What It Takes To Be a Leader in Both Basic Science and Technological Progress

PAUL ROMER

New York University

Statement for House Budget Committee Hearing on Federal R&D

Weds July 8, 2020

Contents

1 The Weakness Revealed By the Pandemic **1**

2 Science, Technology, and That Trade-Off Thing **1**

3 Listening to Other Voices **5**

4 General Principles **7**

4.1 People are what matter, not papers or patents 8

4.2 Achieve robustness via competition 8

4.3 Protect scientific integrity by separating the roles of decision-maker and fact-finder 9

5 Appendix A **11**

6 Appendix B **67**

7 Appendix C **101**

1 The Weakness Revealed By the Pandemic

In a conversation about testing for the SARS-Cov-2 virus, Kari Stefansson, the founder of deCode Genetics, the firm that has been doing population scale genomics in Iceland and which turned its attention to monitoring the spread of the virus, said to me that the technology that every country in the world is now using to identify the presence of the virus was developed in universities based in the US. The puzzle was why the US has been less effective than other countries in taking practical advantage of its basic scientific advances and using them to control the pandemic.

It is a good question. In 2019, a team at Johns Hopkins evaluated how well prepared various nations were to manage a pandemic. They concluded that the United States was the best prepared nation in the world. The United Kingdom was number two. (The results are available at https://www.ghsindex.org/.)

So the question is not simply why the US response was so inadequate, but also why the people tasked with thinking about these issues failed to anticipate how poorly we would do.

For members of the academic community, the convenient answer to this question is that blame for our failed national response lies with our political leaders. But as the warning sign at rail crossing in France cautions: Be careful. One train may hide another.

2 Science, Technology, and That Trade-Off Thing

In the wake of the pandemic, we need to revisit the conversation about the failure of our national innovation system to apply scientific insights and capture the practical benefits we desire. The usual frame for this discussion posits a private business sector that "failed to take up" the insights discovered by the academic scientists.

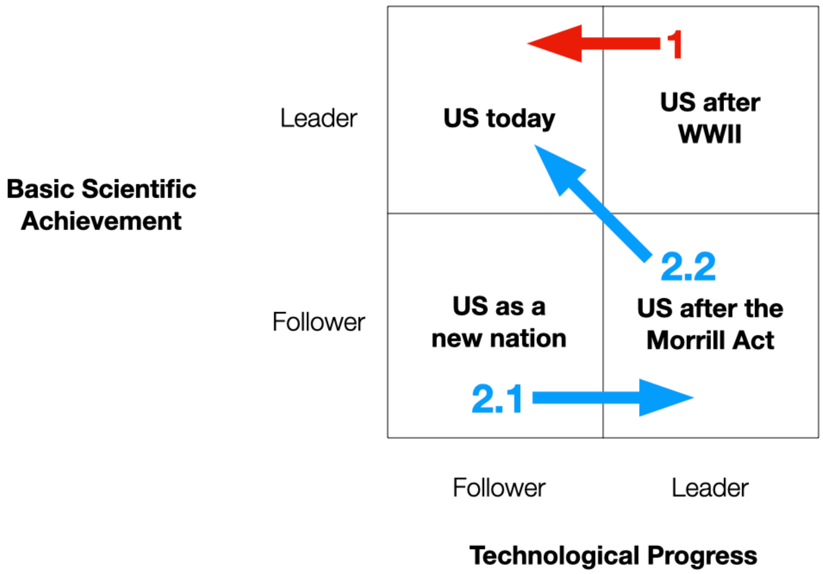


Figure 1: Two narratives about the United States

In the current iteration of this discussion, we need to consider, skeptically, a new version of this narrative in which academics claim that it was the government that "failed to take up" the academic insights.

Many commentators have observed that national leadership in basic science is neither necessary nor sufficient for leadership in technological progress. In terms of the two-by-two table in Figure 1, this means that a nation can end up in either of the off-diagonal boxes.

There is, of course, general agreement that the United States should aspire to leadership along both dimensions and a lingering presumption that leadership in science should naturally translate into leadership in technology. In his famous report *The Endless Frontier*, Vannevar Bush laid out what has come to be known as the "linear model" in which new ideas are produced by basic science, refined by applied science, and translated in practical benefits by product development. According to this model, the rate of new discoveries is limited by the rate of new ideas from basic science which puts new possibilities onto its conveyor belt. Were this model correct, leadership in basic science would be necessary for leadership in technology but not sufficient. The downstream activities are also required. This helps explain the defense by basic scientists, that others have not "taken up" their insights. The problem lies downstream.

