if this development had been a zero-feedback "$\varepsilon = 0$" process, it would have taken about 30 million years, instead of a mere 30, to accomplish this result.

Generally speaking, if the history of computing is an indicator of the future, then the integrated circuit will eventually be replaced by a new technology. Perhaps it will be a modification of the current technology, or will be a totally different technology, or will be new software developments with extant hardware. At this time no one knows what, if anything, will prevail. However, a great effort is presently underway to find the next technology that will continue the remarkable advances of computing as a functional capability.

Figure 3.5 suggests some degree of predictability. The typical textbook view is often stylized as illustrated in Fig. 3.6, where functional capability is described as a sequence of "logistics curves" of the sort prevalent in biology, economics, agriculture, etc. (e.g., {wikipedia}). Each individual curve in Fig. 3.6 indicates a continuous process of technology development from birth through maturity. Overall, this figure implies that each technological instantiation of a useful functional capability is followed by a succeeding technology that progresses along a temporally similar pattern of development.

However, consideration of the innovation histories discussed in Sec. II leads one to doubt that this sort of predictable birth-to-maturity cycle is fully representative of reality: while the mining phases of innovation for a functional capability may be represented by the upper halves of logistic curves, random processes inherent in prospecting make it impossible to predict when relevant new innovations might appear, or when the subsequent mining phases will commence.

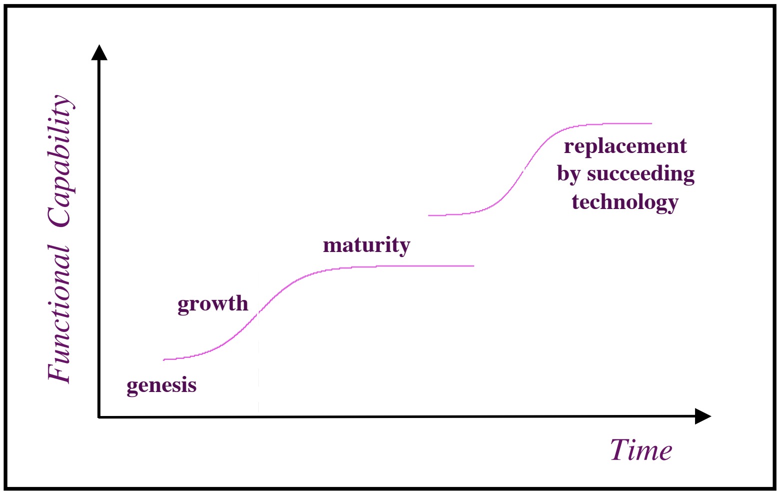


Figure 3.6. Functional capability illustrated by logistics curves.

From the discussion of the major phases of an innovation in Sec II*iii* above, a functional capability curve such as that diagrammed in Fig. 2.2 should replace each of the sequential logistics curves sketched in Fig. 3.6 with a different sort of sequence of growth. This is illustrated in Fig. 3.7.

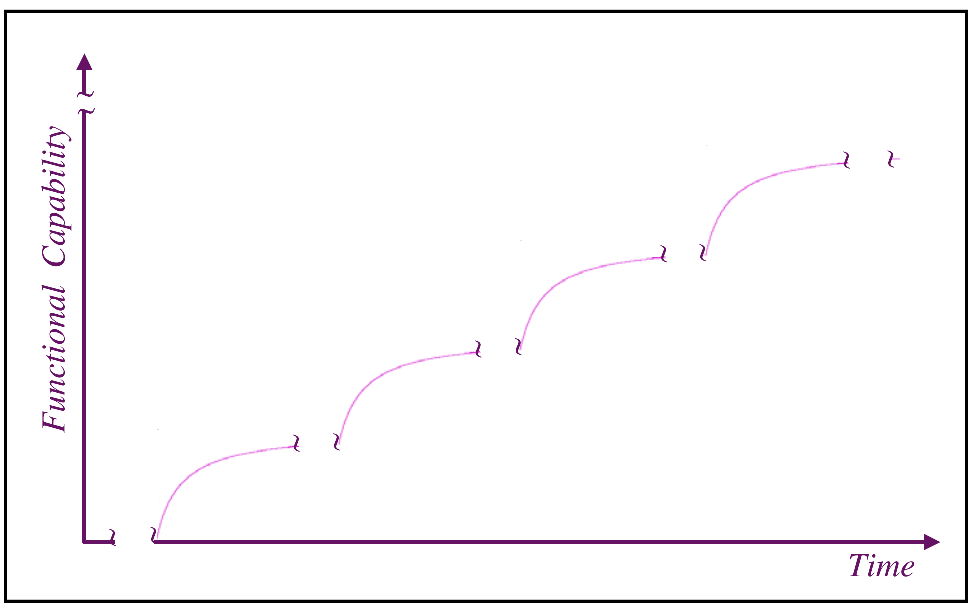


Figure 3.7. Functional capability across multiple technologies.

As depicted in Fig. 3.7, little if any functional capability is available from a particular innovation technology during the prospecting phase; this is specifically indicated in Fig. 2.2, and envisioned as lying between the squiggly lines in the flat parts of the curve in Fig. 3.7. In reality, the prospecting phase for the next mining phase of a functional capability is usually well underway during a previous mining phase (e.g., note the asterisks in Fig. 3.5, which indicate solid-state electronics discovery milestones). When technologies come together to the point that a systematic program can be defined and one enters the mining phase, the process appears from that time to be repeatable, implying that an analytic description should be possible. However, as discussed above, the prospecting phase by definition involves random processes and does not yield to a simple analysis. The entire process that describes increase in functional capability is thus far from deterministic.

iv. Creative Destruction

One mining phase innovation will likely displace another at the point when the new technology cost-effectively offers more functional capability, e.g., GMR magnetic memory read heads (*cf* Sec. II*iif*) as compared with read heads that relied on the MR effect. For simplicity of illustration of this general process, assume that the new technology offers three times the functional capability potential of the technology that it replaces. Assume also that the new technology starts its mining phase with the same process rate coefficient R as the old technology, and that this rate coefficient R remains constant throughout.

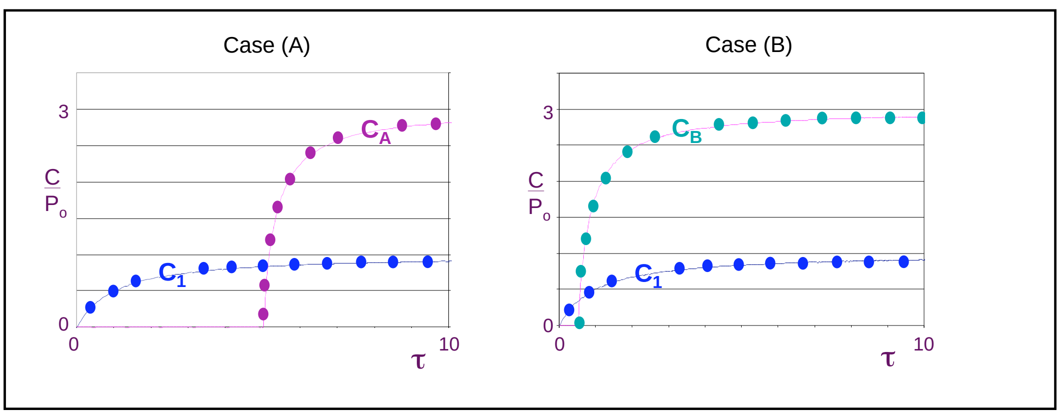


Figure 3.8. Functional capability technology replacement: in Case (A), left, a replacement technology with capability C_A enters the mining phase when the previous technology C_1 is approaching its asymptotic limit; and, in Case (B), right, a new technology with capability C_B enters the mining phase before the potential C_1 of the previous technology is depleted. Case (B) illustrates the process of creative destruction.

Figure 3.8 sketches functional capability time histories for old and new technologies for two scenarios under these assumptions. In Case (A) a new technology with functional capability C_A enters its mining phase as the old technology C_1 is approaching its asymptotic limit. By the time the new technology replaces the old technology, those invested in the old technology would have moved on. However, in Case (B), a new technology C_B enters its mining phase well before C_1 has reached its mature level. This occurs, in the absence of monopoly, when a new technology is cost competitive relative to the old technology. Entrepreneurs will then move to replace the old technology long before the old technology has produced all that it potentially could, and will do so without concern for the profits of those who are vested in the old technology. (In a monopolistic situation there would be little incentive to behave as in Case (B). The monopoly would likely delay the introduction of the new technology, thereby converting Case (B) into Case (A).) Case (B) is an illustration of what Schumpeter calls "Creative Destruction" {Schumpeter, 1942}. Many economists believe creative destruction to be the underlying market-driving force of capitalist systems.

As a specific example of Creative Destruction consider the recent history of magnetic storage for hard disc drives. Figure 3.9 presents the areal data storage density (in Gigabits/ in^2) from 1975 through 2001. The total improvement in areal density over the 25-year period shown in Fig. 3.9 is nearly five orders of magnitude. However, four different technologies contributed to this overall increase in functional capability, with each providing between one and two orders of magnitude improvement.

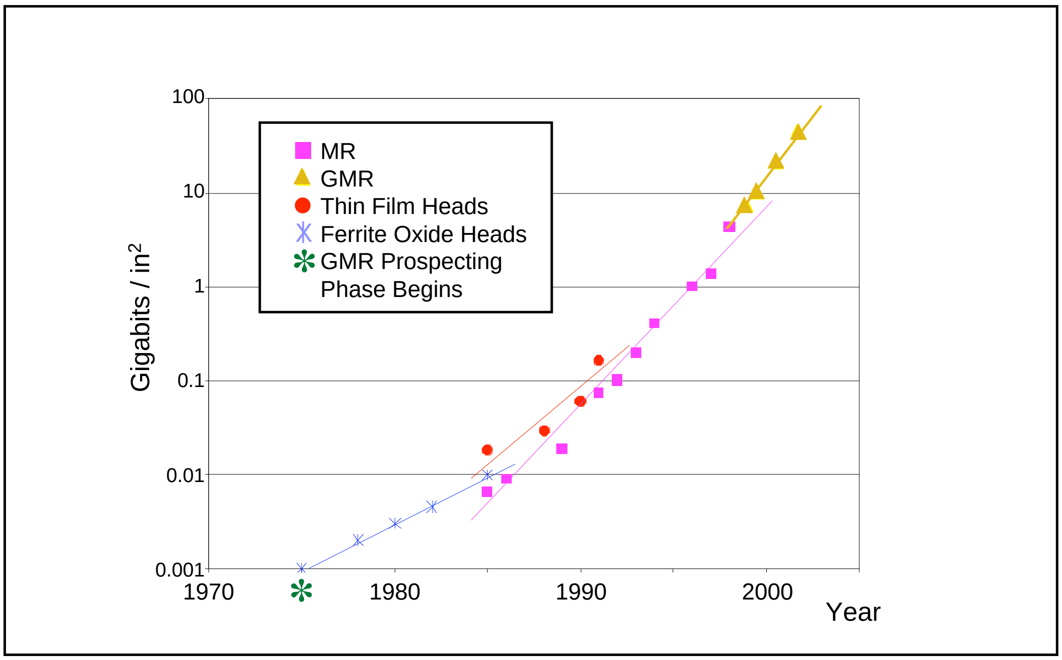


Figure 3.9. Magnetic Storage Density (Gigabits / in^2).

