Underinvestment in Public Good Technologies

Gregory Tassey

ABSTRACT. Although underinvestment phenomena are the rationale for government subsidization of research and development (R&D), the concept is poorly defined and its impact is seldom quantified. Conceptually, underinvestment in industrial R&D can take the form of either a wrong amount or a suboptimal composition of R&D investment. In both cases, R&D policy has not adequately modeled the relevant economic phenomena and thus is unable to characterize, explain, and measure the underinvestment. Four factors can cause systematic underinvestment in R&D-intensive industries: complexity, timing, existence of economies of scale and scope, and spillovers. The impacts of these factors vary in intensity over the typical technology life cycle, so government policy responses must be managed dynamically. In addition to understanding the causes of underinvestment in R&D, the magnitude of the deficiency relative to some "optimum" must be estimated to enable a ranking of technology areas with respect to expected net economic benefits from a government subsidy. Project selection criteria must therefore be based on quantitative and qualitative indicators that represent the nature and the magnitude of identified market failures. The major requirement for management of R&D policy therefore is a methodology that regularly assesses long-term expected benefits and risks from current and proposed R&D portfolios. To this end, a three-stage process is proposed to effectively carry out R&D policy analysis. The three stages are (1) identify and explain the causes of the underinvestment, (2) characterize and assess the investment trends and their impacts, and (3) estimate the magnitude of the underinvestment relative to a perceived optimum in terms of its cost to the economy. Only after all three stages of analysis have been completed can the underinvestment pattern be matched with the appropriate policy response.

Key words: R&D innovation, underinvestment, policy

JEL Classification: O3, O2

1. Introduction

Understanding the role of technology in economic growth requires the development and application of microeconomic theory. This is because each technology has unique characteristics, which interact iteratively over the technology's life cycle with unique industry structures and technical infrastructures. The microeconomic character of such an evolutionary model is further enhanced by the fact that the majority of technology investment decisions are made by individual companies and external financing is often supplied by individuals and small venture capital firms, with much of the financing for particular technologies limited to specific geographic regions.

Economists have explored the numerous elements of the microeconomics of technology-based growth, including the determinants of research and development (R&D) spending, project selection, the tradeoffs between R&D costs and time within a portfolio management context, the interactions of this activity with firm size and market structure, and the responses of R&D investment decisions to different financial and regulatory incentives. In addition, the results of such investments have been estimated using several analytical frameworks.

However, no one has surpassed Mansfield *et al.* (1968, 1971, 1982) in the number of these elements analyzed with respect to both investment and results, the level of disaggregation explored, and the collection and use of industry data to enlighten the proposed frameworks. Mansfield *et al.* (1977) have had a particularly pronounced effect on the analysis of the results of innovation. Their work in the late 1970s in which social and private rates of return were estimated for a range of R&D investments in manufacturing technologies followed Griliches' (1958) path breaking work in agriculture and collectively drew attention to the "gap" between social and private rates of return. This gap has spawned several decades of debate over

Stop 1060
National Institute of Standards and Technology
Gaithersburg, MD 20899
E-mail: tassey@nist.gov

Journal of Technology Transfer, 30 1/2, 89–113, 2005

government roles in supporting various amounts and types of R&D [including Mansfield *et al.*, (1982)]. Economists [including Mansfield, (1980, 1991a, b)] subsequently disaggregated R&D into basic research and applied R&D to examine the differences in rates of return between the "pure public good" (basic science) and the target of applied R&D ("proprietary technology").

Mansfield's broad and highly microeconomic analysis has provided an analytical platform for needed progress in assessing the implications of the "gap" for public policy. This paper focuses on barriers to R&D investment that result in suboptimal distributions of R&D investments across different types of R&D and thereby reduce potential long-term economic growth. The point of departure is the proposition that the gap between the social and private rates of return to R&D investments is not automatically an indicator of underinvestment and, at least as important, it is not an indicator only of potential underinvestment in the amount of R&D. Barriers arise that affect the composition of R&D and this class of market failures can have significant long-term negative effects on economic growth. The resulting added complexity to R&D policy analysis requires a more microeconomic approach in the Mansfield tradition.

2. R&D market failure analysis

Firms conduct R&D for two reasons: (1) to develop new products, services, or processes, and (2) to maintain a capability to identify and assimilate technologies from external sources (Cohen and Levinthal, 1989). Governments of industrialized nations obviously consider these activities to be essential for economic growth because they maintain various sets of R&D support policies. Yet, the theoretical and operational frameworks for managing such policies are incomplete.

Economists have consistently estimated the rates of return from R&D to be considerably above those obtainable from other assets. However, for R&D in aggregate, significant rate of return (RoR) differentials for social (industry) versus private (innovator) R&D investments have been found.[1] Moreover, among types of R&D, basic research and long-term, high-risk technology research have yielded particularly high rates of return (Griliches, 1995). The implication of these persistent rate-of-return differentials is the existence of systemic "market failure." However, the interpretation and policy implications of these RoR differentials are poorly defined because the majority of the economics literature has misspecified the knowledge production and innovation processes, resulting in inaccurate R&D policy prescriptions for dealing with the implied underinvestment.

Assessments of private sector underinvestment have been based largely on the neoclassical concept of externalities (Pigou, 1932) and defined as differences between actual investment patterns and an optimum rate of investment. However, this concept of market failure has not been made operational in the sense that specific roles for government can be deduced (Ruffin, 1996). Coase (1960, 1992) made a huge contribution toward resolving the problem by emphasizing that overcoming the difficulties in estimating the amount and nature of underinvestment mechanisms resulting from externalities requires the definition and then the assignment of property rights and associated transaction costs. The institutional (government) role therefore becomes extremely important and also case specific. Consequently, a microeconomic analysis of each specific case is required.

For R&D, analysis of market failure mechanisms and the consequent need for government intervention require accurate models of the unique set of factors determining this category of investment. If the needed models were straightforward, government R&D strategies would be relatively simple to design and manage. However, private investment incentives respond negatively to the public good content of technology and this content is distributed among elements of the typical industrial technology in more complex ways than generally realized. In particular, most industrial technologies (or, more accurately, elements of them) have a quasi-public good character, which complicates and thus inhibits addressing property rights issues, thereby substantially raising transaction costs, as originally identified by Coase.

Moreover, other factors besides property rights are important in determining R&D

investment patterns, such as the relationship between the nature of a technology and the industry and market structures that deliver it. In addition, defining an optimum rate of investment is a particularly challenging problem for R&D policy, but this is a necessary analytical step independent of identifying the factors determining private sector investment. Assessment of the time-cost trade-off is one of the many contributions by Mansfield to understanding the technology life cycle.

In the Coase tradition, R&D policy requires a set of analytical tools that can efficiently identify and assess the characteristics of an industrial technology that affect investment patterns over time. Thus, this paper seeks to provide more structure and rigor to the analysis of underinvestment in R&D. Specifically, the framework developed emphasizes underinvestment in the major elements of the typical industrial technology and examines how investment patterns change over a technology's life cycle relative to an optimum. The focus is on underinvestment relative to the optimal *composition* of R&D, as opposed to underinvestment in the aggregate *amount* of R&D.[2]

Beginning with Arrow (1962a), economists developed explanations of underinvestment in R&D based on the indivisibility and inappropriability of information and the excessive risk involved in its creation. These factors were developed largely as if technical information were a homogeneous entity and therefore so would be the process (R&D) that produces it. Some research has identified two or more distinct elements of technical information.[3] Tassey (1997) characterizes these elements in terms of uniquely different investment incentives, responding to differing public good content. It remains to develop a framework based on the composition of R&D that identifies specific underinvestment phenomena associated with each element, thereby enabling accurate policy analysis.

That is, to identify, characterize, and understand the sources and impacts of market failures with respect to the composition of R&D investment and eventually match underinvestment phenomena with efficient policy response mechanisms, an analytical framework is needed that emphasizes analysis at the technology-element level, where the elements are distinguished by unique investment incentives and patterns. A "technology-element" framework is different from the more traditional "R&D-phase" approach used by Mansfield *et al.* (1971) and a few others, even though the proposed elements are to a degree the outputs of the conventional phases of R&D. The technology element approach is preferred because it is more closely related to corporate R&D investment decision making, a fact which facilitates investment analysis.

To provide this framework, the analysis takes a three-stage approach: (1) use of conceptual models that allow specification of the causes and effects of underinvestment by major technology element; (2) selection of indicators of suboptimal investment patterns and resultant impact trends; and (3) construction of practical quantitative approaches for defining and estimating the economic impacts of the underinvestment. Collectively, these three steps provide an empirically based "cause-and-effect analysis," which becomes the basis for ranking targets for government funding and matching underinvestment mechanisms with the appropriate policy responses. The following sections discuss these three stages that lead up to the policy response decision.

3. Underinvestment in generic technologies

For policy analysis purposes, a conceptual model must be selected that portrays the typical industrial technology in terms of its public and private good elements and thereby focuses policy analysis on the process creating systematic underinvestment. Such a model is also essential for interpreting the results of this analysis for policy makers, and ultimately helping to select the appropriate policy response.

A quasi-public good model of a technology-based industry

A static representation of a typical technology-based industry is shown in Figure 1. This conceptual approach is significantly different from the "black box" model, which has influenced R&D policy for the past 50 years.[4] The black box model regards technology as a homogenous entity and thereby prohibits consideration of the

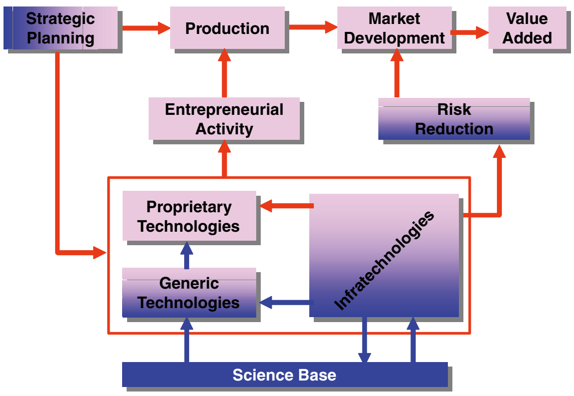

Figure 1. Economic model of a technology-based industry.

existence of distinct elements within the typical industrial technology, each of which responds to distinctly different investment incentives.

In a more realistic model, the typical industrial technology is separated into a set of private and quasi-public elements to reflect the different investment incentives and R&D management practices associated with the provision of each of these elements (Tassey, 1991, 1997). R&D policy is complicated not only by their existence, but also by the fact that these elements vary with respect to their public good content.[5] The specified elements are not arbitrary; rather, they represent distinctly different investment incentives faced by industry and therefore different underinvestment phenomena appear for each element.[6]

This last point is critical because success of a public role in supporting industrial R&D depends on demonstrating both theoretically and empirically that public R&D is a complement rather than a substitute for private R&D. However, disaggregating technology into publicly funded and privately funded components can imply a dichotomous model consisting of a pure public good and a pure private good.[7] Some of the required interaction issues (in particular, complementarity) can be assessed in such a model, but the quasi-public good nature of several common technology elements is obscured and therefore an assessment of investment incentives and subsequent underinvestment behavior is compromised.