There are instances of discovery that support the linear model. Basic mathematical research in number theory translated into applied cryptography. But there are many other instances which do not. The basic science of thermodynamics did not lead to the steam engine as an application. The steam engine lead to thermodynamics as physicists tried to figure out what it was that the problem solvers were doing. As a result, no student of science and technology is willing to defend the linear model.

The historical evidence also undermines the linear model. Nations do find themselves in the off-diagonal boxes, leading in one dimension but not the other. At different points in time, the United States has found itself in both off-diagonal boxes. More controversially, I want to suggest that we may not ever have been in a stable position of dual leadership in the box in the upper right.

There is a narrative about the current state in the US that is informed by the linear model which starts in the post-WWII era, one suggested by the red arrow and the number 1 in Figure 1. According to this narrative, the US was a leader along both dimensions, but that something went wrong in the downstream processes.

But to tell the full story of the United States, we need to go back to the 18th century, when this country was a follower along both dimensions. The pivotal change came with the passage of the Morrill Act of 1862, which established the principle of having a land grant university in every state. The second pivot comes after WWII, is marked by Bush's report. The driving force in this new era was a dramatic expansion of Federal support for basic scientific research and a turn away from the mandate of the Morrill Act: "to teach such branches of learning as are related to agriculture and the mechanic arts, in such manner as the legislatures of the States may respectively prescribe." The post-WWII system was a shift from teaching to research, from problem solving to fundamental inquiry, and state to federal control. To illustrate this practical focus, my late Stanford colleague, Nathan Rosenberg, used to remind his students that the football team at Purdue is known as the Boilermakers because students in its engineering school did research on a fully operational steam locomotive railroad engine kept at the school in the 1890s. Another example that Nate used to emphasize the role of local control over the lines of inquiry at the land grant schools was the development by researchers at the University of Minnesota of the pelletization process that was essential for full exploitation of the state's iron ore deposits.

The possibility suggested in the figure by the blue arrows labeled with the number 2 is that the story of the US after WWII may best be understood as a traverse from the box in the lower right to the box in the upper left, along which, for a period of time, we continued to take advantage of our capital in delivering technological progress without recognizing that it was depreciating.

Another of Nate's revealing illustrations of the sharp change in our national innovation system after WWII comes from Chemistry. In the first few decades of twentieth century, scientists working in the US received almost no Nobel prizes. In Chemistry, the leading nations were Germany and the UK. But it was in this era that the United States moved into a position of worldwide dominance of the petrochemical industry. Moreover, it is not hard to understand how the US was so successful. Our practically oriented universities, which focused on their educational mission, introduced a new discipline, chemical engineering, that was taught in a new type of professional school.

So for me, the story of the US in the last 120 years is the story of the traverse from technological leadership to scientific leadership. From this perspective, the challenge now is not to invent a new strategy for being the worldwide technological leader. It is to revive the strategy that worked before, and in so doing, to find a better balance between the policies that foster basic scientific leadership and the ones that encourage technological leadership. This nation can do both, but it will not do both if the advocates for basic science always get their way in any policy decision.

In uncharted seas, maps once warned "Here, there be dragons." Figure 1 warns that "Here, there be tradeoffs."

3 Listening to Other Voices

There are many details behind the suggestions that I will outline. To keep my remarks short, I summarize very briefly two papers that I wrote on how to restore some of the advantageous features of our pre-WWII national innovation system. Then in the final section, I will offer three general principles that can guide other changes we could make.

The two papers, which are attached as Appendix A and B, try to create mechanisms that will feed insights from industrial users of technology and from aspiring students of science and engineering into the decisions on university campuses. Their insights should supplement the insights provided by the successful scientists who dominate in a system where universities compete for research funds provided by the federal government. Before those well-established scientist crank up their habitual attack on "meddling by outsiders," let me be clear that in these two papers, I do not suggest any encroachment on their domain of influence, nor any reduction in the funds that receive from the federal government. All I suggest is that we provide financial support to universities in some additional ways.