Each of the four distinct trends evident in Fig. 3.9 arises from a different technology. The first trend represents the era of ferrite-oxide heads. This era ended in the mid to late 1980's when the mechanical grinding of the ferrite heads to ever-smaller dimensions had reached the point of diminishing functional return. That asymptotic state led to the second trend, beginning in the mid-1980's, which was the introduction of thin film techniques where photolithography was used to produce ever smaller electromagnets. This introduction is similar to the situation shown in Fig. 3.8a, where the old technology had run its course before the new technology was introduced. However, the thin film induction heads had barely entered the market when they were replaced by the superior technology of magneto-resistive heads shown in the third trend. MR heads involve much thinner films and were able to more easily exploit the photo-lithographic technologies developed by the semiconductor industry. This third trend was clearly Creative Destruction. It lasted about ten years and added two orders of magnitude to the areal storage density. It was then overtaken by a forth trend, that of giant magneto-resistive heads. This again was Creative Destruction, since the MR heads had not reached their asymptotic limit of utility when they were replaced and the insertion was driven by competitive forces. The GMR heads currently dominate in the magnetic memory hard disk drive market.

These trend lines describe the mining phase for each of the technologies. However, as was discussed for MR and GMR in Sec. II*ii*f, the prospecting phase for the technologies began much earlier than the dates attributed to the technologies in Fig. 3.9. For instance, the start date for the prospecting phase of the fourth trend, GMR, is marked with an asterisk in Fig. 3.9. The prospecting phase for GMR began in 1975 and ended about 1992 when the mining phase commenced (see also Fig. 2.1). It is believed that the mining phase for GMR will mature in the 2010 time frame {Prinz, 1995} as the areal densities approach the physical limits set by the thermal stability of small magnetic domains. Hence, the mining phase for GMR heads will last about as long as it took to get GMR through the prospecting phase.

The above discussion sheds some light on the S&T innovation conundrum. The market place and the public are largely oblivious to the (long) prospecting phase of an innovation such as GMR; they focus on, and invest in, the mining phase. As a result they see the magnetic memory functional capability doubling predictably every two years through 2010. This reliable increase is perceived by society as the pace of progress of a revolutionary technology. In fact, the GMR revolutionary phase was over by 1992. Subsequent to 1992, manufacturing technologies were instituted to allow the systematic exploitation of this important technology. Had the manufacturing technologies not been put in place nothing would have been exploited. However, had the work that took place between 1975 and 1992 not occurred, there would have been nothing to exploit. The S&T timescales conundrum has its roots in these two different timeframes (prospecting and mining), and the public visibility of each. Clearly great progress is being made at a rapid pace through the exploitation of technology. However, it is important to keep in mind that innovations like GMR do not just appear out of nowhere.

v. The Flow of Innovation

Consideration of the above discussion suggests that the S&T conundrum has its roots in the complexity of an innovation ecosystem. Fundamental to the ecosystem is the research and development innovation space and the separate but related innovation phases of prospecting and mining. Society at large is the principal driver of innovation; its needs and its resources will determine ultimately what innovation gets done. The relationship with the mining phase is that of a consumer and a short-term investor. The relationship with the prospecting phase involves long-term vision, long-term hopes and long-term investments. A schematic of the dynamic, interdependent flow of innovation is sketched in Fig. 3.10.

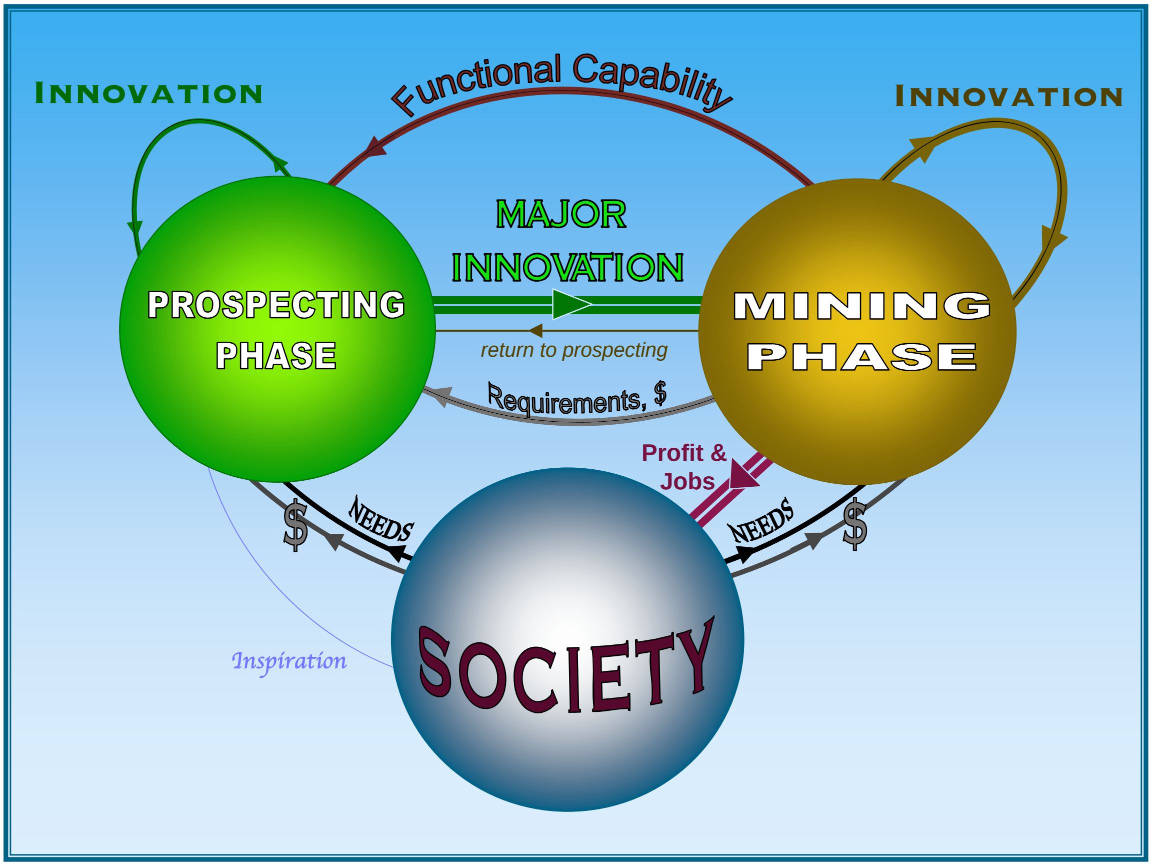


Figure 3.10. The flow of innovation.

Figure 3.10 suggests that major innovations flow from the prospecting phase to the mining phase and that there is a substantial return to the prospecting phase through the functional capability produced in the mining phase, thereby enabling further work in the prospecting phase. This "return flow" provides the prospectors the tools to increase productivity and to address ever more difficult problems. Examples include the vacuum tubes and other commercial technologies used by Young and Page in the invention of early radar, and the introduction of digital lab bench equipment that allowed the automatic recording and analysis of data such as was used for the discovery of DNA fingerprinting. Still other tools result from the advance of modern computers that translate into advances in computer simulation, which enables solution to more complex problems.

There is also a weak innovation return flow from the mining phase to the prospecting phase for innovations that arise in the mining phase but are not ready for exploitation. Sometimes innovations emerge from the mining phase but the state of technology is not advanced to the point where they can be carried productively forward. Those innovations may be passed back to the prospecting phase for further work. Directed energy weapons is an example of an innovation that emerged from the mining phase of lasers but has been returned to the prospecting phase several times, or has co-existed in both phases.

The figure also suggests that there is a substantial amount of innovation that emerges from the mining phase and returns directly to the mining phase for exploitation. Examples of this include GPS receivers, the internet (as distinct from ARPANET), electronic control of automobiles, and the use of microprocessors to replace previous analog controls with digital control systems in general.

Finally, there is a path for innovation that emerges from the prospecting phase and returns directly to the prospecting phase. This last type of innovation is often in the category of practices and techniques that scientists and engineers discover or develop which have little marketability but are very helpful to others in the prospecting phase. An example of this is the work of Prinz noted in Sec. II*ii*f, which helped other prospectors as they progressed towards the discovery of GMR.

This depiction of innovation is considerably dependent on the external society in which the processes of innovation reside. The imperatives of society's needs, of wars, natural disasters, global philosophies, political constructs, economic growth, all contribute to and fundamentally influence the linkages depicted in Fig. 3.10. Economic forces, in that they drive the resources available to innovation, are central to the flow of innovation. We thus need to consider present economic thinking regarding the contributions of science and technology, in order to come to a better understanding of the S&T innovation conundrum facing us today.

IV. Economic Insights

Economic forces that drive the resources available to innovation are central to the evolution of innovation. Over the years various models have been developed that attempt to explain the role technology plays in economic growth. While these models lie well outside the scope of this paper, they do have some bearing on the subject addressed in this paper. For example, in addition to examining economic growth the models are used as guides to strategies regarding what should motivate investments in R&D. They can also provide some ability to anticipate the effect on R&D of various incentives regarding the investment of R&D funds, and may shed some light regarding the direction in which R&D is moving. We will, therefore, discuss briefly some aspects of these models.

i. Solow and Romer: Exogenous versus Endogenous Considerations

It appears that modern economic models that deal with the impact of technology on economic growth trace back to the work of the Nobel Prize winner Robert Solow {Solow, 1956}. Solow assumed that the effects of technology on economic output manifest themselves through the accumulation of knowledge A(t) generated by investments in R&D. This is done by creating an "effective labor" A(t)L(t) where L(t) represents the labor force. Knowledge is viewed as an enhancement of labor productivity. In Solow's model, A(t) and L(t) were imposed (i.e., they were exogenous to the model rather than a product of the model). Imposing A(t) on the model describes the situation where decisions on science and technology investments take place outside of the economic sector, or are accidental spinoffs of the economic sector. This would be the case, for example, where the government or other entity would fund science and technology as a public good.