In Figure 1, the darker shading indicates the degree of public good content for a technology element. In both the black box and technology element models, funding of basic science is a relatively straightforward policy issue because scientific knowledge is generally regarded as a strong public good. Thus, basic research is largely the responsibility of government to fund. This characterization of science simplifies policy management, with the only issues being the amount of funding and the distribution of this funding across fields of science.[8]

The first step in applying scientific knowledge (i.e., conducting *technology* research) is to prove the concept of the new technology. The end point of such an effort is frequently a laboratory prototype or some other form of "proof of concept."[9] If the resulting fundamental or "generic" technology is deemed to have sufficient potential commercial value, companies assign the substantial follow-on R&D to one of the company's line-of-business units to turn the laboratory prototype into a commercial product. Investment decisions in this phase of the R&D cycle use the more quantitative portfolio management tools of corporate finance, such as net present value and internal RoR metrics. The outcome of applied R&D is represented by the proprietary technology box in Figure 1.

The importance for industrial competitiveness of investment in generic technologies is that this first phase of technology research provides the basis for "next generation" or radically new ("disruptive") products, processes, and services.[10] Developing new technology platforms (generic technologies) and market applications derived from these platforms (proprietary technologies) are two very different investments and are managed quite differently within large R&D-based companies.

In addition, a third element is represented by the "infratechnologies" box in Figure 1. Infratechnologies are a diverse set of technical tools that are necessary to efficiently conduct all phases of R&D, to control production processes, and to execute marketplace transactions for complex technology-based goods.[11] These tools are called infratechnologies because they provide a complex but essential technical infrastructure.

Many infratechnologies are adopted as industry standards, emphasizing their public good content (Tassey, 2000). Without the availability of this technical infrastructure, especially codified as standards, transaction costs for all three major stages of economic activity—R&D, production, and marketing—would be much higher, thereby significantly slowing the evolution of technology life cycles.[12] The multiple economic roles of infratechnologies in the typical technology-based industry are indicated in Figure 1 by the several arrows pointing from the infratechnology box to the target areas of impact.

Looking at a technology-based economy more broadly, most small companies do not attempt much generic technology research, relying on assimilating the required technical knowledge from external sources. Large companies may undertake more of this type of research, but a number of factors related to its public good content result in underinvestment.[13] Thus, the proposed taxonomy is essential for R&D policy analysis because it has implications for the sources of radically new technology platforms, access to these generic technologies (property rights issues), process efficiencies and the transaction costs associated with market access.

Dynamics of R&D and underinvestment

The above conceptual model identifies the major technology elements relevant for policy analysis as a static framework. However, beginning with Mansfield *et al.* (1971), economists have shown that complex relationships exit between R&D investment over time and industry structure, risk preferences, strategic and project management capabilities, and the availability of a supporting technical infrastructure. A static framework overlooks the dynamic process by which new technology is created and utilized and thus how underinvestment can occur over the technology life cycle.

Corporate and government R&D investment decisions are made at the microeconomic level. Many attempts have been made to incorporate cause-and-effect relationships for the impact on output of an aggregate (firm- or industry-level) technology variable through the use of production functions, including the acquisition and use of a stock of knowledge. However, most of this literature assumes a continuous, steady-state stream of "opportunities" that, in turn, drive an optimal R&D intensity. The result of such a model is that the "stock of knowledge" grows steadily over time in response to a homogenous R&D process.

In fact, the dynamics of technological change are far from a continuous flow. Rather, "lumpy" advances in generic technical knowledge drive applied R&D for periods of time, after which diminishing returns set in. This requires refurbishment of the generic technology base or the optimal R&D intensity declines (Klevorick *et al.*, 1995).

Such a pattern implies a life cycle model of technological change. The richness of the underlying science base, the pattern of refurbishment of the generic technologies derived from this science base, and the efficiency of the process of creating market applications (innovations) combine to determine the optimum R&D intensity within and across life cycles.[14] Therefore, because the required analytical framework must be based on a technology life cycle concept, it requires a dynamic component.

Collectively, investments in the several phases of R&D determine the technology life cycle's trajectory and longevity and appropriate analysis can occur at several levels of aggregation.[15] The majority of research with a life-cycle dimension has characterized innovation as endogenously generated by competing firms that draw upon a knowledge base where the knowledge base is largely fixed. In such a framework, firms have limited windows of opportunity to apply this knowledge and earn a profit before new technologies mysteriously appear and a subsequent new wave of innovations takes their market shares away.

Schmookler (1966) partially solved this problem by introducing the concept of an accumulating knowledge base over time that innovators utilize to execute the process of creative destruction. Critically important is the proposition that accessing this knowledge base is affected by its public good character, which implies spillovers and increasing returns to subsequent innovators later in the life cycle (Caballero and Jaffe, 1993). However, although R&D or the resulting stock of knowledge has been incorporated into partial and general equilibrium growth models, the mechanisms by which this knowledge base evolves and is drawn upon have not been specified.[16]

Klevorick *et al.* (1995) point out that R&D investment (measured as an intensity ratio to indicate reliance on R&D as a competitive strategy) has been explained by either *search* models (technical knowledge is drawn from a fixed generic knowledge base) or *capital* models (technical knowledge is characterized as an evolving stock, augmented by flows of R&D investment). The former focuses on the efficiency of search mechanisms and the richness of the underlying generic technology base to explain R&D intensity. The latter uses production functions to first relate R&D investment to increases in the stock of applied knowledge and then indicate how this knowledge is used to produce new products/services or increase the efficiency of producing existing products/services.

Search models assume a fixed knowledge base and hence a fixed set of technological opportunities. Thus, diminishing returns are implied. The capital model also embodies diminishing returns through fixed relationships with other factors of production. These fixed relationships imply a constant set of technological opportunities, that is, a fixed generic technology base from which firms can develop market applications (innovations).

In reality, the generic technology base of an industry does not remain constant. The set of technological opportunities available to a firm or an industry changes over time. The rate of change, that is, the refurbishment of the opportunity set, determines the range of expected rates of return from investment in innovations and therefore the R&D intensity of the industry that draws upon this technology base. A central challenge for R&D investment theory is to explain the difference in incentives to invest in the opportunity set versus applications of this set (actual innovations).

Unfortunately analysis of the dynamics of investment in generic technologies (the opportunity set) has been limited. To understand potential constraints on private investment in the generic technology base requires a disaggregated technology model (Figure 1) coupled with a life-cycle framework to show how the major technology elements evolve over time and interact with each other. Investment in the major elements are affected by the functional linkages among these elements (indicated by the arrows in Figure 1), which evolve over the technology's life cycle.[17] The nature of an element or its linkages to other elements can cause underinvestment, as described in the following section.

Causes of underinvestment

Beginning with Nelson (1959) and Arrow (1962a), economists have attempted to explain the causes of underinvestment in R&D. Nelson focused on basic science and its pure public good content. Arrow identified "technical" risk and appropriability (the existence of spillovers) as major causes of underinvestment in R&D generally. The inability of investors to manage technical risk and the inability or unwillingness of property rights systems to prevent spillovers ("market" risk) can combine to create underinvestment.[18] While virtually all investments entail some degree of aggregate risk, the high levels and interactions of technical and market risk intrinsic to individual R&D investments impinge upon the ability to manage or insure against R&D risk.[19]

The sources of technical and market risk need to be specified to eventually provide quantitative cause-and-effect assessments over a technology's life cycle. These sources are more diverse than commonly believed. Specifically, R&D risk arises from

(1) *Technical Complexity:* Complexity and thus technical risk increase with the magnitude of the targeted advance and the complexity of the technology's interfaces with other technologies within a broader technology system;
(2) *Timing:* R&D investment time horizons, based on a combination of acceptable technical and market risk, are often shorter than those required to successfully develop radically new technologies;
(3) *Economies of Scale and Scope:* Some R&D is capital-intensive, which unbalances or limits diversification in R&D portfolios. This accentuates the tendency of R&D strategies to focus on achieving scale efficiencies to reduce both technical and market risk associated with specific applications, at the expense of potential economies of scope enabled by

an emerging technology; the result is to reduce incentives to invest in generic technologies with multi-market potential;

(4) *Spillovers:* Leakage or spillover of technical knowledge to companies that did not contribute to the research project creating the knowledge is typically greater the earlier in the R&D cycle an investment is undertaken, thereby increasing market risk.

Any one of these four sources of R&D risk can have serious negative impacts on the composition of private-sector R&D investment. Moreover, the severity of their impacts varies over technology life cycles and across technologies. Their impacts are particularly pronounced in the early phases of R&D aimed at next generation or radically new technologies and at supporting technology elements with a strong infrastructure character (infratechnologies). The four categories are discussed in more detail below.

Complexity. Complexity affects risk in a number of ways. The greater the targeted advance or the more multidisciplinary the R&D process, the greater the technical risk. The result can be market failure of the Arrow risk management type. Complexity is an especially vexing problem early in the technology life cycle because such research typically requires multidisciplinary research teams and unique research facilities that do not exist. Making such investments requires a positive assessment of a distant and uncertain potential market. Thus, firms react to research efficiency barriers by not undertaking more radical research projects, even though the potential RoR is high. As the technology life cycle evolves, companies often are caught between inefficient R&D processes and the intense competitive dynamics of high-tech markets, with the result that products are rushed to market with multiple performance defects.[20]

Moreover, most technologies driving advanced economies are complex systems in that a number of different technologies come together to eventually meet final demand.[21] The response has been specialization of private sector R&D on specific components of the overall system technology. Specialization creates inefficient R&D investment at the component level due to inadequate system-level performance specifications and system integration requirements. These economic impacts result from an increase in the public good content of both generic technologies and infratechnologies at the systems level.

Biotechnology is a dramatic example of the impacts of complexity leading to market failure. Using Figure 1 as a taxonomy, Table I lists multiple areas of bioscience (column 1) that have had to advance before a larger set of generic product and process technologies (columns 3 and 4) could be developed. These generic technologies have evolved over the past 25 years and slowly attracted private R&D funds, which are just now beginning to yield significant numbers of proprietary market applications (column 5).

However, lack of focus on and research tools for advancing complex generic technologies before attempting the development of specific innovations (new drugs) led to attempts to leap from advances in basic science to product development.[22] The result has been a very low success rate for biotechnology firms. In fact, 25 years after Genentech became the first public biotech company, only 12 of the 50 largest companies are profitable and the industry as a whole is still losing money.

Table I also shows the other category of industrial technology with significant public good content—infratechnologies (column 2). The demand for infratechnologies is derived from the demand for the generic technologies and its applications. However, the public good character of these technical tools results in significant underinvestment, thereby further reducing the productivity of R&D.[23]

In summary, dealing with complexity is made more difficult by the lumpy nature of technological advance (periodic quantum advances in the generic technology) and the increasing systems nature of modern technologies. These factors are pushing companies toward more focused R&D and market strategies within technology life cycles. Efficient technology systems require multiple components to be developed simultaneously and optimized so that maximum system performance is achieved.