Subsequent to Solow's work it became increasingly clear that technological progress was a significant (perhaps the most significant) factor in long-term economic growth. This led to research regarding models in which technological change is included as arising from intentional decisions made to maximize profits (e.g., {Romer, 1990}). Within these models a trade-off occurs between those doing research and those doing production engineering so as to maximize profit. Romer considers three economic sectors. The first is involved in producing new designs (considered as research) and is characterized by perfect competition. The second sector produces intermediate capital goods and operates in monopolistic competition. This sector purchases "designs" from the research sector. The third sector is involved in final output and is characterized by perfect competition. This sector uses unskilled labor and technical human capital and competes for that human capital with the research sector. In this model, generation of technical knowledge A(t) is assumed to satisfy the relationship

$$dA/dt = \delta H_A A(t), \tag{15}$$

where δ is a productivity parameter and H_A is the human capital working in the research area. Equation (15) implies that the rate of change of knowledge is proportional to the number of people generating knowledge and to the total knowledge available. We will discuss the ramifications of this later point in the next section. Romer defines the total human capital H involved in technical matters as

$$H = H_A + H_Y, \tag{16}$$

where H_Y is the human capital working on production design. Under his assumptions Romer shows that the profits are maximized when $H_Y = (\Lambda/\delta)r$. Hence, the knowledge growth rate g_A is related to the interest rate r by the equation

$$g_A \equiv [dA/dt]/A = \delta H - \Lambda r, \tag{17}$$

where Λ is composed of fixed parameters of the model.

It is important to note that Eq. (17) indicates a negative connection between the knowledge growth rate and the interest rate (i.e., the growth rate falls as the interest rate increases). It is also important to note that the growth rate increases as the total human capital increases. Furthermore, since g_A cannot be negative by definition, the total human capital H must exceed $\Lambda r/\delta$ for there to be any growth. Under the balanced growth assumptions employed by Romer, the output of the economy is proportional to the knowledge A(t). Hence, the knowledge growth rate is the economic growth rate.

Romer concludes, among other things, that within this endogenous technology investment model:

- The stock of human capital determines the rate of economic growth.
- Too little human capital is devoted to research when the economy is in equilibrium.
- The rate of technological change is sensitive to the interest rate.
- A subsidy to physical capital accumulation may be a very poor substitute for direct subsidies that increase the incentive to undertake research.
- Integration into world markets will increase growth rates.

The Romer model provides some insight as to how R&D investments change course due to economic conditions, when they are determined endogenously by profit maximization. In the next section we attempt to describe prospecting characteristic timescales using Romer's nomenclature, with a purpose of evaluating ramifications of applying an endogenous perspective to this phase of innovation.

ii. The Characteristic Time for Prospecting

Here we use our research experience to examine the characteristic time for innovations emerging from the prospecting phase of S&T. It seems reasonable to assume that the innovation rate from the prospecting phase is proportional to the rate at which knowledge A(t) is generated in this phase.

If we employ Eq. (15) that expresses Romer's model for knowledge generation, we find that

$$A(t) = A_o \exp\{\int \delta H_A(t)dt\} \tag{18}$$

where A_o is the initial value of A(t) and $H_A(t)$ is the total human capital involved in research at time t. Equation (18) requires that the rate of knowledge generation per person a′(t) is

$$a'(t) = (dA/dt)/H_A(t) = \delta A_0 \exp\{\int \delta H_A(t)\, dt\}. \quad (19)$$

The scaling implied by this equation is interesting. It states that the rate of knowledge generation per scientist and engineer increases exponentially with the number of scientists and engineers as well as with time. If true, this could have profound implications for how economic regions manage their knowledge generation.

We can gain some insight into the above choosing a specific time dependence for $H_A(t)$. NSF data for the period 1980-2000 is consistent with the scientific and engineering workforce roughly tracking the GDP {NSF1, 2004}. Therefore, for simplicity, assume that

$$H_A(t) = H_0 \exp\{\alpha t\} \quad (20)$$

where α is the GDP growth rate. The GDP data since 1953 can be fit with $\alpha = 0.0334$, when time is measured in years. Upon substituting this value into Eq. (20) and evaluating the integral in Eq. (19) one finds that

$$a'(t)/a'(0) = [\exp\{\delta H_o\}]^n \quad (21)$$

where

$$n = 29.9(\exp\{0.0334\, t\} - 1) \quad (22)$$

The quantity $[\exp\{\delta H_o\}]$ is the annual growth factor for knowledge at the initial time when time is measured in years. Equation (21) can be viewed as a compound interest calculation. We do not know the value of the growth factor. We can, however, assign values to this factor to get some sense for the scaling. This is done in Table 4.1, where we have included the initial knowledge doubling times associated with the selected factors and evaluated Eq. (21) for 20 years and 50 years.

Table 4.1. Rate of Knowledge Generation Ratio a ' (t)/a ' (0).

Initial doubling time (years)	$[\exp\{\delta H_o\}]$	a′(20)/a′(0)	a′(50)/a′(0)
2	1.350	5,023.000	6×10^{16}
7	1.100	15.000	216,530.000
20	1.035	2.700	84.000
70	1.010	1.330	3.600
140	1.005	1.150	1.900

Under any of the values of $[\exp\{\delta H_0\}]$ given in Table 4.1, the knowledge production rates for individual scientists and engineers show substantial increases. We would expect knowledge to grow at least as fast as the science and engineering work force. That corresponds to a doubling time of about 20 years. For this case, Table 4.1 predicts that individual scientists and engineers would increase their knowledge production rates by a factor of 2.7 in 20 years and by a factor of 84 in 50 years. Such factors would be difficult to miss. If such an effect exists it should be visible in such knowledge measures as patent activity and publication activity. However, an examination of the available data indicates that, in the United States, the number of patents per scientist and engineer has actually been declining for the past 50 years; see Fig. 4.1 {Wilson, 2003}.

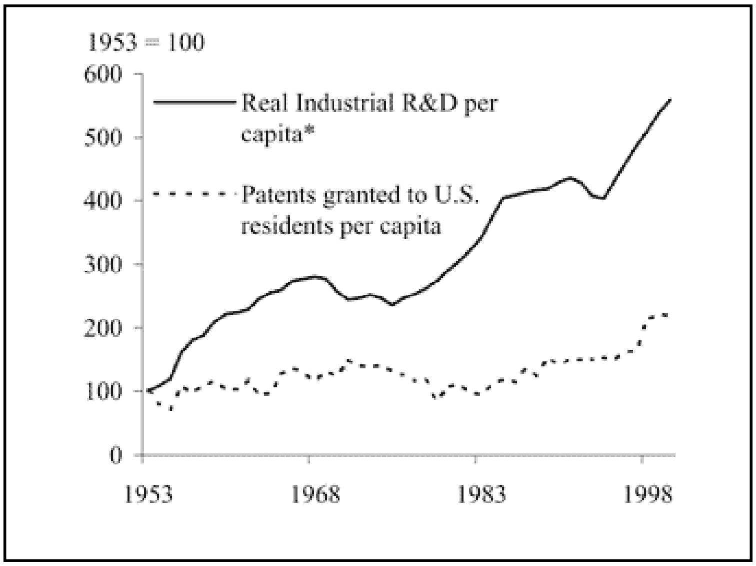


Figure. 4.1. Per Capita Patents and R&D, 1953-2000. Industrial R&D refers to all company-performed R&D, including that funded by the federal government. {Wilson, 2003}

Similarly, U.S. scientific and technical papers per scientist and engineer have been declining for at least the past 15 years; see, e.g., Fig. 5-30 of {NSF2, 2004}. The scientific and technical publication output for the OECD countries (excluding the United States) grew at an annual rate of about 2 % over the same period. However, world GDP grew at a rate of about 3 % over this period. If we assume that science and engineering employment tracked the GDP (as it did in the United States) then the publication per scientist and engineer in the OECD countries also declined over the past 15 years. There may be reasons why these rates have declined (such as the production of better patents) but it is unlikely that they could overcome the scaling shown in Table 4.1.

Other indicators should also be considered. For example, learning calculus in high school still takes one or two years. It still takes four years to get a BS degree and five to seven years to get a PhD. Of course, the PhD-level independent research problems that are undertaken are matched to the current functional capability and so have increased in complexity, but still, time to solution is about the same. Finally, as discussed in Sec. II, the time scale for truly new innovations to emerge from the prospecting phase has not changed in the past one hundred years.

From the above discussion it appears that the rate of knowledge generation per scientist and engineer must have a much weaker time dependence than that suggested by Eq. (15), or δH_0 is a very small number. This is important because Eq. (15) contributes to the perception of ever more rapid production of knowledge at the individual scientist and engineer level and therefore contributes to the apparent S&T conundrum.

Knowledge generation in science and engineering is complex and is not describable by a simple function. An examination of history suggests that knowledge generation is episodic. The appearance of figures like Newton, Maxwell and Einstein have precipitated periods of enormous productivity that are often followed by more quiescent or routine periods of knowledge production. Our inclination is to not be too specific regarding the details. It does, however, seem reasonable that the rate of knowledge generation should be proportional to the number H_A of scientists and engineers working on knowledge generation. It should be recognized that only a small fraction of the S&T workforce produces revolutionary innovations. Within reasonable bounds, we expect that this fraction will remain constant and therefore that the number of true innovators will scale with the total science and engineering workforce. This suggests the general equation

$$dA/dt = H_A(t)R(t). \qquad (23)$$

The rate R(t) could be viewed as the average over all individuals of the individual rates of knowledge generation. In the abstract, R(t) could be any function including that given by Eq. (15). If one replaces Eq. (15) with Eq. (23) and repeats the Romer analysis one finds that the growth rate g given by Eq. (17) is replaced by the following expression

$$g = HR(t)/A(t) - \Lambda r. \qquad (24)$$

This is not a surprising result and leads to Romer's expression when R(t) is given by Eq. (15). Therefore, Romer's conclusions would seem to follow if R(t) were some function other than that used in Eq. (15) including it being a function that is slowly varying in time or even independent of time. Of course, if the quantity H(t)R(t) grows more slowly than A(t), and the above equation were to be used for long periods of time, then the growth rate would eventually go to zero. However, because of the complexity involved with knowledge generation, expressions such as given by Eq. (24) are most valuable for what they contribute to understanding how things scale with parameters such as interest rate rather than for what they contribute to absolute determination of growth rates. Applying Eq. (24) or Eq. (15) over long periods of time without appropriate adjustments would not be wise. Nevertheless, the insights provided by these equations into such matters as sensitivity to interest rates and to human capital when technology investments are made in an endogenous fashion should be helpful.

The reason that Eq. (15) predicts such a dramatic rise in the rate of knowledge production with time lies in its assumption that all knowledge is available to each individual scientist and engineer and, all knowledge therefore increases the knowledge producing capacity of the individual scientist and engineer. While this assumption is correct in the abstract it is not correct in practice. It is important to not confuse productivity enhancement with increase in the rate of knowledge generation.

For example, the functional capabilities developed over the past 100 years have dramatically increased productivity related to the conduct of operations that are well understood. In science and engineering the process of data collection and analysis has been transformed. Much of the drudgery has been eliminated, results are analyzed as they are produced and entire job categories involved with data collection and analysis have been eliminated. However, this type of productivity enhancement does not translate directly into knowledge generation. The rate limiting process of understanding new results and turning that understanding into new knowledge and new functional capability remains to be done by human beings who are no more intelligent than they were 100 years ago.