Timing. Because technologies evolve in cyclical patterns, timing of R&D is critical. Underinvestment mechanisms can affect the rate of market

Table I
Interdependency of public—private technology assets: biotechnology

		Generic Technologies		
Science Base	Science Base	Product	Process	Commercial Products
• cellular biology	• bioinformatics	• antiangiogenesis	• cell encapsulation	• coagulation inhibitors
• genomics	• biospectroscopy	• antisense	• cell culture	• DNA probes
• immunology	• combinatorial chemistry	• apoptosis	• DNA arrays/chips	• drug delivery
• microbiology/ virology	• DNA sequencing, profiling	• bioelectronics	• fermentation	• inflammation inhibitors
• molecular biology	• electrophoresis	• biomaterials	• gene transfer	• hormone restorations
• nanoscience	• fluorescence	• biosensors	• immunoassays	• mRNA inhibitors
• neuroscience	• gene expression analysis	• functional genomics	• implantable delivery systems	• nanodevices
• pharmacology	• bioinformatics	• gene delivery systems	• nucleic acid amplification	• neuroactive steroids
• physiology	• magnetic resonance spectrometry	• gene testing	• recombinant DNA	• neuro-transmitter inhibitors
• proteomics	• mass spectrometry	• gene therapy	• separation technologies	• protease inhibitors
	• nucleic acid diagnostics	• gene expression systems	• transgenic animals	• vaccines
	• protein structure modeling/analysis	• monoclonal antibodies		
		• pharmacogenomics		
		• stem-cell		
		• tissue engineering		

penetration *within* technology life cycles ("life cycle evolution") or the transition *between* two distinctly different technologies and their respective life cycles ("life cycle transition"). Life cycle evolution is determined both by macroeconomic factors (general capital market risk, which primarily affects the amount of R&D investment) and by the efficiency of R&D processes.

Life cycle transition is an R&D efficiency issue is affected by the public good characteristics of the generic technology and associated infratechnologies. At life cycle initiation, corporate decision making targets investments in new generic technologies in order to prove concepts before deciding to commit much larger funding to actual innovation efforts. Generic technology research not only involves greater technical risk but also requires longer R&D investment time horizons. Hence, excessive discounting for both risk and time results in underinvestment during this early phase of R&D.

More specifically, a company evaluating the risk of investing in a new technology faces a set of projected performance/cost ratios, represented by curve 2 in Figure 2. Initially the performance/price ratio for the new technology (point B) will be below the current ratio for the defender technology (point A on curve 1).[24] The consequent risk of lower technical and cost performance, possibly for some time, pushes estimates of positive rates of return into the future, which in turn lowers the estimated net present value of the R&D investment.

Moreover, the innovator's risk assessments are compounded by the fact that the defender technology seldom gives up without a fight. For example, in the face of a challenge from flat-panel displays, manufacturers of cathode-ray tubes continued to reduce costs and improve picture quality, thereby raising the hurdle for the invading technology (by continuing to move up Curve 1, even as the curve flattens). Thus, although diminishing returns for the defender technology have typically set in when the new technology appears, potential advances contrib-

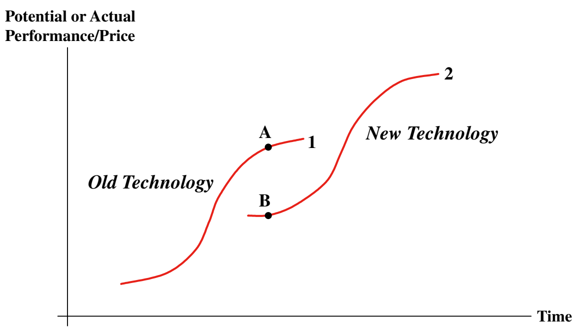

Figure 2. Translation between two technology life cycles.

ute to the risk of investing in the new technology at an early point in its life cycle.[25]

Transitioning *across* technology life cycles is a particularly difficult issue for R&D policy process to address. A number of high-tech companies manage transitions among successive *product* life cycles quite effectively. However, the transition to a radically new *technology* life cycle is seldom achieved by the majority, if any, of the firms applying the defender technology. Many of these companies lose out to small, innovative firms that are willing to take on the high risk posed by the initial performance/price gap. In cases of radically new technologies, the established companies may be replaced by new industries—either domestic or foreign.

This aspect of creative destruction occurs in part because current suppliers are focused on enhancements of the defender technology in an attempt to maximize short-term cash flows. They consequently wait too long to undertake research targeting the new technology. This pattern also occurs in part because the established industry is organized to develop and market the defender technology and does not have all of the required R&D skills and facilities needed to migrate to the new technology.[26] In summary, the *composition* of R&D investment is inadequate from a long-run (multi-cycle) perspective.[27]

Companies with long investment time horizons and high risk tolerances (and hence relatively low discount rates) might be able to make sufficient investments in the new technology (and associated R&D capabilities) before the old technology matures. This would allow innovation to take place higher on performance/price curve 2 than point B. Such a strategy seems logical because market penetration can occur more rapidly. However, with the typical R&D life cycle about 10 years (Branscomb and Auerswald, 2002; Cannon, 2002) and increasing competitive pressures raising discount rates, even large, R&D-intensive companies with substantial central research labs are decreasing investment in the longer-term research that produces most new technologies.

The problems in moving up the performance/price curve from point B are evident in the biotechnology example. The evolving structure of this industry (small firms focusing on individual drug candidates) accentuates an inefficient tendency exhibited in all high-tech industries to varying degrees toward a forced approach to product development; that is, attempting to develop proprietary products with an inadequate underlying generic technology and supporting infratechnologies. This raises product failure rates due both to inadequate proof of concept (generic technology) and to inefficient research methods that increase R&D costs leave excessive uncertainty with respect to research outcomes. These higher failure rates discourage risk capital needed for follow-on investments.

Economies of scale and scope. Arrow (1962) first identified the problem of indivisibility of technical information and its consequent effect on R&D investment. All other factors aside, the typical firm tries to fully utilize its resource base; that is, it seeks to capture the benefits from the entire scope of potential market applications that arise from a generic technology and supporting infratechnologies. Time and considerable resources are needed to acquire a unique set of technology assets, so maximum efficiency in utilizing these assets is essential. First mover and sustained leadership strategies lead firms to seek to reduce risk by concentrating on the development of a relatively limited set of technology assets and related market applications (so-called "core competence" strategies).

In contrast, portfolio theory preaches diversification into multiple product fields to diversify market risk. However, firms have found that this strategy can entail substantial technical risk due to the increased level of resources required to achieve and maintain competency in multiple technologies, each requiring somewhat different sets of research skills and facilities. Moreover, diversified portfolios also can require multiple sets of production and marketing assets. Such risk presents a disincentive to diversify horizontally or vertically. Increased global competition in technology-based markets has accentuated this risk. Thus, firms seek to reduce risk by concentrating on the development of technology assets for a limited set of market applications (so-called "core competence" strategies).

A narrower technology focus often includes downsizing central corporate research laboratories and shifting the composition of the research to more applied topics that directly support the remaining lines of business. Less generic technology research is conducted because a narrower market focus implies less opportunity for economies of scope. Many firms, especially small and medium ones, do little or no generic technology research, instead relying on external sources for new technology platforms.

In summary, such strategic shifts can be a two-edged sword. On the one hand, concentrating R&D resources on fewer technologies and their market applications can increase economies of scale. Such a strategy reduces technical risk (core competence is emphasized) and possibly market risk (due to specialization). On the other hand, market risk may actually be increased because product/service diversification has been reduced; that is, the firm's portfolio of technologies and subsequent market applications is more narrowly focused and therefore more dependent on a smaller number of markets.

Evolving industry structures offer a partial adjustment mechanism. In the last two decades, forms of collaboration for cost sharing and risk pooling have proliferated, including multi-company consortia and partnerships with universities and government. However, these complex entities have contractual, organizational, and management problems that can cause inefficiencies in the R&D undertaken and thus can lead to suboptimal results (Miotti and Sachwald, 2003).

Thus, while specialization may enhance short-term performance, it accentuates "path dependence" and thereby reduces firms' and entire industries' ability to adapt to new technology life cycles by reducing generic technology research capabilities. Such propositions are reflected in evolving "evolutionary theories" of technology-based growth (Nelson, 1995) in which acquired assets and supporting infrastructure tend to maintain technological pathways as a conscious attempt to fend off the Schumpeterian process of creative destruction. However, the life cycle nature of technological change eventually leads to the very destruction that such strategies seek to avoid. The incumbents and their defender technology inevitably lose out to the emerging technology.

Spillovers. In the economics literature, market failure in the technology-based sector of the economy is most often related to the existence of externalities called "spillovers," defined as technical knowledge and thus presumably economic benefits that are not captured by the firm creating the knowledge. That is, portions of these benefits accrue (spill over) to other firms, so that these firms "free ride" on the R&D of the innovating firm.

Unfortunately, economists have not reached a consensus on a taxonomy for characterizing and analyzing R&D spillovers. The literature identifies several types of R&D spillovers that are not necessarily distinct from one another. Most directly associated with R&D investment are "price" or "market" spillovers and "knowledge" spillovers (Griliches, 1979; Mohnen, 1996). In both cases, technical knowledge passes from the originator to other economic agents without full compensation.

Price spillovers occur when the increased value of new or improved products or services is not fully reflected in the price differential between the old and the new versions of the product or service. The portion of the benefits captured, but not paid for, by the user diminishes the reward to the supplier (innovator). In both cases, the return on investment for the innovator is reduced, which, if sufficiently pronounced, can constrain incentives for further R&D investments.

Knowledge spillovers occur when technical knowledge itself leaks or spills over from the innovator to competing firms without compensation to the innovating firm. The competing firms use the knowledge to imitate the innovation and thereby potentially reduce the innovating firm's return on investment. To the extent that the firms imitating the technical advance do not compensate the innovating firm (such as through licensing arrangements), a "reward-capture" problem can arise which inhibits private investment in R&D.[28]

Finally, this literature also identifies "production" spillovers in which knowledge diffuses into an industry in purchased capital goods embodying new technology developed in the capital goods industry. Such spillovers result from synergistic interactions within the purchasing industry (such as learning by doing) and are argued to be independent of any knowledge spillovers in the capital goods industry or price spillovers in the custody exchange due to market dynamics. Empirical studies have not shown that production spillovers present a significant disincentive or occur with any regularity.[29]

Spillovers create several problems for R&D policy analysis. On the one hand, they are viewed negatively because, when excessive, the expected RoR to the prospective innovator is lowered to the point that the contemplated investment in R&D is not undertaken. On the other hand, it is this unique characteristic of technology—the fact that it diffuses widely—that enables large and diffuse economic impacts to be realized.[30] Moreover, spillovers do not have to mean one-for-one subtraction of benefits from the innovator. For example, network externalities can result from imitation by other firms coupled with the establishment of an efficiency-enhancing infrastructure that enable increasing returns for many firms from serving a larger user population.

The form of technical knowledge influences the spillover pattern. Technical knowledge has been classified as either ''embodied'' or ''disembodied''. Production spillovers are a market transfer of embodied technology. Price spillovers usually refer to this type of technical knowledge, as well. In contrast, technical knowledge developed within or somehow acquired by an industry can be disembodied in that the knowledge is known and used independently of any physical structure. The industry's own generic product and process technologies fall into this category, as do many types of nonproprietary technology infrastructure (infratechnologies). Such disembodied knowledge tends to ''leak'' or spillover, causing the frequently referenced appropriability problems.