Similarly, the functional capability produced over the years has allowed research into previously inaccessible regimes. Nanotechnology is a good example of this. The "machine shop of the 21st century" now allows work at nanometer dimensions. However, the rate limiting process involved in sorting out what is going on in this regime will still determine the rate of progress in producing new functional capability from this regime. The fact that data and information move at an ever-increasing rate does not mean that new knowledge follows this rate. Indeed it may have the opposite effect. The Nobel laureate economist Herbert Simon summarized this situation as follows {Simon, 1995}: "What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention, and a need to allocate that attention efficiently among the overabundance of information sources that might consume it." Anyone who has coped with email over the past few years understands of what Simon speaks.

The practical reality is that the creative individual scientist or engineer can deal creatively with only an infinitesimal fraction of the total knowledge that is in principle available to them. If they try to expand beyond this they become overwhelmed and their creative productivity drops. This is the crux of the explanation of why significant innovations seem to take 15 to 20 years in the prospecting phase regardless of when they occur. Fifteen to twenty years represents a few characteristic times for knowledge generation by an individual or group of scientists and engineers. This is the manifestation of the rate limiting process discussed Sec. III*ii* above. When individuals are confronted with very new information—i.e., not preprogrammed as in the mining phase—they sort it out in their own characteristic time (which is related to the time it takes for those individuals to grasp a new concept, and thus are more closely correlated with the latent intelligence of those individuals than to the level of sophistication of, e.g., the laboratory equipment or computers available to them).

If this workforce is subject to the pull and tug of endogenously driven swings of emphasis (e.g., H_A becoming H_Y (research to production)) as a function of market forces as discussed by Romer, and if these swings occur on timescales that are more rapid than the 15-20 years that are characteristic for a given prospecting activity, then it will not be likely that prospecting will yield many results. Hence, when prospecting is at the mercy of endogenous forces, there is a significant and dangerous probability that the creative time required for discovery will be insufficient for success.

iii. General Purpose Technologies

Many of the technologies for which the S&T innovation conundrum seems most pronounced involve solid-state electronics or technologies derivative thereof. For example, the rapid advances in computer power are directly related to the clock speed and transistor density of electronic chips. The magnetic memory read heads discussed in Secs. II*ii*f and III*iv* above have at least two relationships to solid-state electronics: one deals with increased demand for computer memory; and, the other deals with the fabrication technologies required for the read heads. The MR and GMR heads were able to draw very heavily upon the thin film deposition techniques and the photolithographic technologies that were developed to support the semi-conductor electronics industry. This allowed the read heads development trajectory to start with a very large process rate coefficient (R_0 of Sec. III*i*) that was determined by the manufacturing technology and the accumulated knowledge resulting from 40 years of work by the semi-conductor industry. Furthermore, the initial functional capability potential (P_0) for each of these magnetic storage technologies was such that each increased the functional capability by a factor of from 10 to 50 as compared to the factor of 10^6 provided by the Si, SiO_2 MOSFET in the solid-state electronics industry (but each was economically competitive even at such reduced factors). This underlying, encompassing importance of solid-state electronics suggests that there is something very special about that innovation.

This observation has also been made by researchers who develop economic models to investigate economic growth. In recent years, that community has introduced the term "general purpose technologies" (GPTs) to capture the impact of certain special technologies on economic growth (see, e.g., {Helpman, 1998}). Lipsey, Bekar and Carlaw {Helpman, 1998} suggest that GPTs share the following characteristics:

- wide scope for improvement and elaboration;
- applicability across a broad range of uses;
- potential for use in a wide variety of products and processes; and,
- strong complementarity with existing or potential new technologies.

Our central interest in GPTs is not with the impact that they have on economic growth but rather with the impact that they have on the progress of R&D as effected by investment of R&D funds. A recent paper by Dowling {Dowling, 2003} studies the behavior of entrepreneurs and the role of GPTs regarding cycles of economic growth. Dowling notes that "a GPT has the endearing quality of being able to "plug into" many existing avenues for technological exploitation." He introduces the term "genesis innovation," which we interpret to relate to the prospecting phase, and the phrase "post-genesis innovation," which we interpret to apply to the mining phase. He defines "GPT search" as pursuing a genesis (prospecting) innovation, and "GPT exploitation" as pursuing a post-genesis (mining) innovation. He notes that "a majority of entrepreneurs will opt for the pursuit of innovation within an existing GPT (GPT exploitation) in preference to GPT search until the number of financially viable product innovations is nearly exhausted." This latter behavior is due to the assumption that the probability of finding a financially viable innovation is higher in the mining phase than in the prospecting phase.

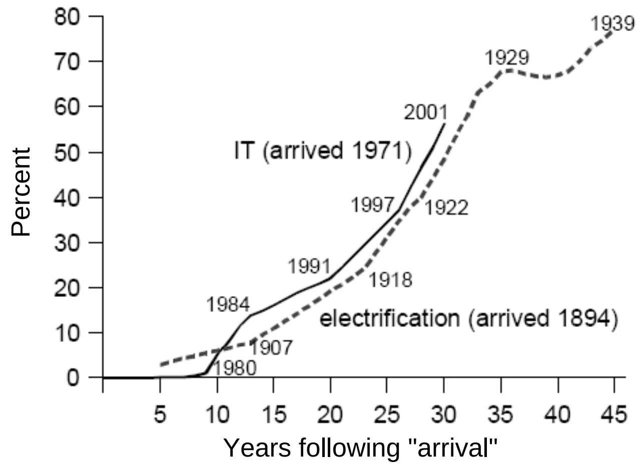
When a mining phase is associated with a GPT then it has the following characteristics identified by Bresnakan and Trajtenberg {Bresnakan & Trajtenberg, 1995}:

- It is pervasive and spreads through most sectors of the economy.
- It gets better over time and the therefore reduces costs to users.
- It makes it easier to invent and produce new products and processes.

The innovation represented by solid-state digital electronics or its derivative innovation now referred to as information technology (IT) certainly satisfies the above characteristics. As a result, in addition to the planned innovation mentioned above, there arise great opportunities for additional innovation as the technology improves, the infrastructure grows, and the technology spreads throughout the economy. Hence, obvious opportunities develop for rapid insertion as a result of the progress of a GPT. The microprocessor is a case in point. Its advent permitted the rapid replacement of complex control systems with systems that required much less power and turned out to have much greater potential. This led to a large number of rapid insertions and innovations. The automobile, for example, has gone from a primarily mechanically controlled/ regulated system to an electronically controlled system, greatly modifying the performance and the maintenance of the automobile. The electronically controlled automobile then in turn precipitated a new series of innovations for electronic control modules and diagnostic software and systems. These types of the innovations have their time scales influenced by insertion opportunities (e.g., new models, retooling costs and competitive pressures).

Jovanovic and Roussau {Jovanivic & Roussau, 2003} assert that electricity and IT are perhaps the two most important GPTs to date. Solid-state electronics is, of course, the technology that underlies IT and certainly satisfies the criteria developed by Lipsey et al.

In Fig. 4.2, Jovanovic and Roussau compare the diffusion of personal computers throughout American households with the diffusion of electric service throughout American households. It is interesting that the two GPTs plotted in Fig. 4.2 follow very similar curves. The diffusion of the electric service provided a path for rapid innovations when the associated technology allowed the innovations to become cost competitive (e.g., washers, dryers, TVs, air-conditioning, electric lighting, etc). Similarly the diffusion of personal computers allowed rapid innovation in a number of areas (Internet, e-Bay, on-line banking, on-line shopping, tax preparation software, etc). Many of these types of innovations will occur quickly because they do not require new input from the prospecting phase. They emerge from the mining phase and rapidly re-enter it. For example, the prospecting phase of the Internet had been done (ARPANET) before the PC had diffused into the economy, so there was no need to go back to the prospecting phase. The prospecting phase for TV occurred in parallel with the diffusion of electric service. Other developments just used the availability of the new infrastructure to innovate regarding the way normal things were done (e.g., banking, shopping, etc).



{Figure 8 of Jovanovic & Rousseau, 2003, "General Purpose Technologies"}

Figure 4.2. Diffusion of a major GPT (in %) as a function of the years following its arrival: electrification [dashed]; and, information technology [solid]. The ordinate indicates the "*cumulative percentage of households that obtained electric service and that owned a personal computer in each year following the 'arrival' of the GPT*" {Jovanovic & Rousseau, 2003}.

Note that IT itself is not a technology in the sense that we have considered technology in this paper (e.g., vacuum tubes, transistors, integrated circuits, thin film deposition, photolithography, magneto-resistance, giant magneto-resistance, etc). This technology actually involves an array of technologies, some of those just mentioned, as well as fiber optics, lasers, RF devices, software, displays, etc. IT, however, because of its broad impact as a GPT, has a profound influence not just on society but also on the scientific and technical disciplines that support its advancement or that desire to do so. This GPT precipitates and “floats” other technologies in its energetic and nonlinearly steepening swell. The infrastructure that is put in place as a result of the GDP diffusion shown in Fig. 4.2 enables rapid insertion of new innovations that displace already-established technologies.

Data like that shown in Fig. 4.2 are often fitted with Sigmoid functions (S-curves). A simple sigmoid function fit to the data shown in Fig. 4.2 suggests that IT will reach the 80 % diffusion level about 36 years after its introduction as compared with 45 years for electrification, and that this 80 % level of diffusion will be reached by 2007.

At 80 % diffusion, IT will likely be viewed as a routine utility much as electrification was when it achieved a comparable level of diffusion into society. By such a time, the present societal focus on IT diffusion into an expanding market, which is integral to the diffusion period of a major GPT, will in all likelihood begin to diminish. The introduction of additional functional capabilities enabled by this new GPT infrastructure will continue for some time. However, next steps are murky with respect to the next GPT: the overwhelming emphasis on rapid technology exploitation and product development during this IT GPT growth period may have seriously impacted the long-term investments needed to bring new technologies out of the prospecting phase and to a point where they can be mined effectively. To address the central issue of the U.S. global technological positioning and national security in a looming post-IT GPT era, we now turn to a discussion of S&T investments and associated governance issues.

V. U.S. S&T Investments: Financial and Human

The entrepreneurial behavior that has resulted in a focusing of funding around the IT GPT, so as to maximize the number of innovations produced, has significant implications regarding the funding of R&D. In this section we consider implications of the recent U.S. S&T fiscal and human capital investment history in light of the developing international situation.

i. The Funding of S&T

The total amount of funding that the United States is prepared to put into R&D is an approximately steady percentage of the gross domestic product (GDP). Figure 5.1 {NSF3, 2002} shows that, after World War II, R&D grew as a percentage of GDP until it reached about 3 % in about 1965. Since then it has oscillated about a value of 2.5 %. It is probably impossible to predict from first principles the percentage of GDP that the United States can spend on R&D. It may be, however, that the Nation has determined the answer experimentally, and that number is 2.5%.