Knowledge has also been classified as tacit or codified.[31] These terms refer to the form of technical knowledge, that is, whether or not the knowledge has been formalized as a commodity in the sense defined by Arrow. Much tacit technical knowledge is embodied in human capital. Romer (1990) refers to such tacit knowledge as ''rival technical knowledge'' in that its use in one application (assigning human research capital to a particular R&D project) prevents its use in alternative applications. Such knowledge only diffuses slowly from one person to another usually and requires close contact between parties executing the transfer. Thus, spillovers occur slowly. Generic technology is largely of this type (Darby and Zucker, 2003).

The previously discussed concept of ''technological opportunity'' is a means of characterizing the potential inflows (inward spillovers) of technical knowledge. For example, Klevorick *et al.* (1995) used the Yale Survey of R&D managers to analyze inter-industry differences in technological opportunity. They concluded that these sources lie primarily outside the industry and that levels of opportunity result from intrinsic differences in knowledge spillovers.

Finally, adoption of the disaggregated technology model summarized in Figure 1 allows the observation that the infratechnology element is a case where spillovers are largely desirable. The strong infrastructure character of infratechnologies, frequently manifested in the form of standards, means that significant economic benefits are realized only when all participants in a market (both demand and supply sides) have access to the same infrastructure (Tassey, 2000).

For example, using the same acceptance test method to consummate a market transaction for a technologically complex product reduces transaction costs and thereby increases the rate of market penetration of the new technology. Thus, policy makers must ensure the availability of an efficient standards infrastructure to achieve utilization of these infratechnologies. However, the optimal achievement of such is not easy, as timing and content of standardization (and hence the derived demand for the underlying infratechnologies) are difficult investment objectives to define and hence to manage. ''Free riding'' with respect to the standard is actually encouraged to increase economic efficiency, but this objective guarantees underinvestment in the underlying infratechnology.

In summary, the evolving literature on spillovers implies a model in which technical information diffuses in specific channels and at non-negligible cost. Rate and direction depend on the degree to which the information is generic or applied and on its ultimate use (proprietary product/process development or infrastructure). The implied disaggregated model leads to a differentiated public policy perspective. That is, spillovers of generic technologies and infratechnologies are desirable because of their public good content, whereas proprietary technologies require assignment of property rights to provide incentives for market applications. This critical distinction allows a significant improvement in R&D policy over that implied by the Arrow-type conclusion in which property rights result in underutilization of this information.

Defining underinvestment for government policy development

Having identified the causes of underinvestment in R&D, a method for estimating underinvestment can now be discussed.

A major analytical challenge for R&D policy analysis is to assess the extent to which the innovator (private) rate of return (PRR) falls below the hurdle rate (HR) for corporate R&D investment. The analysis also must determine the reason for the sub-threshold PRR. A low projected PRR could simply be due to inadequate technical and/or market appeal, which means the technology has a relatively low potential social rate of return (SRR). Alternatively, intrinsic or externally imposed market failures of the type discussed above could be substantially suppressing the PRR below a relatively high SRR.

While recognizing the existence of a rate-of-return gap and noting that the magnitude of underinvestment can be related to this gap, a more accurate representation of underinvestment phenomena is possible. In Figure 3, four R&D projects are shown.[32] The vertical arrows represent the possible SRR and the horizontal arrows indicate the possible PRR. Hence, the lengths of the arrows represent the ranges of potential outcomes from a particular R&D investment. Risk is the probability of an outcome (RoR) being below the HR.[33]

These two required elements of policy analysis are incorporated in Figure 3. The first three projects (A, B, and C) have ranges of expected PRR below the corporate HR and therefore would likely not receive significant or sustained private-sector funding. Project A is an early-phase (generic) technology research project, which has significant economic potential. However, at this point in that technology's development, the actual (or uncertainty about the actual) levels of technical and market risk are beyond the ranges acceptable to corporate R&D managers. That is, the risk-adjusted PRR estimate is lower than individual firms' HRs, and thus a corporate commitment to funding this project in response to conventional decision criteria will be limited.

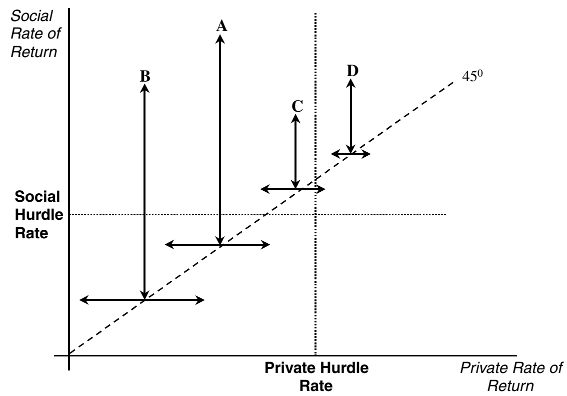

Figure 3. Government R&D project selection: rate-of-return criteria.

Such a project is a candidate for government financial assistance, usually in the form of cost-shared funding for the conduct of generic technology research. Cost sharing reduces risk enough to stimulate initial private investment in applied R&D (in Figure 3, Project A shifts to the right toward the innovator's HR). If successful, this early-phase technology research will reduce uncertainty about the probability distribution of technical and market outcomes and shift the distribution above the innovator's HR. Conversely, this research could result in an insufficient increase in the expected PRR and might also reduce the expected SRR below its HR, causing termination of R&D in the particular technology by both industry and government.

Project B also has a very low expected PRR, while at the same time exhibiting a high SRR. A policy analyst might conclude that too large an investment of government funds would be required to reduce both uncertainty and risk and thereby raise the expected PRR above the private-sector HR. However, such projects are typical of elements of an industrial technology that have a strong infrastructure character and exhibit a high SRR. For example, the network externalities resulting from standardization greatly increase the SRR for this type of infrastructure, but simultaneously prevent ownership of the embedded intellectual property by any one firm. In such cases, the PRR may never be permanently raised above the HR. Thus, the R&D may be funded largely by government to capture the high SRR from eventual adoption by industry of the resulting infratechnology as an industry standard.

Projects C and D are quite different from either A or B. Project D is a typical R&D project found in a business unit of an R&D-intensive company. Its range of expected PRR is more than high enough to receive adequate private R&D funding. Project C is a marginal project under existing corporate R&D investment criteria; that is, its expected RoR does not quite exceed the private-sector HR, so this project is not funded.

However, projects such as C can have sufficient SRR (above the social HR) to warrant government investment incentives. In such cases, because C falls within the general category of applied R&D projects that a firm's business unit would typically consider, the R&D policy response should be directed at conventional corporate R&D investment criteria (as opposed to the distinctly different criteria used to evaluate long-term, exploratory technology research). For situations such as project C, a tax credit rather than direct funding is the correct policy response.[34]

R&D cycle analysis

The above discussion focuses on relationships between the nature of the R&D project and its risk-adjusted expected RoR at different points in the technology life cycle. Such comparative static analysis is important for identifying the factors that determine underinvestment.

In addition, the analyst must also analyze the dynamic element of underinvestment; that is, how the *pattern* of underinvestment affects the evolution of the technology. In essence, this analysis is an assessment of both technical and market risk management by industry over the technology life cycle.

As previously defined, generic technologies (proof of concept) provide the basis (technology platforms) for an array of market-specific applications (collectively, the technology trajectory). Although amounts invested in generic technologies are small relative to those for applied R&D, these investments are the first step in *technology* research. The distinction between scientific research and technology research is extremely important for market failure analysis. Scientific research seeks to advance *knowledge*. The concept of risk is relevant only in the sense that funding decisions are based on the expectation that scientific knowledge will be advanced.

In contrast, risk assessments are continually made by industry over the remainder of the technology life cycle based on estimates of the probability distributions of future market returns. These estimates drive R&D investment decisions. The nature of this risk is twofold: the probability of achieving the technical objectives perceived to be required for the target markets and the probability that when commercialization occurs, the market will be receptive.[35] As long as reward-risk ratios exceed private and social HRs, R&D funds will continue to be allocated to the technology's development. Patterns of R&D by which technology becomes progressively more applied is evident in the technology life cycles underpinning the major technology drivers of the last 60 years.[14]

However, the conventional characterization of a technology's evolution as a steady increase in its applied character masks a fundamental discontinuity in the pattern of risk reduction that strongly affects investment behavior. As discussed above, the technical complexity of the proof-of-concept objective of generic technology research, its frequent mismatches with existing corporate market strategies and internal research capabilities, and its distance in time from potential commercialization combine to cause a spike in both technical and market risk. By occurring early in the technology research cycle, this "risk spike" can block substantial private R&D investment.

Such a discontinuity in risk reduction is portrayed in Figure 4 for two hypothetical corporate R&D projects, A and B. In both cases, the assumption is that the underlying science base has been advanced over time to a point at which corporate managers can make initial assessments of the technical and market risk associated with the development of technologies derived from on this science.

Based on the body of scientific knowledge acquired through basic research, the decision making process might be said to face a purely "technical" risk, R_0, that summarizes estimates of the probability that a technology could be developed that performs some generalized function. However, the characteristic that distinguishes technology research from scientific research is the fact that the ultimate intent is commercialization. Thus, an additional amount of technical risk must be estimated and added to R_0 because the scientific principles presented now have to be proven capable of conversion into specific technological forms with specific performance attributes that meet specific market needs. Several of the factors cited in Section "Dynamics of R&D and underinvestment" can raise this technical risk. Moreover, total risk is now reassessed in view of the need to meet production cost objectives.

A "market" risk also must be estimated to allow for the significant probability that demand for the new technology will be overestimated or that market penetration will be slower than projected due to the factors cited in Section "Dynamics of R&D and underinvestment" such as improvements in the defender technology. Slower market penetration could mean that production costs will be high relative to the defender technology for a longer period of time due to failure to realize economies of scale or scope. This market risk must be added to technical risk.

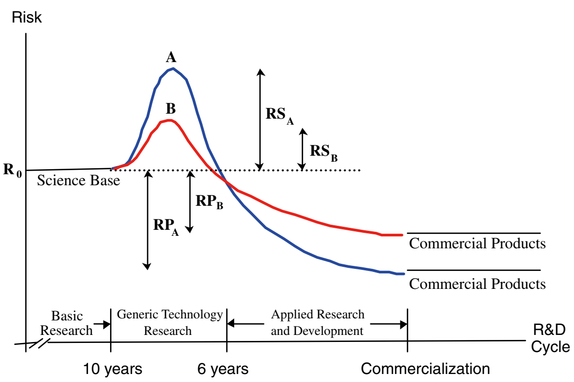

Figure 4. Risk reduction in the R&D process.

In Figure 4, the risk spikes RS_A and RS_B represent the combined increase in technical and market risk for projects A and B, respectively. Such risk spikes might be thought of as the "public (or social) risk" component because they occur in the early generic technology research phase, which has the public good dimension described earlier. Project A is the more radical innovation and so it presents a greater initial risk spike, RS_A.

Without the risk spike, firms would be faced only with a reduction in the "private risk" component, RP, associated with applied R&D. Project A also requires greater risk reduction during applied R&D, RP_A.

In the absence of the risk spike, conventional R&D investment criteria would deal with RP because it falls within acceptable reward-risk ratios. Thus, if at the level of pure technical risk, R_0, conventional corporate R&D criteria result in private investment based on risk-adjusted rate-of-return estimates, the policy problem is to overcome the risk spike so that these criteria can be applied. The importance of overcoming the risk spike for the more radical technology, A, is increased by the fact that the overall risk associated with project A will actually decline to a lower level than that for Project B. This occurs because, if Project A is successful technically, the resulting set of market applications will likely have a larger collective value than B, hence increasing the probability of a RoR above the firm's HR.[36] This is depicted in Figure 1 by a greater decline in private risk for Project A, indicated by RP_A.

Current R&D policy tools do not fully recognize the large discontinuity in the risk reduction process occurring at the transition between scientific and technology research. If this risk spike did not occur, the risk curves in Figure 4 would have steadily declining slopes and would support proponents of little government R&D funding beyond basic science. However, the substantial jump in total risk caused by the potential divergence between technical and market requirements on the one hand and research capabilities, time discounting, and corporate strategy mismatches on the other can and do lead to substantial underinvestment.[37]

In summary, the advancement of basic science sufficient to allow technology development to begin does not guarantee immediate or even eventual commitment of adequate private sector funds. Reducing risk spikes through early-phase generic technology research is necessary to enable optimal private sector R&D spending. However, the fact that specific elements of an industrial technology are quasi-public goods means that their efficient development over the entire life cycle requires a mixture of public and private funding, distributed according to the magnitude and duration of various market barriers.[38]

4. Defining and measuring underinvestment

Having identified the critical quasi-public elements embedded in the typical industrial technology, the policy analyst ideally should then estimate an optimal level of investment for specific technology elements against which to assess actual investment behavior. For identified underinvestment mechanisms (the "cause"), such analysis requires the selection of impact metrics and measures that enable quantification of the economic impact (the "effect") of identified barriers. The resulting quantitative estimates are combined with qualitative analyses of the identified barriers to enable appropriate policy responses to be selected and then to allocate resources for those responses across technologies. Ranking for policy purposes will be a function of the estimated SRR and the difference between the PRR and the private HR.

However, economic research has made only partial progress in developing and applying the required analytical tools. An overall optimal R&D intensity has been correlated with assessments of technological opportunities (Klevorick *et al.*, 1995) and estimated at the national level by Jones and Williams (1998, 2000).[39] At the microeconomic (technology) level, the opportunity set determines the distribution of possible rates of return from applied R&D (innovation investments) and thus implies an optimal R&D intensity for that technology-based industry. Based on initial rankings of technological opportunity, the policy analyst ideally would estimate SRRs and PRRs for each technology and compare them with appropriate HRs. However, arriving at an optimal R&D intensity to use as a benchmark for underinvestment analysis, estimating the appropriate investment "gap," and then explaining this gap in terms of a market failure mechanism are extremely difficult analytical and empirical steps.

Rate-of-return analysis

To some extent, the use of return-on-investment techniques (borrowed from corporate finance) has proved useful for demonstrating retrospectively the range of economic impacts from specific technology investments.[40] However, this approach is burdened with several conceptual and empirical issues, especially when applied to prospective (strategic planning) exercises.

One problem with prospective analyses for R&D policy decisions is the fact that several significant and uncertain phases of R&D must take place followed by investments in production and marketing capabilities before commercialization can occur. This problem is particularly severe for basic and early-phase (generic) technology research. In the latter case, the analysis is complicated by the fact that the additional R&D following investment in the generic technology is typically much larger, but its amount, composition, and timing are driven by the direction and timing of generic technology research.

Moreover, the time to commercialization from generic technology research is quite long. Rate-of-return analysis is sensitive to the assumed time series of R&D investment and the subsequent commercialization and market penetration patterns. Moreover, use of rate-of-return techniques or discounted cash flow/net present value measures produce a negative bias for prospective R&D investments through their treatment of risk and reward. Specifically, these techniques do not ascribe value to the intellectual capital created by the R&D cycle as it reduces technical risk (Boer, 1998, 2000). That is, at best only an incomplete market exists for technology capital.

Nevertheless, even for prospective analyses, the conventional rate-of-return method can be useful for the development and management of government R&D support roles.[41] To this end, rate-of-return analysis can be applied in two ways. One way is to compare rates of return between major categories of investment. If the SRR from a category of investment (say, R&D) is higher than that from another category, the suspicion is that some barrier is preventing efficient allocation of resources in the private sector. That is, underinvestment is occurring in the category with the higher SRR due to a suppressed PRR. Otherwise, funds would flow into the investment lowering the SRR until the marginal RoR was reduced to the opportunity cost of capital.

The relationship between the rate at which the expected PRR declines with increased R&D investment and the HR (the opportunity cost of capital for private sector investors) is shown in Figure 5. The two negatively sloped lines, P and P′, are marginal efficiency of investment (MEI) curves. Firms rationally invest in the R&D projects with the highest expected PRR first and then select projects with the next highest PRRs and so on until the opportunity cost of capital (HR) is reached. The actual slope of an MEI curve can be steep or shallow depending on technological and market opportunity, efficiency of R&D, etc. The entire curve can shift up or down depending on intrinsic characteristics of the technology and its associated infrastructure. Figure 5 also shows two alternative HRs, HR and HR′.

For an MEI curve P and a hurdle rate HR, this industry will invest an amount Q in R&D. If one or more of the types of market failure discussed in Section "Causes of underinvestment" intensifies, the entire MEI curve could shift downward to P′ causing a reduction in R&D investment to Q′. Similarly, if market conditions change so that either the PRR on alternative investments or general risk averseness increases, the HR could shift upward to HR′ causing an equivalent reduction in R&D investment to Q′.

A second application of rate-of-return analysis is to represent the amount of underinvestment in R&D by the "gap" between the SRR and PRR for the same category of investment. The implication is that if industry can only earn the PRR, then less R&D investment will be forthcoming.

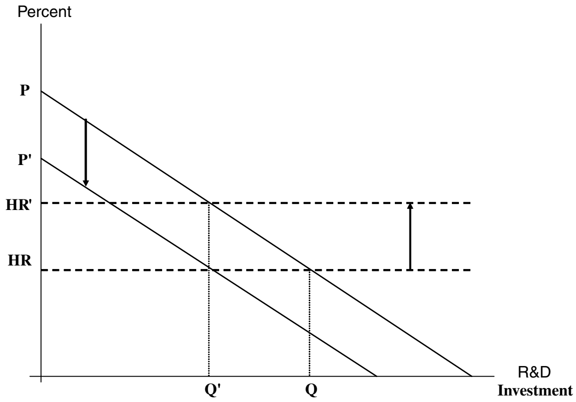

Figure 5. Marginal efficiency of R&D investment (expected private rate of return).

This conceptual approach would be represented in Figure 5 if innovating firms are assumed to capture all the returns generated by their R&D; that is, the curve P is also the SRR. If some barrier (typically assumed to be spillovers) causes a downward shift in the PRR curve to P′, then a "gap" appears leading to a reduction in R&D of QQ′ (for a HR).

However, this characterization of underinvestment is simplistic along several dimensions, especially with respect to the implied assumption of a fixed gap and the implications of a gap for policy:

(1) The "gap" is important only to the extent that it pushes the PRR below the HR. Obviously, the closer the PRR is to the HR, the fewer additional projects will be undertaken before the marginal PRR declines to the HR. The slope of the MEI curve is clearly important in this regard.
(2) From an economic growth perspective, "outward" spillovers can actually have a beneficial effect because they create opportunities for new firms to enter the market for the technology, which thereby expands competition and hence the range and amount of economic benefits delivered. That is, the reverse concept of "inward" spillovers to other firms in the industry needs to be taken into account. Moreover, to the extent that inward price or production spillovers are

realized from supplying industries, the innovating industry benefits in the form of higher PRR, thereby leading to more R&D in that industry.[42] The implication is that the size of the gap and hence the PRR relative to the HR will likely vary over the technology life cycle.

(3) A larger market resulting form spillovers can also allow network externalities to be realized or a general risk reduction to occur that opens that market to more risk-adverse groups of buyers. The result may be an increasing PRR for the initial innovators, as well as for imitators. Moreover, less steep MEI curves resulting from additional and larger markets derived from the generic technology lead to more applied R&D. Such effects usually are significant in the middle and later portions of the technology life cycle, where applications become "commoditized" and competition is based more on cost (personal computers are an example). Thus, the dynamics of technology-based markets can keep the PRR above the HR and maintain reasonable levels of private R&D investment for a longer period of time (extending the technology life cycle).

(4) Whereas the SRR is a single aggregate return on an industry's R&D investment, the PRR varies among firms in that industry, including the primary innovator, which makes the estimate of a PRR for the purpose of "gap analysis" difficult. In the end, what counts for policy analysis is the number of firms for which the PRR exceeds the HR at each phase in the R&D cycle.

Estimating underinvestment

Ideally, estimates of the PRR associated with each technology element would be obtained at different points in time as part of the underinvestment analysis. Such estimates would be compared with both the industry's average HR and the estimated SRR to determine if a government policy response might be warranted and, if so, over what period of time.

However, conventional applications of rate-of-return analysis, especially in the early phases of the R&D cycle, are difficult to execute. For example, because generic technology research is the first phase of technology R&D and is aimed at a next generation or an emerging technology, significant markets typically do not yet exist. Point estimates for dates in the future are highly uncertain and, if relied upon as the sole basis for initiating an R&D project, can lead to expensive long-term commitments with frequent negative outcomes.

Modified techniques are therefore required. One alternative approach involves regular assessments of progress toward technical goals and preliminary estimates of market opportunities. Repeated revisions of assessments of the extent to which technical and market risks are being reduced in effect segment the R&D funding processes with zero-based decisions occurring at each check point. Such "real options" tools can continue to be used until individual companies are able to apply conventional investment criteria to estimate PRRs.[43] As these PRRs rise above HRs, government subsidies can be phased out.

The relative funding contributions of government and industry will be determined by the subsidies required to elevate the PRR above the HR (or to reach a termination decision). Real options models exist that are highly quantitative, although application of this approach to early-phase generic technology research will likely entail some compromises. The major management requirement is periodic review of technical risk reduction and observance of industry's behavior with respect to its internal R&D strategies (evidence of reductions in both technical and market risk).

Still, the more quantitative the selected approach, the better. For infratechnologies, a quantitative approach can frequently be used. Here, markets are often established for a high-tech product or service, but inadequate technical infrastructure of some type is creating significant inefficiencies and thereby retarding market penetration.[44] Thus, direct estimates of the cost of the inadequate infrastructure are obtainable from industry surveys. Removing or at least reducing the costs constitute an economic benefit and such estimates can be related to projected research program cost estimates to calculate an expected RoR.[45]

With respect to identifying, characterizing, and estimating the magnitude of underinvestment in infratechnologies, NIST has conducted several prospective economic analyses of the adequacy of existing infratechnologies in order to provide inputs for its strategic planning function. Table II summarizes five studies of the economic costs of underinvestment in infratechnologies.[46] The choice of topics for such studies is based on a somewhat unstructured process of background analyses, consultations with industry, reviews of other studies (if they exist), and higher order policy directives.

The scope of such studies can vary significantly, making their use in ranking underinvestment phenomena difficult. For example, the first study in Table II examined the costs of inadequate standards for the exchange of electronic product design data between automobile companies and their parts and subsystems suppliers. The estimated annual cost is only for that supply chain, even though other manufacturing supply chains (such as aerospace) are known to have similar interoperability problems. In contrast, the software testing study conducted two larger case studies—one in manufacturing and the other in services—which provided a sufficiently broad base (along with some other data) to rationalize extrapolation to the national economy level. Hence, this study not only produced cost estimates for the two case studies, but also a much larger (national) estimate of the costs of underinvestment in testing infratechnologies.