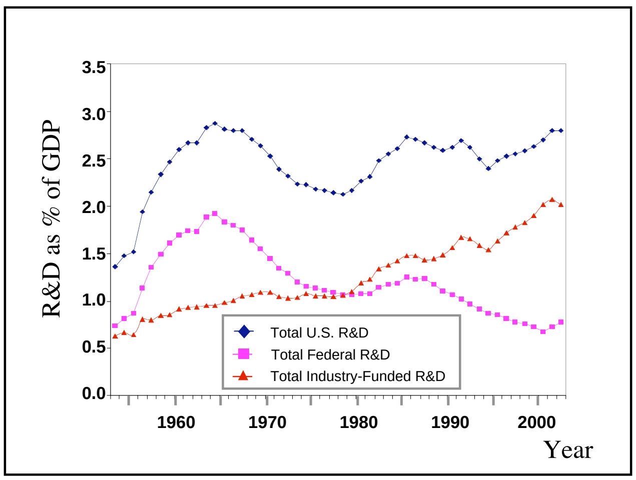


Figure 5.1 R&D expenditure as % GDP: 1953 - 2002.

For at least the past 20 years the total number of scientists and engineers in the country has tended to track the GDP. This appears to be reasonable. It also seems reasonable that the total amount of R&D that a nation invests will be related to the size of the economy (i.e., GDP) that it is intended to support. Some fraction f_m of the R&D activity represented by Fig. 5.1 is mining activity and some fraction f_p is prospecting activity so that $f_m + f_p = 1$. When a major GPT such as IT arrives it will grow for some period faster than the GDP and will have a growing requirement for scientists and engineers to work in the mining phase of the GPT since that is where the jobs and profits are. If the R&D funding is, overall, a fixed percentage of GDP, and the R&D workforce tracks the GDP, then the conservation of the number of scientists and engineers requires that f_m must increase and f_p must decrease while the GPT is in a rapid growth phase. This has the beneficial effect of stimulating the rate of innovation in the mining phase related to the GPT, since the rate of innovations emerging from a particular GPT is quite likely proportional to the number of scientists and engineers working on matters related to the mining phase of that GPT. However, since innovation in the prospecting phase is also likely to be proportional to the number of scientists and engineers working there, this distribution results in less innovation in the science and technology prospecting planes when measured relative to the GDP. This raises the question of how funding of R&D should be determined.

The field of endogenously vice exogenously determined technology growth is one of active research and therefore one cannot draw final quantitative conclusions. However, the purpose of the discussion in this work is not to provide a quantitative description of endogenously determined R&D investments but rather to investigate the potential for volatility of R&D investments determined this way. Of particular concern is how endogenously driven funding affects personnel and therefore progress in what we have called the prospecting phase of R&D. We have shown in Sec. IV*ii* above that this phase has long periods of latency while the enablers for future mining opportunities are falling into place. Such activities, which may take 1-2 decades and require focused and individual or group research "puzzle-solving" effort, need sustained investments, and thus do not do well in volatile investment environments. The mining phase would seem to be much better suited for endogenously determined R&D investments than is the prospecting phase. It is only in the mining phase that the output of R&D investments is contributing to economic growth and producing profits in the near term. There is no short-term profitability from investments in the prospecting phase. In the long term, of course, a few of the R&D investments in the prospecting phase will pay off handsomely, but most of them will not. Unfortunately, it is not possible to determine precisely from where in prospecting space the long-term future breakthroughs and GPTs will emerge. This argues for an exogenous approach to funding the activities in the prospecting phase. Funding in the prospecting phase is based on the historical experience that every so often singularities of scientific and technical change emerge and transform society and the military. It takes deep pockets, and the ability and interest to take a long-term perspective, to sponsor such activities. Since private entities are increasingly driven to maximize profits, an endogenously determined investment strategy would seem to be most reasonable for investing their R&D funds. That leaves governments and perhaps a few very large companies to provide for the exogenous investments needed to retain a viable prospecting phase of R&D. Unfortunately many governments, including the U.S. government, are increasingly turning to endogenous thinking for R&D investments. This leads to placing a metric on an R&D proposal where the value is the functional capability produced divided by the funds invested. As we discussed earlier, functional capability in the prospecting phase is often zero for many years. Such a metric, therefore will drive the prospecting phase investments toward shorter range undertakings which would be better supported out of the mining phase. What really happens in such cases is that mining phase R&D consumes the prospecting phase, thereby jeopardizing the long-term future. It is important to note that many of the capabilities that we now take for granted—for example radar, GPS, solid state electronics, GMR, etc.—would likely not have been funded in their early prospecting phase had they been subject to the above endogenous metric. The characteristic times for the prospecting phase and the mining phase are simply determined by different factors, and for success, the funding of these phases should recognize this.

In considering the U.S. Government's role in supporting long term R&D it is helpful to ask whether or not there is any evidence that the government's investments in long-term R&D have had any measurable influence on the economy. A recent study supported by the National Science Foundation provides some insight in this regard. The study examined the sources of the research papers cited by U.S. patents issued in the years 1987 and 1988 and 1993 and 1994. The approximately 79,000 patents considered referenced research papers from 430 institutions. The four most heavily cited categories were physics, chemistry, engineering and biomedicine, areas that have historically been very significant to DoD; e.g., as noted by the Senate Armed Services Committee in May, 2003 "…the recent display of the armed forces' technological advantages, such as precision weaponry, unmanned systems, smart munitions and increased situational awareness…stand on the shoulders of decades of investment in core scientific disciplines such as chemistry, physics, materials research and information technology" {SASC, 2003}. Table 5.1 shows in rank order the 10 institutions most cited in the fields of Physics and Engineering and Technology {Narin et al, 1997}.

Table 5.1. Linkage Between U.S. Scientific Research and Patents.

Top Ten U.S. Institutions in Rank Order

Physics Papers	Engineering & Tech. Papers
1. AT&T Bell Labs	1. AT&T Bell Labs
2. IBM Corp.	2. IBM Corp.
3. Stanford University*	3. Univ. CA Berkeley*
4. Bellcore	4. MIT*
5. NRL*	5. Stanford Univ.*
6. Lincoln Labs*	6. General Electric Co.
7. MIT*	7. Tex. Inst. Co.
8. Univ. of Illinois*	8. NRL*
9. UC, Santa Barbara*	9. N. Carolina State Univ.*
10. Cornell Univ.*	10. Bellcore

(*) = clearly government funded research.

[1]Institutional research papers cited in U.S. patents; ('87, '88, '93,' 94 patents surveyed, over 430 research institutions cited).

This table makes several interesting points. The first is that AT&T Bell Labs shows up as No. 1 in both categories. This illustrates the huge impact that AT&T Bell Labs had even through the mid 1990's. It is noteworthy that AT&T's historical policy for funding its research program was to treat it as a public good and make it widely available to the economy at large. In the parlance of the economic models this was exogenous funding of R&D. It is generally agreed that the age of U.S. industry playing such a role is now a thing of the past (see {Friedman, 2005}).

The second noteworthy observation from the table is that at least 70 % of the institutions listed under Physics and at least 50 % of the institutions listed under Engineering and

Technology were most certainly supported by government funds. The papers referenced were quite likely published prior to 1990. While the pressure was building by that time to shorten the time horizon of government-funded research the investment was still determined largely by seeking long-term public good. It seems reasonable to assume these investments to be exogenous in the economic sense. It is not clear that such an assumption would be valid for today's government funded research. This development begs the question as to who is looking out for the long-term public good? The question is important because, as the present study shows, while today's technologies can be exploited faster both by increasing near-term investment, and due to the diffusion of the electronics GPT, they all offer only a finite number of financially viable innovations. Eventually they all mature. The faster money is provided the faster they mature. When this happens what technologies will be ready to take their place? These replacement technologies will likely emerge from the prospecting phase. That, however, involves undertaking a long term, low probability of success program of prospecting. If the scientists who would be prospectors have been cultivated as, and have turned into, miners instead, then the future is very worrisome indeed.

The third noteworthy observation from Table 5.1 is that two DoD laboratories (Lincoln Laboratory and the U.S. Naval Research Laboratory) are among the top 10 contributors to the research underlying U.S. patent productivity involving physics and engineering papers. This speaks well of the DoD funded research programs at least through the mid-1990s, and contradicts the oft-made assertion that DoD no longer influences the direction of science and technology.

The exogenous versus endogenous funding of S&T is especially interesting in the case of the DoD. In general, it seems that the nation would be wise to assign the responsibility to industry for the endogenous funding of R&D with the purpose of maximizing profits and to assign the responsibility to the government for exogenously funded long-term R&D with the objective of having future science and technology innovations available for exploitation when they are needed. If the government does not take this responsibility then it will not get done. In the case of national defense, Article 2 of the U.S. Constitution makes clear that this is totally the government's responsibility. Science and technology has always played an important role in warfare and national defense. It is DoD's responsibility to ensure that the science and technology needed for national defense is available when it is needed and that there are no surprises in this regard. DoD is, therefore, confronted with the full spectrum of R&D issues. It must undertake to exploit technology that is in or is entering the mining phase so as to equip the forces in the near term, yet it must also undertake to provide the prospecting necessary to ensure the availability of scientific and technical innovations needed for future defense and warfare. DoD must, therefore, manage two funding allocation processes. For the mining phase undertaking it would seem that a process that determines R&D funding endogenously would be appropriate. However, a process that maximizes profits such as is done in the economic models would not seem to be desirable (although some would argue with this conclusion). It would seem that optimizing military effectiveness subject to funding constraints, political constraints, manpower constraints and so forth would be appropriate. For the prospecting phase, DoD has a more significant problem than others.

The prospecting phase is unpredictable: a breakthrough that could affect DoD can come from almost anywhere in prospecting space. It is, therefore, difficult for DoD to narrowly specialize, yet DoD cannot afford to fund research in all areas. DoD has its own S&T conundrum. The resolution of this conundrum lies in the scientific and technical competence of the DoD's own work force. This is not something that DoD can ignore under the guise of outsourcing. This involves the government's brain trust and is truly a case of if you want it done right then do it yourself. We will consider this further in a discussion of staffing of scientific and technical organizations, below.

ii. Staffing for Innovation

The consequences of the rate-limiting nature of human creative timescales is clear regarding staffing requirements for organizations that produce innovations or are dependent on innovations to meet their missions. In the prospecting phase it is not possible to make precise predictions on where in the prospecting planes important innovation will emerge.

At any point in time one can, of course, target general areas that might be promising. For example, today nanotechnology is believed to be promising for future innovations. The reason that nanotechnology is promising is that as a result of progress made in fabrication technologies to support the semiconductor industry and other developments such as tunneling tip microscopes, functional capability has emerged that allows serious technological efforts to be conducted at scales approaching nanometers. This allows one to fabricate structures and devices at the nanometer scale. The problem remains for the prospectors to identify viable material systems and to invent and test various device concepts. This will occur on the timescales characteristic of the prospecting phase. To accomplish this, a large number of disciplines need to be brought to bear if progress is to be made. These include among others: physics, chemistry, biology, material sciences, electronics, information technology, metrology, health and safety, and so forth. The buzz-word "nanotechnology" is useful for raising interest and attracting funding. This increases the number of prospectors, which increases the probability of discovery and innovation, but it is still a probabilistic undertaking.