Examples of government responses to the "risk spike"

Several decades of large-scale funding of molecular biology research by NIH were required before private investment kicked in and spawned a biotechnology industry. A recent analysis of U.S. patent citations in biotechnology found that more than 70% of them were to papers originating solely at public research institutions (McMillan *et al.*, 2000). And, 20 years after the first biotechnology company went public, NIH still provides research funding to dozens of the more than 300 biotechnology companies.

The tremendous growth in health care productivity being made possible by a radically new technology also is creating a new industry with substantial economic growth potential. This phenomenon is occurring to a greater extent in the United States in large part because the U.S. Government funded both the science base and portions of the subsequent early phases of technology research (i.e., the quasi-public good element). This funding coupled with other factors such as clustering and a robust venture capital market has allowed U.S. industry and U.S. capital markets to reach positive investment decisions ahead of the rest of the world.[47]

Table II
Economic studies of costs due to inadequate infratechnology investment

Focus of study	Infrastructure studied	Industries covered	Estimated annual costs $billions
interoperability costs	product design data exchange	automotive supply chain	1.0
deregulation	metering, systems monitoring/control	electric utilities	3.1–6.5
software testing	all stages of the testing cycle	transportation equipment	1.8
		financial services	3.3
		extrapolation to U.S.	59.5
interoperability costs	business data exchange: demand, production, inventory, procurement, & distribution	automotive supply chain	5.0
		electronics supply chain	3.9
medical testing	quality of measurement assurance (calcium)	medical testing laboratories	0.06–0.199

In the past, non-market motivations (primarily national defense) allowed Federal funding of major new technologies at above-threshold levels, which subsequently drove economic growth for decades. The fields of computing and communications provide compelling examples of how government funding plays a critical role in advancing generic technologies and achieving minimum thresholds of R&D capability necessary to stimulate takeoff in private sector investment. Federal funding for electrical engineering in areas such as semiconductors and communications technologies (major components of computing technologies) has fluctuated between $800 million and $1 billion since the 1970s. Funding for computer science increased from $10 million in 1960 to approximately $1 billion in 1995. Over these time periods, such amounts have represented a major fraction of all early-phase technology research funding in information technology.[48]

The majority of this funding went to industry and university researchers. Not only did the government-sponsored research advance key areas of the underlying science and technology, but it also fostered a broad and deep R&D capability that leveraged follow-on private investment by industry. An extremely important aspect of this support is the extension of Federal funding beyond basic scientific research to generic technology and even experimental deployment. For example, before 1970, the Federal government sponsored individual researchers who developed generic network technologies, such as queuing theory, packet switching, and routing. During the 1970s, experimental networks, notably the ARPANET, were constructed. These networks were primarily research tools, not service providers. Most were federally funded because, with a few exceptions, industry had not realized the potential of the technology.[49]

During the 1980s, networks were widely deployed, initially to support scientific research. The National Science Foundation (NSF) was the major supporter of networking during this period, primarily through the NSFNET, which evolved into the Internet. At this point in networking technology's evolution, industry began to see the enormous economic potential. Companies such as IBM, Digital Equipment Corp., and Compuserve established proprietary networks. These networks were rapidly utilized worldwide for email, file transfers, and electronic funds transfers.

However, as often happens in the evolution of a major new technology, companies with a large share of the initial proprietary applications displayed little interest in the even greater potential of the generic technology. To be broadly successful and thereby have large economic impact, systems technologies such as the Internet have to be based on open architectures. This type of technical infrastructure greatly expands markets for system technologies by facilitating network externalities. Unfortunately, such a requirement presents a negative investment incentive to firms with substantial commitments to proprietary networks. Moreover, telephone telecommunications companies resisted computer networks, including the Internet, because the nature of voice communications networks is strikingly different from data networks.[50]

Similarly, IBM pioneered the concept of relational databases but did not pursue commercialization of the technology because of its potential to compete with established IBM products. NSF-sponsored research at UC-Berkeley allowed continued exploration of this concept and brought the technology to the point that it could be commercialized by several start-up companies and then by more established suppliers, including IBM. This pattern was also evident in the development of reduced instruction set computing (RISC). Though the concept was originally developed at IBM, RISC was not commercialized until DARPA funded additional research at UC-Berkeley and Stanford as part of its Very Large Scale Integrated Circuit (VSLI) Program in the late 1970s and early 1980s.[51]

Other examples of critical government funding of generic technology research include expert systems, speech recognition, and image processing. Industry began to invest in these and other areas of artificial intelligence (AI) in the 1960s but scaled back when the long time periods required for commercialization became apparent. Federal investments continued in these generic technologies for a decade or more until conventional industry R&D criteria could rationalize investments in applied R&D (i.e., projected PRRs exceeded HRs). Now, private investment is driving the commercialization of many AI technologies.

Thus, government R&D is frequently the mechanism to reduce uncertainty through support of basic research and then technical and market risk (the risk spike) in the early (generic) phases of a technology's development. However, this policy response does not guarantee commercialization in a fixed period of time, if ever. Technical and market risk may be reduced more slowly than anticipated, or market risk may actually be increased by unforeseen developments in competing technologies. In the case of gallium arsenide, the Department of Defense spent $570 million on research between 1987 and 1995, but the overall market had not reached $1 billion in annual sales by 1996.[52] Two reasons for this pattern were slower-than-anticipated progress in R&D and repeated extensions of the silicon-based semiconductor life cycle.

5. Summary

The analytical framework presented provides an approach to explaining and estimating R&D underinvestment trends in a cause-and-effect framework. Emphasis is on underinvestment in the early phases of technology R&D (generic technology research) and in technology infrastructure (infratechnologies); that is, the focus is on R&D composition distortions. The final stage of the policy process, not addressed here, is the selection, design, and implementation of policy responses—primarily the forms and management of government-funded research. The complexity of this last stage requires separate treatment. Also not analyzed are market failures associated with the aggregate amount of industrial R&D, which means primarily applied R&D. The cause-and-effect phenomena are significantly different from those associated with composition failures and therefore so are the appropriate policy responses.

Economic theory needs advancement in the area of R&D underinvestment phenomena. The underinvestment mechanisms are not well defined, with the literature focusing largely on spillovers. Even spillovers are not sufficiently treated, as several types exist and the relationships to private sector R&D investment behavior are not adequately specified. Moreover, the economic impact of spillovers has positive as well as negative dimensions. This complexity, coupled with the quasi-public good nature of technology research, has limited progress in developing models and useful tools for R&D policy analysis.

The approach taken in this paper involves three analytical steps. First, the identification and analysis of the causes of underinvestment is undertaken to enable ultimate selection of a policy response mechanism. Second, indicators of underinvestment in R&D are constructed to allow a scan of relevant private R&D investment patterns and supporting infrastructure trends and thereby identify broad investment categories for more intensive analysis. These indicators are then compared to data on economic performance in the relevant industries to prioritize and focus subsequent analysis. Third, based on these analyses, technology-specific impact metrics and measures are selected to allow quantification across technologies of the net benefits of past policy implementations and the expected net benefits from candidate policy initiatives.

If these three stages of analysis are successfully carried out, the results can significantly enhance the selection of policy response mechanisms and the levels of resources required for candidate initiatives; that is, a match can be achieved between different types and modes of underinvestment and the most efficient policy responses.

Acknowledgement

The author is indebted to David Leech and Andrew Wang for comments on previous drafts.

Notes

1. Early studies at the microeconomic level (individual innovations) are Griliches (1958) and Mansfield *et al.* (1977). This literature has been reviewed by Griliches (1995) and Hall (1996).
2. Because of this dichotomy in underinvestment phenomena, government policies fall into two major categories—tax incentives and direct funding. Aggregate underinvestment in R&D is usually the result of broad capital market failure characterized by excessive aversion to risk and/or the lack of appropriate institutional mechanisms for funding high-risk investments (Hall and van Reenen, 2000; Hall, 2002). The appropriate policy response is a tax incentive and is distinctly different from the direct funding response that addresses

underinvestment phenomena in the form of suboptimal composition of R&D. However, tax incentives and direct funding mechanisms often are incorrectly considered interchangeable (Tassey, 1997).

3. Link and Tassey, (1987), Tassey, (1982, 1991, 1997), Nelson, (1992, 1993), Nadiri, (1993), Teubal *et al.* (1996).

4. The ''black box'' model has been discarded when a social objective (national defense, energy independence, health care) is the ultimate objective. However, when economic growth is the target, its philosophical hold has been much stronger.

5. Public good content gives an element an infrastructure character. As with any type of infrastructure, technological infrastructure can be defined in terms of institutional capabilities rather than the type of technical knowledge. The former approach has been taken by Justman and Teubal (1995). Here, the emphasis is on the characteristics of the specific infrastructure element, which, in turn, implies a set of optimal institutional capabilities.

6. An alternative four-phase R&D cycle has been proposed by Branscomb and Auerswald (2002). Their model (pp. 30–34) uses different labels for the four phases because their focus is on small R&D firms and the supply of venture capital. However, the end points of the four phases in their model are similar, so their framework is functionally identical to the one used here for corporate R&D generally.

7. See David *et al.* (2000) for a comprehensive review of this funding-based approach.

8. However, the distribution of government funds across areas of science can be a hotly debated issue, as has been the case with the skewed distribution of U.S. Government funding toward life sciences. For example, in fiscal year 2001, Federal funding for life sciences was about $13 billion greater than funding for all other areas of science combined.

9. Generic technology research is typically done in central corporate laboratories using techniques based on real options models to manage the research portfolio. Data on corporate investment in generic technologies is poor. What data are available indicate that companies conducting such research allocate anywhere from 1 to 12% of total R&D spending for this phase of R&D.

10. Kim and Morbougne (1997) provide evidence of the economic importance of investment in such technologies. R&D-intensive companies in the United States and Europe were surveyed to obtain sales and profits data on investments in incremental improvements in the current generation of technology (product line extensions) and in radically new products based on next generation (i.e., new generic) technologies. For the average firm in the survey, product line extensions dominate both in terms of number and sales. This result is hardly surprising, as companies focus most of their resources on extracting value from their current technology portfolios. However, a majority of profits were found to be attributable to the relatively few ''discontinuous'' (new generic technology-based) innovations.

11. Examples are measurement and test methods, process and quality control techniques, evaluated scientific and engineering data, and the technical basis for product interfaces (Tassey, 1997). As quasi-public technology goods, they are co-supplied by industry and government. In the United States, the major government source is the National Institute of Standards and Technology (NIST).

12. The typical high-tech industry depends on a large number of infratechnologies whose collective economic impact is substantial. For example, a NIST study (Finan, 1998) estimated that the U.S. semiconductor industry would spend $5.5 billion in 2001 on measurement (an important type of infratechnology, much of which ends up as industry standards).

13. In other words, firms are largely ''takers'' with respect to the longer-term and higher-impact ''natural trajectories'' (Nelson and Winter, 1977) that sustain high R&D intensities and these trajectories usually derive to a significant extent from external sources (Klevorick *et al.*, 1995). This literature implies the model proposed here. Once a technology life cycle is initiated (i.e., proof of concept is established), subsequent applied R&D investments are driven by the generic technology. Even though the generic technology can itself evolve to a degree over successive sub-cycles, it provides the platform that determines the trajectory of applications.