Organizations that are in the business of actually making things or actually using things need to staff their organizations with this in mind. Since information technology has not reached the point where it can or should replace human creativity it is necessary to actually engage the relevant prospecting scientific communities. The most effective way to do that is to have staff who are card-carrying members of those communities. This is necessary in order to understand what the state of the art is and to have access to it. If one is not a card-carrying member one gets no respect in such communities, and therefore no serious entry. Furthermore, since science and technology is a global undertaking (and becoming more so) it is essential to engage on a global scale. This is true for DoD as well as commercial businesses.

The U.S. is becoming a relatively smaller player in this global enterprise. It is not possible for any one organization, including DoD, to conduct science and technology efforts in all areas of this global science and technology undertaking. It is possible, however, for large organizations like DoD to employ a large enough science and engineering work force to gain entry into most areas in prospecting space. This will provide the best window on what is going on out there and for evaluating its importance. The strategy should be to have a workforce that has the stature to be welcome and the broad competence (scientific and military) to recognize something that is important to DoD when it sees it. This should be DoD's science and engineering brain trust.

There is considerable literature regarding the staffing requirements needed to accomplish the above. Much work was done in this regard during the 1970's and 1980's. One such study by Roberts and Fusfeld {Roberts & Fusfeld, 1981} identified five key staff requirements for technologically innovative organizations as follows:

- The creative scientist or engineer. The source of creativity within the organization.
- The entrepreneur, who pushes the technical idea forward in the organization.
- The project manager, who can focus upon the specifics of the new development and indicate which aspects will go forward.
- The sponsor. The senior individual who provides coaching, backup, and large skirts behind which entrepreneurs and creative scientists can hide. His role is that of protector and advocate.
- The gatekeeper, who brings essential (technical or market) information into the technical organization.

These staff requirements remain valid today although one might make explicit the need for information technology expertise to support the staff requirements. It is interesting that the DOD, which at one time had a staff that measured very well against Roberts requirements, has allowed this staff to dwindle over the years by choosing to outsource increasingly those science and technology functions that were traditionally done in-house {Marshall, Coffey et al., 2004}. This essentially eliminates the scientific and technical brain trust that understands DoD problems within the context of science and technology. This results in the DoD itself becoming increasingly less technically competent and innovative, losing its card-carrying membership in the scientific and technical communities upon which its future depends thereby becoming more vulnerable to scientific and technical surprise, becoming less able to even understand the complex scientific and technical issues of modern warfare and therefore vulnerable to scientific and technical hucksterism by its contractors, and finally unable to even manage and exercise stewardship over those contractors. This outcome is not good news for anybody except our potential adversaries.

iii. Implications

The long-term consequences of the above are troubling for today's top tier nations. History supports the conclusion that there are only a finite number of financially viable innovations within a given GPT. Indeed, most economic models would seem to explicitly employ this assumption. This finitude results in a behavior for innovation similar to that predicted by Eq. (5), which predicts functional capability for a given technology. The GPT attracts funding (at the expense of other R&D activity), which increases the rate of innovation, which attracts more funding (at the expense of other R&D activities), which further increases the rate of innovation for the particular GPT, and so forth. This results in an ever-faster depletion of the remaining innovations until there is nothing left, and leads to a nonlinear rollover in the innovation rate, following which the process stops. During the decades that this has been taking place, the number of prospectors has declined and those remaining have been strongly influenced by the major GPT dynamics.

As was described in the sections above, there is a time delay between the emergence of the prospecting phase and the onset of the mining phase for the introduction of a new technology. This time delay is typically 15-20 years. Furthermore, there is no way to predict where in prospecting space the next major technology will emerge. One increases the probability of finding something by increasing the number of prospectors. However, since economic forces of the past few decades have acted to reduce the number of prospectors (relative to GDP) in order to increase as much as possible the productivity in the mining phase, the rate or likelihood of innovation in prospecting space has also been reduced (relative to GDP). *This is the core of the S&T innovation conundrum.* Today's technologies are indeed arriving at an ever more rapid rate, but they may be doing so at the expense of future potential. "In short, we are eating our proverbial seed corn."{Welch et al., 2005}

The developing international situation is driving much of the jeopardy (perceived and real) associated with the S&T innovation conundrum. The U.S. response to these developments will determine its long-term economic and military strength. We will consider briefly some of these developments as they relate to the innovation conundrum. It is always risky to extrapolate today's trends for the purpose of making long-term projections. However, we will do so in order to obtain a projection of what the world might look like in 25 years if today's trends prevail.

For consistency we use data provided by the World Resources Institute {earthtrends}. Other databases provide somewhat different absolute values of GDP but the trends are similar. We are interested principally in the trends. In 2000, the United States was responsible for about 26 % of the world's GDP, Europe was responsible for about 33 %, and Asia (Japan, China, India, Korea and Taiwan) was responsible for about 23 %. The trends in GDP suggest that by 2030 the United States will be responsible for about 26 % of the world's GDP, Europe will be responsible for about 24 % and Asia will be responsible for about 33 %. It seems reasonable to assume that Asia (especially China and India) will use its newfound wealth to produce an innovation system modeled after the successful U.S. system. That means, among other things, Asia will invest substantially in research facilities and in the production of scientific and technical human capital.

As discussed in Sec. IV*i*, Romer's work notes the importance of human capital to long-term economic growth. It is therefore, interesting to examine the current trends in human capital production. Figure 5.2a provides the time history for the annual bachelor degrees in science and engineering granted to U.S. citizens by U.S. educational institutions and similar degrees granted to Asians by Asian institutions for the period 1975 through 1998. Figure 5.2b provides the same information for doctoral degrees of the period 1986 to 1998 {NSF3, 2002}.

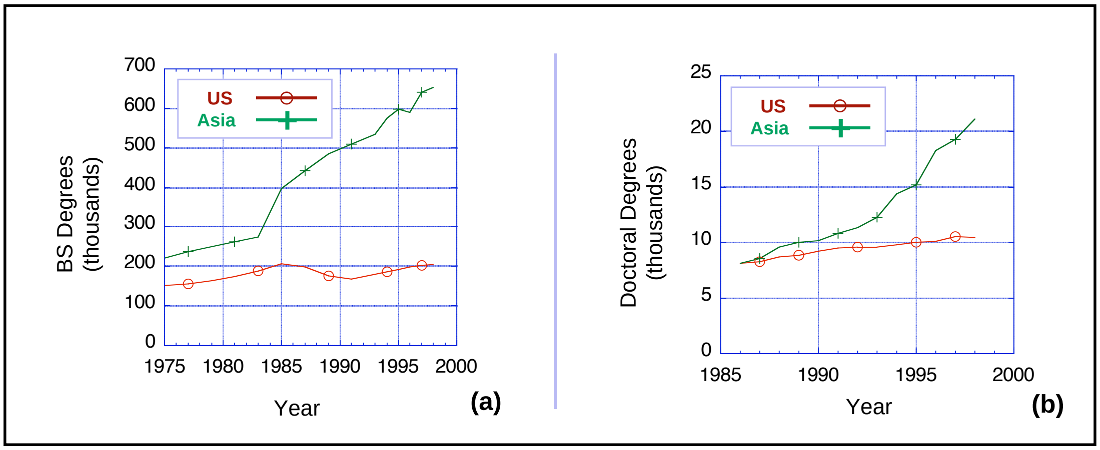


Figure 5.2. Time histories of degrees in science and engineering granted to U.S. citizens by U.S. educational institutions and similar degrees granted to Asians by Asian institutions for: (a) annual bachelor degrees; and, (b) annual doctoral degrees.

Assume that all current scientists and engineers will retire within the next 25 years. Assume also that the production rates will remain constant over this time. In this situation, by the year 2030 the United States will have produced about 3 million bachelor level scientists and engineers and 235,000 PhD level scientists and engineers. During the same timeframe Asia will produce 16 million bachelor level scientists and engineers and 528,000 PhD level scientists and engineers. It is not clear how many non U.S. citizen scientists and engineers will be permanently working in the United States at that time. Nevertheless, it is clear from the above numbers of a significant shift in knowledge generation towards Asia will take place over the next 25 years. It is also clear that the United States, no matter what action it takes, cannot produce scientists and engineers at rates that compete with Asia over the next 25 years. Asia will have perhaps five times the number of scientists and engineers as does the United States by 2030. It also appears that Asia will develop a GDP that is adequate to support a first-class research and development infrastructure.

It seems inevitable, therefore, that Asia will within the next decades surpass the United States in the rate of knowledge production thereby positioning itself to become the dominant player in R&D prospecting space. For the past 100 years of the United States has been the dominant player in R&D prospecting space and has been remarkably adept at moving important prospecting innovations into highly successful functional capabilities. International developments over the next 25 years will challenge if not overturn this U.S. dominance. How this turns out will depend on the U.S. strategy for dealing with the situation.

The U.S. must develop a realistic strategy for coping with these potential developments. The key word in the last sentence is "realistic." Much of what we hear does not strike us as being realistic or up to the challenge that faces us. A strategy that states we will simply innovate our way out of this problem sounds nice but strikes us as being without content—what does it mean? A strategy based upon a belief that we are more innovative than others and will therefore win strikes as being silly and dangerously arrogant.

VI. Recommendations

The recommendations from this study fall into two categories: "governance of the nation's R&D investment" and "knowledge generation and human capital." Each category is discussed below.

i. Governance of the Nation's R&D Investment

The study suggests that the United States revisit its governance for R&D investment and construct a new governance that recognizes the mutual dependencies but distinctly different natures of the prospecting phase of R&D innovation and the mining phase of R&D innovation. Different governance is required for each of these phases.

The private sector is very effective at optimizing the short term R&D investment in the mining phase in situations where market forces dominate. The private sector should therefore be responsible for this governance. However, the very forces that make the private sector so effective at the governance of mining phase R&D also make it ineffective for the governance of the long-term R&D associated with the prospecting phase, where profitability is very uncertain and can only be measured in hindsight after many years of sustained investment.

The proper role for the government in R&D is to ensure the health of the prospecting phase R&D (mix of basic and applied research, and exploratory development) that is crucial for long-term economic growth and military power but is not going to get done by the private sector. This role is so important to the long term economic and military health of the nation that the government must be staffed with the world class scientists and engineers needed to carry out this responsibility. This responsibility cannot be carried out by functionaries or administrators whose jobs are simply to send public moneys to non-governmental entities. We have chosen the term "governance" quite deliberately in this regard. It should be carried out by government employees who are card-carrying members in the scientific and technical communities deemed to be of long-term importance for the nation's future. They must have the respect from these communities that is only earned by peers. The communities must accept the government's scientists and engineers as scientific and technical peers in order for the required long term planning and steadfast direction to occur and so that the required advocacy is in place both within and outside of the government. At one time the federal government was staffed to carry out this function. It is not clear that this is true today, especially in the DOD sector. This deficiency must be remedied. Excuses for not dealing with this matter such as asserting that the government cannot hire or retain the required talent are not acceptable since it is a problem that can be fixed. It is not an overstatement to say that the Nation's long-term economic and military strength may be at stake.