14. Case studies of the major technologies driving modern economies show this pattern quite clearly; for example, digital computers (Flamm, 1988), network communications (National Research Council, 1999), and biotechnology (Henderson, *et al.*, 1999).

15. At one extreme, Mansfield *et al.* (1971), Abernathy and Utterback (1978) and Utterback (1979) provide a detailed framework and analysis of the impact of the evolution of individual technologies on specific product life cycles. At the other, proponents of broad-based ''long waves'', such as Mensch (1979) and Graham and Senge [1980], have documented the bunching of related technologies that appear after a period of advances in the underlying science and when existing capital stocks (and their associated technologies) have significantly depreciated.

16. As pointed out earlier, the conventional approach to microeconomic analysis of analyzing R&D investment by phase (basic, applied, development) is inadequate because it ignores the output of each phase and the implications for investment in subsequent or complementary phases.

17. For simplicity, Figure 1 omits feedback loops, which are critical sources of information relevant for investment decision making over a technology's life cycle.

18. Mansfield *et al.* (1971) made a major contribution to the understanding of R&D investment patterns by disaggregating risk into three types: technical, commercialization (innovation), and market. The conceptual framework developed in this paper considers commercialization and market risk to be the same.

19. A distinction should be made between ''uncertainty'' and ''risk.'' The two terms are frequently used interchangeably, but they have distinctly different meaning. Risk is the probability of not achieving a technical or market goal; it implies a known probability distribution (for each objective). Uncertainty refers to the lack of knowledge of the probability distribution. In the R&D cycle, technical uncertainty is progressively reduced as knowledge is gained about the probability distribution associated with the technical objectives being pursued. Continued progress in defining this probability distribution and a similar distribution for market success will eventually permit the conventional decision tools of corporate

finance to be applied and a risk-adjusted expected rate of return estimated (based on the product of technical and market risk estimates).

20. Software is a prime example. A NIST study by RTI International (2002) estimated that limited and ineffective testing of software during the R&D stage costs the U.S. economy at least $60 billion per year, with the users of software absorbing over 60% of these costs. This estimate is for direct costs only (the costs of detecting and fixing errors) and does not include estimates of the ultimate losses in revenues and value added.

21. Kash and Rycroft (1998) define complexity as the degree to which one individual cannot have a total grasp of all elements of a technology. The implications are that communications (including risk assessments) among specialists is important and research coordination is hard to achieve. Locating and accessing technical infrastructures also becomes more difficult.

22. For example, a company may develop a drug candidate that is effective in terms of the intended impact, say, blocking a specific RNA pathway that influences unwanted protein production by a tumor, only to discover after multiple phases of expensive clinical trials that the specific pathway (one of many within the tumor cell) is not solely or even particularly instrumental in the growth of the tumor. The market failure results from the fact that the generic technology research on the relationships between different pathways and tumor growth ("antisense" technology) is too broad and complex for individual companies to undertake and involves significant spillovers. Moreover, the associated infratechnologies are frequently not adequate, reducing the efficiency (increasing the time and expense) of conducting this research. Companies, therefore, often guess at the underlying generic technology (the overall pathway structure and the interacting roles of all proteins), as well as the infratechnologies for mapping these relationships. The result that the efficiency of R&D at the industry level is reduced. The lack of an explicit disaggregated R&D model such as the one presented here appears to have led to underinvestment in generic technologies and subsequent wasted efforts by biotechnology firms in applications (drug) development.

23. Using biotechnology as an example is particularly revealing because it is heavily supported by government funding, but the complexity of the technology still appears to result in mismatches in the rates of progress among the three major technology elements identified in Figure 1.

24. This phenomenon can result from several factors: (1) technical problems that typically are present in newly commercialized technologies; (2) higher cost due to a manufacturing process that has not been optimized for the new technology (small initial markets do not provide sufficient incentive or cash flow to invest in the necessary process R&D); and (3) interfaces between the new technology and other components in the broader technology system are typically not defined (i.e., as standards) early enough in the technology's life cycle, thereby raising system integration (and, in effect, product) costs.

25. For example, reaching the physical limits of silicon-based semiconductor technology has been predicted for some time, but advances in the underlying generic technology keep pushing that occurrence farther into the future (new sub-cycles keep appearing), thereby raising the risk of investing in radically new semiconductor technologies such as gallium arsenide.

26. From an economic growth policy perspective, one set of firms replacing another is not a negative result under "creative destruction" models of technology-based competition. However, if the new set of firms resides in another economy, a loss of domestic value added occurs in the first economy and it experiences a reduction in its growth rate.

27. Of course, having the generic technology in place does not guarantee market success for the innovating firm or even the entire domestic industry. The video cassette recorder (VCR) is one of the best known examples, but there are many others. A major type of semiconductor manufacturing equipment called a stepper was invented in the United States, but market share is now almost totally Japanese. Oxide ceramics, which every commercial wireless communication system incorporates, was discovered in the United States but Japanese industry today dominates commercial markets. See Tassey (1999, pp. 29–31) for additional examples.

28. One distinction between these two types is that price spillovers occur through the market mechanism, while knowledge spillovers do not (Jaffe, 1996).

29. For example, Jorgenson and Stiroh (2000) and Stiroh (2001) find little evidence of production spillovers from purchases of IT capital goods.

30. In particular, some price spillovers are desirable if users are to benefit substantially from innovations. Thus, market dynamics are relied upon to effect some reasonable distribution of benefits.

31. See Dasgupta and David (1994).

32. Adapted from Tassey (1997). The figure assumes that the minimum expected SRR for an R&D project is at least equal to the PRR (the SRR is above the 45° line). This assumption follows Jaffe (1996, 1998), who used a similar diagram to analyze the relationships between spillovers and rates of return on government funding decisions. See also Link and Scott (2001, pp. 776–777). This assumption holds over the technology life cycle, even though new technologies often impart negative externalities to the technologies they are replacing during the transition period (see Figure 2 and associated discussion). Note that the range of expected SRR (vertical arrows) straddles the social HR. This simply reflects the fact that government R&D investment entails risk, just as private investment does.

33. Conventional underinvestment analysis simply makes a judgement that the PRR has been suppressed relative to the SRR. This is an inaccurate approach for two reasons. First, the role of the private HR is not accounted for and, second, the dynamics of the technology life cycle is ignored, one aspect of which is that the PRR changes over this cycle. See the previous discussion on timing (pp. 96–97) and the next section.

34. Unlike direct funding of R&D, a tax credit is not applied to particular projects such as C, but rather provides an incentive for all such projects by nudging their expected RoR above the corporate HR. A tax incentive therefore has the advantage of being technology neutral. However, this

mechanism has the disadvantage of significant leakage (excess tax expenditures) because all projects (not only like C but also like D) can potentially benefit from the tax incentive—whether or not these projects require the incentive. The conditions under which each of these two policy mechanisms (direct funding and tax incentives) should be used are discussed in more detail in Tassey (1997).

35. The relationship between technical and market risk was defined and examined empirically by Mansfield *et al.* (1971, pp. 50–54). The probability of commercialization of an R&D investment is determined by the product of the probability of technical completion and the probability of commercialization (given technical completion).

36. For example, Mansfield *et al.* (1971, p. 53) found that the probability of commercialization was higher for "large or medium" technical advances than for "small" ones.

37. The position and slope of the risk reduction curve vary depending on a number of R&D efficiency factors, in particular, the availability of a range of infratechnologies.

38. This phenomenon is recognized to varying degrees by virtually every industrialized nation, as evidenced by the existence of technology research support programs. Examples are the Framework Program in Europe and NIST's Advanced Technology Program in the United States.

39. Such estimates are averages across all technologies and are useful as general targets. However, a macroeconomic target does not take into account varying levels of technological opportunity among individual technologies.

40. Griliches (1988, 1995), Mansfield (1991b), Nadiri (1993), Hall (1996), and Cameron (1998) have assessed the literature on the rates of return from R&D investment and find both the private rate of return (PRR) and social rate of return (SRR) to be high relative to other types of investment.

41. For further discussion of these tools and applications to analyses of specific government R&D programs, see Tassey (2003).

42. The SRR for the generic technology is also likely to be higher when the contributions of supplying industries are taken into account. See Schankerman (1981), Hall and Mairesse (1995), and Jones and Williams (1998). This relationship increases the importance of the supply chain (as opposed to a single industry) as the unit of analysis for R&D policy.

43. See Angelis (2000) and Boer (2000).

44. Biotechnology is a prominent example of such a situation, as inefficiencies in technical infrastructure supporting both R&D and manufacturing abound and collectively are significantly restraining the evolution of this industry. See Lagace (2003).

45. Once a research program has been completed, retrospective impact assessments use just this approach. See Tassey (2003) for detailed descriptions of impact assessment methodologies and examples of both retrospective and prospective impact studies of NIST research programs.

46. These studies can be accessed at http://www.nist.gov/public_affairs/budget.htm.

47. Approximately $1 billion of NIH research funding goes directly to industry each year.

48. National Research Council (1999, p. 2).

49. National Research Council (1999, p. 169).

50. For example, voice traffic is handled by a continuous connection (a circuit) for the duration of the transmission, while computers communicate in bursts. Unless a number of these bursts or "calls" can be combined on a single transmission path (seldom the case in complex, high-capacity transmission systems), line and switching capacity is wasted. Telecommunications engineers were primarily interested in improving the voice network and were skeptical of alternative technologies. Thus, although telephone companies provided point-to-point communications in the ARPANET, the industry switching technology was not used. National Research Council (1999, p. 172).

51. National Research Council (1999, p. 9).

52. Elias and Hinzmann (1996, pp. 2, 5).

References

Abernathy, W. and J. Utterback, 1978, 'Patterns of Industrial Innovation,' *Technology Review* **80** (June-July), 40–47.

Angelis, D., 2000, 'Capturing the Options Value of R&D,' *Research Technology Management* **43** (July–August), 31–34.

Arrow, K., 1962, 'Economic Welfare and the Allocation of Resources for Invention,' in *The Rate and Direction of Inventive Activity: Economic and Social Factors,* Princeton: Princeton University Press, pp. 609–625.

Boer, F.P., 1998, 'Traps, Pitfalls and Snares in the Valuation of Technology', *Research Technology Management* **41** (September–October), 45–54.

Boer, F.P., 2000, 'Valuation of Technology Using 'Real Options,'' *Research Technology Management* **43** (July–August), 26–30.

Branscomb, L. and P. Auerswald, 2002, *Between Invention and Innovation: An Analysis of Funding for Early-Stage Technology Development,* National Institute of Standards and Technology, Advanced Technology Program (November).

Caballero, R. and A. Jaffe, 1993, 'How High Are the Giants Shoulders: An Empirical Assessment of Knowledge Spillovers and Creative Destruction in a Model of Economic Growth' in O.J. Blanchard and S. Fischer (eds.), *NBER Macroeconomics Annual 1993,* Cambridge, MA: The MIT Press.

Cameron, G., 1998, 'Innovation and Growth: A Survey of the Empirical Evidence' Neufield College, United Kingdom (July).

Cannon, P., 2002, 'Tell Your Legislator the Truth about R&D,' *Research Technology Management* **45** (November–December), 9–11.

Coase, R., 1960, 'The Problem of Social Cost,' *Journal of Law and Economics* **3**, 1–44.

Coase, R., 1992, 'The Institutional Structure of Production,' *American Economic Review* **82**, 713–719.

Cohen, W. and D. Levinthal, 1989, 'Innovation and Learning: The Two Faces of R&D,' *The Economic Journal* **99** (September), 569–596.

Darby, M. and L. Zucker, 2003, 'Grilichesian Breakthroughs: Inventions of Methods of Inventing and Firm Entry in

Nanotechnology', NBER Working Paper 9825 (July). Cambridge, MA: National Bureau of Economic Research.

Dasgupta, P. and P. David, 1994, 'Towards a New Economics of Science,' *Research Policy* **23,** 477–521.

David, P., B. Hall, and A. Toole, 2000, 'Is Public R&D a Complement or Substitute for Private R&D? A Review of the Econometric Evidence,' *Research Policy* **29** (April), 497–529.

Dosi, G., 1988, 'Source, Procedures, and Microeconomic Effects of Innovation,' *Journal of Economic Literature* **26** (September), 1120–1171.

Elias, E. and B. Hinzmann, 1996, *The Hype Curve: New Technology as the Modern-Day Gold Rush*. Menlo Park, CA: SRI International.

Finan, W., 1998, *Metrology-Related Cost in the U.S. Semiconductor Industry, 1990, 1996, and 2001* (NIST Planning Report 98-4). Gaithersburg, MD: National Institute of Standards and Technology.

Flamm, K., 1988, *Creating the Computer: Government, Industry, and High Technology*. Washington, DC: The Brookings Institution.

Graham, A. and P. Senge, 1980, 'A long-Wave Hypothesis of Innovation,' *Technological Forecasting and Social Change* **17** (August), 283–311.

Griliches, Z., 1958, 'Research Costs and Social Returns: Hybrid Corn and Related Innovations,' *Journal of Political Economy* **46** (October), 419–431.

Griliches, Z., 1979, 'Issues in Assessing the Contribution of Research and Development to Productivity Growth,' *Bell Journal of Economics* **10**, 92–116.

Griliches, Z., 1988, 'Productivity Puzzles and R&D: Another Non-explanation,' *Journal of Economic Perspectives* **2**, 9–21.

Griliches, Z., 1992, 'The Search for R&D Spillovers,' *Scandinavian Journal of Economics* **94** (1992, Supplement), 29–47.

Griliches, Z., 1995, 'R&D and Productivity: Econometric Results and Measurement Issues,' in P. Stoneman, (ed.), *Handbook of the Economics of Innovation and Technological Change*, Malden, MA: Blackwell Publishers, Ltd. pp. 52–89.

Hall, B., 1996, 'The Private and Social Returns to Research and Development,' in B. Smith and C. Barfield, (eds.), *Technology, R&D, and the Economy*. Washington, DC: Brookings Institution and the American Enterprise Institute, pp. 140–183.

Hall, B., 2002, 'The Financing of Research and Development,' NBER Working Paper No. 8773 (February). Cambridge, MA: National Bureau of Economic Research.

Hall, B., and J. Mairesse, 1995, 'Exploring the Relationship between R&D and Productivity in French Manufacturing Firms,' *Journal of Econometrics* **65**, 263–293.

Hall, B. and J. van Reenen 2000, 'How Effective are Fiscal Incentives for R&D? A New Review of the Evidence,' *Research Policy* **29** (April), 449–469.

Henderson, R., L. Orsenigo, and G. Psiano, 1999, 'The Pharmaceutical Industry and the Revolution in Molecular Biology: Interactions among Scientific, Institutional, and Organizational Change,' in D. Mowery and R. Nelson, (eds.), *Sources of Industrial Leadership,* Cambridge, MA: Cambridge University Press.

Jaffe, A., 1996, *Economic Analysis of Research Spillovers: Implications for the Advanced Technology Program* (report prepared for NIST's Advanced Technology Program). Gaithersburg, MD: National Institute of Standards and Technology.

Jaffe, A., 1998, 'The Importance of 'Spillovers' in the Policy Mission of the Advanced Technology Program,' *Journal of Technology Transfer* **23**, (Summer), 11–19.

Jones, C. and J. Williams, 1998, 'Measuring the Social Returns to R&D,' *Quarterly Journal of Economics* **113** (November), 1119–1135.

Jones, C. and J. Williams, 2000, 'Too Much of a Good Thing?: The Economics of Investment in R&D,' *Journal of Economic Growth* **5** (March), 65–85.

Jorgenson, D. and K. Stiroh, 1999, 'Raising the Speed Limit: U.S. Economic Growth in the Information Age', *Brookings Papers on Economic Activity* **31**, 125–211.

Justman, M. and M. Teubal, 1995, 'Technological Infrastructure Policy (TIP): Creating Capabilities and Building Markets,' *Research Policy* **24** (March), 259–281.

Kash, D. and R. Rycroft, 1998, 'Technology Policy in the 21st Century: How Will We Adapt to Complexity?' *Science and Public Policy* **25** (April).

Kim, W. Chan and R. Mauborgne, 1997, 'Value Innovation: The Strategic Logic of High Growth', *Harvard Business Review* **75** (1) (January–February), 102–112.

Klevorick, A., R. Nelson, and S. Winter, 1995, 'On the Sources and Significance of Inter-Industry Differences in Technological Opportunities,' *Research Policy* **24** (March), 185–205.

Lagace, 2003, 'Innovation and Change: Making Biotech Work as a Business,' *HBS Working Knowledge* (http://hbswk.hbs.edu/item.jhtml?id = 3247&t = innovation).

Link, A., and G. Tassey, 1987, *Strategies for Technology-based Competition,* Lexington, MA: D.C. Heath.

Link, A., and J. Scott, 2001, 'Public/Private Partnerships: Stimulating Competition in a Dynamic Market,' *International Journal of Industrial Organization* **19**, 763–794.

Mansfield, E., 1968, *Industrial Research and Technological Innovation*. New York, NY: W. W. Norton.

Mansfield, E., 1980, 'Basic Research and Productivity Increase in Manufacturing,' *American Economic Review* **70** (December), 863–873.

Mansfield, E., 1991a, 'Academic Research and Industrial Innovation,' *Research Policy* **20**, 1–12.

Mansfield, E., 1991b, 'Social Returns from R&D: Findings, Methods, and Limitations,' *Research Technology Management* **34** (November–December), 24–27.

Mansfield, E., J. Rapoport, J. Schnee, S. Wagner, and M. Hamburger, 1971, *Research and Innovation in the Modern Corporation*, New York: W. W. Norton.

Mansfield, E., J. Rapoport, A. Romero, S. Wagner, and G. Beardsley, 1977, 'Social and Private Rates of Return from Industrial Innovations,' *Quarterly Journal of Economics* **91**, 221–240.

Mansfield, E., A. Romeo, M. Schwartz, D. Teece, S. Wagner, and P. Brach, 1982, *Technology Transfer, Productivity, and Economic Policy*, New York: W. W. Norton.

McMillan, S., F. Narin, and D. Deeds, 2000, 'An Analysis of the Critical Role of Public Science in Innovation: The

Case of Biotechnology,' *Research Policy* **29** (January), 1–8.

Mensch, G. 1979, *Stalemate in Technology*, Cambridge, MA: Ballinger.

Miotti, L. and F. Sachwald, 2003, 'Co-operative R&D: Why and with Whom? An integrated Framework of Analysis,' *Research Policy* **32** (September), 1481–1489.

Mohnen, P., 1996, 'R&D Externalities and Productivity Growth,' *STI Review* **19**, 39–66.

Nadiri, I., 1993, 'Innovations and Technological Spillovers,' (NBER Working paper No. 4423), New York: National Bureau of Economic Research (August).

National Research Council, 1999, *Funding a Revolution: Government Support for Computing Research*, Washington, DC: National Academy Press.

Nelson, R., 1959, 'The Simple Economics of Basic Scientific Research,' *Journal of Political Economy* **67**, 297–306.

Nelson, R., 1992, 'What Is 'Commercial' And What Is 'Public' About Technology, And What Should It Be?,' in N. Rosenberg, R. Landau, and D. Mowery (eds.), *Technology and the Wealth of Nations*, Stanford, CA: Stanford University Press.

Nelson, R., (ed.), 1993, *National Systems of Innovation*, New York: Oxford University Press.

Nelson, R., 1995, 'Recent Economic Theorizing about Economic Change,' *Journal of Economic Literature* **30** (4), 48–90.

Nelson, R. and S. Winter, 1977, 'In Search of a Useful Theory of Innovation,' *Research Policy* **6**, 36–76.

Pigou, A.C., 1932, *The Economics of Welfare*, 4th ed. London: Macmillan.

RTI International, 2002, *The Economic Impacts of Inadequate Infrastructure for Software Testing* (NIST Planning Report 02-3). Gaithersburg, MD: National Institute of Standards and Technology. (http://www.nist.gov/director/prog-ofc/report02-3.pdf)

Romer, P., 1990, 'Endogenous Models of Technological Change,' *Journal of Political Economy* **98**, S71–S102.

Rosenberg, N., 1982, *Inside the Black Box: Technology and Economics*, Cambridge, UK: Cambridge University Press.

Ruffin, R., 1996, 'Externalities, Markets, and Government Policy,' *Federal Reserve Bank of Dallas Economic Review* (Third Quarter), 24–29.

Schankerman, M., 1981, 'The Effect of Double Counting and Expensing on the Measured Returns to R&D,' *Review of Economics and Statistics* **63**, 454–458.

Schmookler, J., 1966, *Invention and Economic Growth,* Cambridge, MA: Harvard University Press.

Stiroh, K., 2001, 'What Drives Productivity Growth,' *Federal Reserve Bank of New York Economic Policy Review* **7** (March).

Tassey, G., 1982, 'Infratechnologies and the Role of Government,' *Technological Forecasting and Social Change* **21**, 163–180.

Tassey, G., 1991, 'The Functions of Technology Infrastructure in a Competitive Economy,' *Research Policy* **20**, 345–361.

Tassey, G., 1997, *The Economics of R&D Policy,* Westport, CT: Greenwood Publishing Group (Quorum Books).

Tassey, G., 1999, *R&D Trends in the U.S. Economy: Strategies and Policy Implications* (NIST Planning Report 99-2), Gaithersburg, MD: National Institute of Standards and Technology. (http://www.nist.gov/director/prog-ofc/report99-2.pdf)

Tassey, G., 2000, 'Standardization in Technology-Based Markets,' *Research Policy* **29**, 587–602.

Tassey, G., 2003, *Assessing the Impacts of Government Research: Methods, Results, and Interpretations*. (NIST Planning Report 03-1), Gaithersburg, MD: National Institute of Standards and Technology (June).

Tassey, G., 2004, 'Policy Issues for R&D Investment in a Knowledge-Based Economy,' *Journal of Technology Transfer* **29**, 153–185.

Teubal, M., D. Foray, M. Justman, and E. Zuscovitch, (eds.), 1996, *Technological Infrastructure Policy: An International Perspective*, Boston: Kluwer.

Utterback, J., 1979, *The Dynamics of Product and Process Innovation in Industry*, Cambridge, MA: The MIT Center for Policy Alternatives (February).