A special situation exists for Defense R&D where the beneficial effects of the free market do not apply due to the small market size and the specialized nature of warfare. In this case the United States Constitution assigns implicitly governance for the full spectrum (prospecting phase and mining phase) of R&D to the federal government. The same principles articulated above apply here.

ii. Knowledge Production and Human Capital

It seems clear that the rate of new scientific and technical knowledge generation is related to the number of scientists and engineers who are working on new knowledge production. Global demographics suggest that this could result in the shifting of the center for new scientific and technical knowledge generation from the West to Asia over the next 25 years. If, in the long term, knowledge determines economic growth and military power then such a shift has profound implications for U.S strategy. In this regard, it would be in the long term interests of the United States to advocate a position that knowledge generation should be viewed as a public good available to all countries so that all countries, including the United States, can benefit from the global production of scientific and technical knowledge. This must be done within reasonable constraints associated with intellectual property and national security considerations. The knowledge exploitation phase is where controls will be most effective rather than the knowledge generation phase. Furthermore, the United States should increase its production of scientists and engineers so that it can more easily interact with the global S&T community. This will require that the U.S. S&T communities continue to be viewed as leading players in science and technology, otherwise they will not have entry into the larger community. Efforts to provide incentives to increase the number of scientists and engineers must be approached carefully. Romer provides a viewpoint of some of the subtleties inherent in this topic {Romer, 2000}. Finally, the United States should move to attract as many foreign scientists and engineers as possible to become U.S. citizens or permanent residents so as to maximize the U.S. ownership of the global scientific and technical community and, thereby, of its knowledge generation. Emigrant scientists and engineers have historically made huge contributions to U.S. economic and military strength.

VII. Conclusions

The major conclusion of this study is that we have lost sight of some key realities. We have become so mesmerized by our enormously successful exploitation of past S&T breakthroughs that we have forgotten how they happened in the first place. Since society is primarily interested in the creation of functional capability (e.g., computing power) this memory lapse becomes problematic with respect to maintaining a pipeline of future breakthroughs.

The rapid advances in electronics and computer products over the past 50 years have created a general impression of continuous scientific breakthroughs. In reality, the breakthrough S&T innovations for electronics and computers took place in the 1940's and 1950's. The subsequent rapid advances in functional capability were the result of a brilliant and enormously successful program to exploit those early breakthroughs. An unfortunate byproduct of this success was the impression that these rapid advances in functional capability also represented the time scale for breakthroughs in science and technology.

An examination of the histories of a number of major S&T innovations covering the past 100 years indicates that today a breakthrough innovation takes 15-20 years to progress through the early phases, just as it did 100 years ago. There is often no functional capability produced during this early phase. Our fixation with exploitation and near term profits is incompatible with the realities of these time scales. This is problematic due to its effect on investment strategies.

There are two distinct phases in S&T innovation. For a successful innovation, once the underlying S&T is in hand as a result of the early phase work, a second phase can be undertaken where rapid technical progress resulting in significant new functional capability is possible with the application of adequate financial and human capital. The time scales for progress are much faster than in the early phase and are based on the potential of the technology and the resources applied.

The characteristics of these two separate phases are captured by the descriptors "Prospecting" and "Mining" respectively. There is a dynamical relationship between the prospecting phase and the mining phase. The long-term health of one depends on the health of the other. Both phases also involve conducting basic and applied research as well as exploratory development (but with a different mix). However, the two phases are fundamentally different and require different governance. Unfortunately we seem to have forgotten this thereby contributing to the S&T innovation conundrum.

An essential aspect of governance is the allocation of resources (e.g., people and funding). Economists, understanding the important role that technology plays in economic growth, have begun developing theories regarding the impact on economic growth when R&D investments are determined so as to maximize profits. This is referred to as an endogenous investment strategy. These theories are helpful in discussing how economic conditions combined with an endogenous investment strategy for R&D affect the scientific and technical talent pool and the generation of knowledge as well as economic growth. They also shed light on how knowledge affects long-term economic growth. It seems clear that a solely endogenous approach to determining R&D investments results in too little long-term research being funded. Talent and resources gravitate to the mining phase at the expense of the prospecting phase and at the expense of the knowledge generation needed to sustain economic growth in the long term. In the short term, however, the private sector, using profit maximization techniques, is extremely effective at introducing innovations that exploit science and technology developments thereby maintaining the U.S. competitive advantage. Part of the S&T conundrum is related to the balancing of these conflicting attributes. Proper balancing of these competing outcomes should be a byproduct of the separate governances required for the prospecting and mining phases of innovation. This suggests a national imperative for an exogenous (i.e., not determined by near term economic and profit maximizing considerations) determination of the R&D investment for the prospecting phase.

An examination of R&D funding data since the 1950's suggests that the United States is comfortable with a steady state investment in R&D of about 2.5 % of GDP. The total investment seems to have oscillated about this value for 50 years. It seems reasonable that a nation's need and ability to support R&D should be proportional to the economy that the R&D is intended to support. The relationship to GDP, therefore, is not surprising.

The federal share of R&D (measured relative to GDP) has mostly declined since 1965 while the industry share has mostly increased over the same period. Since the industry investment in R&D is (and should be) mostly endogenous, this has raised concern about the long-term investment in the prospecting phase of R&D. This concern is responsible for part of the conundrum. The severe competitive environment created by globalization has left the federal government as one of the few entities in a position to take responsibility for the long-term knowledge necessary for long-term economic growth. Unfortunately the federal government has also moved towards an endogenous approach for the determination of its R&D investment. This seems to be driven by the reasonable objective of justifying the government's R&D investment. However, it is interesting to note that many of the S&T innovations upon which today's societies are based would never have been funded in their early phase had they been subject to a purely endogenously determined R&D investment strategy. It may be that the very attempt to measure the output on too short a time frame is producing a program that operates on too short a time frame, hence eliminating its true value.

In the long term, knowledge is a fundamental pillar of economic growth and military power. Therefore, understanding how knowledge grows is important to proper governance. Here again a reality check is needed. It seems that we have concluded that, in today's world, knowledge grows at an ever-increasing rate and actually feeds on itself. This can only be possible if individual scientists and engineers are themselves producing knowledge at an ever-increasing rate. However, an examination of patent and publication data for the past 50 years indicates that the rate of new knowledge production at the level of the individual scientist and engineer has not increased and may have actually declined. This reality should place significant constraints on models that purport to predict the rate of growth of scientific and technical knowledge. Failure to do so results in unrealistic expectations and further contributes to the conundrum. Knowledge cannot grow far more rapidly than does the scientific and technical workforce without any increase in the rate of knowledge generation at the level of the individual scientist and engineer.

This part of the conundrum may be related to the fact that the enormous advance in automation of routine functions such as data collection and data analysis has greatly improved productivity with respect to these routine tasks. However, such improvements should not be confused with increases in the rate of knowledge generation.

A related contributor may be that new functional capabilities such as computers, advanced fabrication technology and increasingly sophisticated analytical instruments have enabled scientists and engineers to work in new regimes (such as at the nano scale). The ability to now work in these new regimes is not equivalent to an increase in the rate of knowledge production. It is human beings who produce new knowledge. New knowledge production for these regimes is still paced by the rate limiting step of human cognition and understanding that has evolved slowly over many millennium. This rate limiting step is analogous to those that occur in many dynamical systems such as chemically reactive systems. Speeding up the flow of information and data does not automatically translate into new knowledge production. Indeed it can divert the attention of scientists and engineers resulting in the opposite effect.

Part of the conundrum is also related to the differing nature of technologies. There are certain technologies, called general purpose technologies (GPTs), that are characterized by broad applicability over many segments of society and act as enablers for societal development. Examples of such technologies are electrification technologies and information technologies. The infrastructure that accompanies the diffusion of GPTs throughout societies create opportunities for further technology development and innovation and focus financial and human capital around these opportunities. This has the positive effect of accelerating certain types of innovation (mostly in the mining phase) and the long-term potentially negative effect of reducing other types of innovation (mostly in the prospecting phase) especially those that are not closely related to the GPT.

Another major contributor to the conundrum relates to global trends that have emerged over the past decade. Recent economic theory suggests that long-term economic growth is a result of the scientific and technical human capital involved in knowledge generation. In this regard, trends over the past decade suggest that by 2030 Asia will have a scientific and technical workforce that may be as much as five times the size of the U.S. scientific and technical workforce. It will be very difficult for the United States to take steps that will allow it to match the current Asia rates for the production of scientists and engineers. The global trends also suggest that Asia could surpass the United States in GDP by 2030.

It is reasonable to assume that Asia will use its new wealth and scientific and technical human capital to create a world class R&D infrastructure probably modeled after the highly successful U.S. R&D infrastructure. This could position Asia to become the global center for knowledge creation by the middle of the 21st Century. If, in the long term, knowledge production is truly the foundation of economic growth and military power then these global trends confront the United States with profound challenges. The United States needs to move quickly and correctly to meet these challenges.

Market forces and competition within the United States have been very effective at focusing scientific and engineering work so as to exploit scientific and technical discoveries, inventions and innovations. This has been responsible for much of the U. S. economic growth over the past several decades. However, this focus on exploitation has led to concerns regarding the generation of knowledge required for continued economic growth in the longer term when the current crop of technologies have run their courses.

These concerns are exacerbated by the realization that globalization and demographics will likely “flatten” the world over the next 25 years in terms of economic competitiveness. A likely consequence of this is that the United States will move from the position of “Chairman of the Board” for the global economy to a partner in the global economy. This topic is explored further in a recent book by Thomas Friedman {Friedman, 2005}.

In January 2001, the Hart-Rudman Commission on National Security stated that "the inadequacies of our systems of research and education pose a greater threat to U.S. national security over the next quarter century than any potential conventional war that we might imagine. American national leadership must understand these deficiencies as threats to national security. If we do not invest heavily and wisely in rebuilding these two core strengths, America will be incapable of maintaining its global position long into the 21st century." {Hart-Rudman, 2001}. This remains a valid summary of our present situation.

VIII. References

Allison DK, 1981, "New Eye for the Navy: The Origin of Radar at the Naval Research Laboratory," NRL Report 8466, http://torpedo.nrl.navy.mil/tu/ps/doc.html?dsn=32770402800973138054.

ASCR, 2005, U.S. DOE SC ASCR Strategic Plan, "Advanced Scientific Computing Research: Delivering Computing for the Frontiers of Science," page 38, http://www.sc.doe.gov/ascr/.

Australia Prize, 1998, "1998 Australia Prize: Professor Sir Alec Jeffreys (UK)," https://sciencegrants.dest.gov.au/SciencePrize/Pages/Doc.aspx?name=previous_winners/Aust1998Jeffreys.htm.

Baibich MN, Broto JM, Fert A, Nguyen Van Dau AF, Petroff F, Eitenne P, Creuzet PG, Friedric A, & Chazelas J, 1988, "Giant Magnetoresistance of (001) Fe/(001)Cr Magnetic Superlattices," *Phys.Rev.Lett,,* **61**, 2472.

Bello, PA, (Applied Research Laboratory of Sylvania), 1963, "Characterization of Randomly Time-Variant Linear Channels," *IEEE Trans. on Comm. Systems,* **CS-11,** 360.

Binasch G, Grunberg P, Saurenbach F, & Zinn W, 1989, "Enhanced Magnetoresistance in Layered Magnetic Structures with Antiferromagnetic Interlayer Exchange," *Phys.Rev.B,,* **39**, 4828.

Borsuk G, & Coffey T, 2003, "Moore's Law: A Department of Defense Perspective," *Defense Horizons,* **30,** http://www.ndu.edu/inss/DefHor/DH30/DH30.pdf.

Branscomb L, 2004, "Where Do High Tech Commercial Innovations Come From?" http://www.law.duke.edu/journals/dltr/articles/2004dltr0005.html.

Bresnakan T, & Trajtenberg M, 1995, "General Purpose Technologies: Engines of Economic Growth," *Journal of Economics,* **65,** 83.

cnrs, 2003, http://www.cnrs.fr/cw/en/pres/compress/Albert_Fert.htm.

computinghistorymuseum, History of Computing: Evolution of the Major Computer Storage Devices; www.computinghistorymuseum.org/teaching/papers/research/StorageDevices-Zeyt.

Crace J, 2004, "Alec Jeffreys: Pennies from Heaven," The Guardian [Education Guardian Weekly] (Sept. 14, 2004); http://education.guardian.co.uk/egweekly/story/0,5500,1303418,00.html.

Dowling B, 2003, "Is the Long Wave Getting Shorter?" WP 08-2003; http://www.jims.cam.ac.uk/research/working_papers/abstract_03/wp0308.pdf.

earthtrends, http://earthtrends.wri.org/.

Economist, 2001, Case Study (from the Economist Print Edition Dec 8, 2001), "The Art of the Quantum Leap," Economist **361***(8251)* Special Section p 19; also Science Technology Quarterly, 2005, http://economist.com/science/tq/displayStory.cfm?Story_ID=885080.

fermat-s-last-theorem, http://www.answers.com/topic/fermat-s-last-theorem.

Fert A, 2000, private communication.

Friedman TL, 2005, "The World is Flat: A Brief History of the Twenty-first Century," Farrar, Straus and Giroux: New York, 488 pages, ISBN: 0374292884.

Grunberg P, 2001, “Layered Magnetic Structures: History, Highlights, Applications,” *Physics Today,* 54, 31; http://www.physicstoday.org/pt/vol-54/iss-5/p31.html.

Hart-Rudman, 2001, Road Map for National Security: Imperative for Change. Phase III Report of the U.S. Commission on National Security/ 21st Century, Jan 2001 p. *ix*; mil.wa.gov/hls/documents/PhaseIIIFR.PDF.

Helpman E, 1998, “General Purpose Technologies and Economic Growth,” MIT Press.

Hertz H, 1889, “Die Krafte electrischer Schwingungen behandelt nach der Maxwell’schen Theorie” *Annalen der Physik,* 1.

Hindsight, 1969, “Project Hindsight,” Office of the Director Defense Research and Engineering, DTIC N. AD495905.

Hollman M, 2001, “Chriasian Huelsmeyer, the Inventor,” http://www.radarworld.org/huelsmeyer.html.

Howeth LS, 1963, “History of Communications-Electronics in the United States Navy,” Washington, http://earlyradiohistory.us/1963hw.htm.

itrs, 2004, http://public.itrs.net/.

Jeffreys AJ, Wilson V, & Thein SL, 1985, “Hypervariable ‘Minisatellite’ Regions in Human DNA,” *Nature,* **314,** 67.

Jovanovic B, & Rousseau P, 2003, “General Purpose Technologies,” http://www.econ.nyu.edu/user/jovanovi/GPT.pdf.

Julliere M, 1975, “Tunneling Between Ferromagnetic Films,” *Phys.Lett.,* **54A,** 225.

Koestler A, 1964, “The Act of Creation,” The McMillan Company: New York, 751 pages.

Kurzweil R, 1999, “The Age of Spiritual Machines,” Viking Penguin.

Kurzweil-data: http://www.penguinputnam.com/static/packages/us/kurzweil/excerpts/chap1/ch1note19.htm.

Lewis WD, 1960, “Coordinated Broadband Mobile Telephone System,” *IRE Trans*., 43.

Makofske M, 2005, “25 Years of Really Cool Stuff,” Lemelson-MIT Program (January 14, 2005); http://web.mit.edu/newsoffice/2005/top25.html;
also, http://www.cnn.com/2005/TECH/01/03/cnn25.top25.innovations/.

Marshall M, Coffey T, Saalfeld F, & Colwell R, 2004, “The Science and Engineering Workforce and National Security,” Defense Horizons, **39,** http://www.ndu.edu/ctnsp/defense_horizons/DH39.pdf.

Maxwell JC, 1892, “Treatise on Electricity and Magnetism,” Clarendon Press.

Merriam-Webster, 1999, Mirriam-Webster’s Collegiate Dictionary, Merriam-Webster Inc, Springfield, MA.

Mesny R, 1934, “Constantes de temps, durees d’etablissement, decrements,” *L’Onde Electrique,* **13,** 237.

Moore GE, 1965, “Cramming More Components onto Integrated Circuits,” *Electronics,* 38(*8).*

Narin F, Hamilton K, & Olivastro D, 1997, "The Increasing Linkage Between the U.S. Technology and Public Science," *Research Policy,* **26***(3)*, 317.

nobelprize, "The Nobel Prize in Physiology or Medicine 1968," http://nobelprize.org/medicine/laureates/1968/.

NSF1, 2004, http://www.eh.net/hmit/gdp/; http://www.nsf.gov/sbe/srs/seind04/c3/fig03-02.htm.

NSF2, 2004, NSB Science and Engineering Indicators (Figure 5-30); http://www.nsf.gov/sbe/srs/seind04/c5/c5s3.htm.

NSF3, 2002, "Science and Engineering Indicators 2002, Vol 2: Appendix Tables," http://www.nsf.gov/sbe/srs/seind02/pdf/volume2.pdf.

Parkinson BW, Stansell T, Beard R, & Gromov K, 1995, "A History of Satellite Navigation," Navigation: The Journal of the Institute of Navigation, **42***(1)*.

Pasteur L, 1854, [University of Lille].

Philips, http://www.extra.research.philips.com/pressmedia/background/optrec/.

Pohlmann KC, 1992, "The Compact Disc Handbook, 2nd Edition," A-R Editions; also see, http://www.ieee.org/organizations/history_center/sloan/DAR/timeline.html; and, Despain J, http://www.oneoffcd.com/info/historycd.cfm.

Prinz G, 1995, "Spin-Polarized Transport," Physics Today (April, 1995).

privateline, http://www.privateline.com/PCS/history.htm.

Ramsey NF, 1989, "Experiments with Separated Oscillatory Fields and Hydrogen MASERS; Nobel Lecture," http://nobelprize.org/physics/laureates/1989/ramsey-lecture.pdf.

Reed IS, & Solomon G, 1960, "Polynomial Codes Over Certain Finite Fields," *J.SIAM,* **8**, 300; Cipra BA, 1993, "The Ubiquitous Reed-Solomon Codes," *SIAM News,* **26**(1).

Riordan M, & Hoddeson L, 1997, "Crystal Fire—The Birth of the Information Age," W. W. Norton & Co: NY.

Roberts E, & Fusfeld A, 1981, "Staffing the Innovative Technology-Based Organization," *Sloan Mgmt. Review,* 19.

Roessner Carr DR, Feller I, McGeary M, & Newman N, 1998, "The Role of NSF's Support of Engineering in Enabling Technological Innovation: Phase II Final Report to the National Science Foundation," Arlington VA: SRI International.

Romer P, 1990, "Endogenous Technological Change," *Journal of Political Economy,* **98**(5), S71.

Romer P, 2000, "Should the Government Subsidize Supply or Demand in the Market for Scientists and Engineers?" Working Paper 7723; http://www.stanford.edu/~promer/Mkt-forSE.pdf.

Saiki R, Gelfand RD, Stoffel S, Scharff S, Higuchi R, Horn G, Mullis K, & Erlich H, 1988, "Primer-directed Enzymatic Amplification of DNA with a Thermostable DNA Polymerase," *Science,* **239**, 487.

Saiki R, Scharf S, Faloona F, Mullis K, Horn C, Erlich H, & Arnheim N, 1985, "Enzymatic Amplification of b-globin Genomic Sequences and Restriction Site Analyses of Sickle Cell Anemia," *Science,* **230**, 1350.

Sanger F, Nicklen S, & Coulson AR, 1977, "DAN Sequencing with Chain-Terminating Inhibitors," *Proc.Natl.Acad.Sci.,* **74**, 5463.

SASC, 2003, Senate Armed Services Committee in Defense Authorization Committee Report (108-46, May 2003).

Schulte HJ, & Cornell WA, 1960, "Multi-area Mobile Telephone System," *IRE Trans.*, 49.

Schumpeter J, 1942, "Capitalism, Socialism and Democracy," Harper press edition 1975.

Simon H, 1995: in Varian H, "How Much Will Two Bits Be Worth in the Digital Marketplace?" *Scientific American,* **273***(3),* 200.

Solow R, 1956, "A Contribution to the Theory of Economic Growth," *Quarterly Journal of Economics,* **70**, 65.

Southern EM, 1975, "Detection of Specific Sequences among DNA Fragments Separated by Gel Electrophoresis," *J.Mol.Biol.,* ***98(3),*** 503.

Task Force on the Future of American Innovation, 2005, "The Knowledge Economy: is the United States Losing its Competitive Edge? Benchmarks of our Innovation Future," http://www.futureofinnovation.org.

TRACES, 1968, "Technology in Retrospect and Critical Events in Science," prepared for NSF by Illinois Institute of Technology.

Welch LE, Mote CD, Bacioccco AJ, Borsting JR, Deutsch JM, Good ML, Hermann RJ, McGroddy JC, Moore CB, O'Connor JG, Powell RC, Ulaby FT, Wilson BA, & Wilson JE, 2005, "Assessment of Department of Defense Basic Research," p. 23, National Academies Press: Washington, DC, 54 pages, ISBN 0-309-09443-7.

wikipedia, http://en.wikipedia.org/wiki/Logistic_function.

Wilson D, 2003, "Are We Running out of New Ideas? A Look at Patents and R&D," FRBSF Economic Letter 2003-26 (Sept 12, 2003); http://www.frbsf.org/publications/economics/letter/2003/el2003-26.html.

Wyman AR, & White R, 1980, "A Highly Polymorphic Locus in Human DNA," *Proc.Natl.Acad.Sci.USA,* **77**, 6754.

Zhirnov V, Cavin R, Hutchby J, & Bourianoff G, 2003, "Limits to Binary Logic Switch Scaling—A Gedanken Model," *Proc IEEE,* **91,** 1934.