The way I propose for giving an industry some influence over decisions made on campus is to give it the power to levy a tax on all firms in the industry and to use the proceeds to support research and educational programs that will benefit the industry. The corresponding way to give students a voice is to bring back the system that prevailed before WWII, in which students who paid tuition could decide where they studied and what they studied. Under this system, universities competed for tuition revenue by introducing such new courses of study as chemical and electrical engineering. The way to do this today and simultaneously to encourage more US citizens who pursue a graduate education is for the federal government to create a large number of portable fellowships that are rewarded to the most talented undergraduates. These fellowships should cover both living expenses and generous tuition charges for three full years of graduate education. The fellowships should be portable in the sense that a recipient could use it to pursue any degree program in science and engineering at any university.

I know from the discussions that followed my paper on portable fellowships, which included discussions with staff and members of Congress that led to the introduction of a bill that provided for these fellowships, that the leading research universities and the advocates for basic science opposed this new additional source of funding. This is why I suggested before that the right decision for the nation might mean that the advocates for basic science do not get their way. The National Defense Education Act (NDEA), passed in 1958, created a fellowship program that was similar in spirit to the one I propose. Over time, that program was replaced by one that the professors in research universities, as the principle investigators on research grants, preferred. Instead of giving funds directly to students and letting them make their own decisions, the funding agencies gave money to professors, who supported graduate students by hiring them to work on the research projects that the professors wanted to work on. I am not persuaded, but a reasonable person could claim that the shift to unilateral control by principle investigators is the best way to support basic science. But what cannot be denied is that giving the professors unilateral control deprived aspiring scientists and engineers, many of whom would end up working in industry, of any say in the graduate educational programs that they could pursue on the nation's university campuses. It deprived the nation of the type of educational innovation that led to the creation of schools of chemical engineering.

Even if the Congress comes up with new money to provide the type of fellowships I propose that empower aspiring scientists and engineers, even if the Congress holds constant the funds that agencies can give as research grants to professors, I can assure you that the leaders of the scientific community, who are based in a few dominant research institutions, will not welcome these new fellowships. My message to them: "There be tradeoffs here." What is best for you may not be best for the nation.

4 General Principles

I will close by offering some general principles that could help guide us back toward leadership in both basic science achievement and in technological progress, back toward a system that is a hybrid between the one that brought the US to worldwide preeminence in industrial technology before WWII and the one that displaced when the federal government took control and gave so much autonomy to professors.

The first is a general principle that I inferred after my first inquiries into science and technology policy. The next two are more recent. They reflect the lessons I learned from the failure of our regulatory system in the run-up to the financial crisis. (In Appendix C, I attach a third paper prompted by the financial crisis.) These last two have, to my surprise, been reinforced by the failures that are now evident in our failed response to the pandemic.

4.1 People are what matter, not papers or patents

The most important outputs produced by the nations universities are well trained people. People working in many different organizations can write papers and apply for patents. The unique contribution that universities can make to the nation is to give the most talented young people a chance to acquire the skills, the confidence, the habits of mind that will allow some of them to make outstanding contributions either to basic science or to technological progress, or to both. The Morill Act and the NDEA rewarded universities for making these investments in people. Today, we should look for new ways to reward these investments once again and we should make sure that the educational programs offered by universities respond to the opportunities that students perceive.

4.2 Achieve robustness via competition

Part of the genius of the Morrill Act is that it did not try to create a world beating national university. It build a system of many different competing universities. After the introduction of the centralized federal role in WWII, this has evolved toward one dominated by a handful of winners. The NDEA included a system for allocating its fellowships that prevented the leaders of the day from growing stronger at the expense of the followers. In effect, the NDEA fellowship program included a competition policy that kept a few powerful institutions from dominating the national innovation system.

We should copy this feature of the NDEA fellowship program in any new system that supports universities. And we should extend the insight that motivated this approach and the design implicit in the Morrill Act. We should rely on the states. We should encourage many voices. One of the problems that we now see with the federal system for protecting health is that it gave monopoly powers to the FDA and the CDC. In so doing, we created what the engineers call single points of failure in our defense against viral pathogens. Instead of further concentrating power in these dominant agencies, we should have the courage to bet on the states as the Morrill Act did, to build up stronger departments of health in each of the states and defer to the states on such critical decisions as what types of tests to deploy for identifying a dangerous viral pathogen.

If we can trust state governments to run universities, we can trust them to regulate tests.

4.3 Protect scientific integrity by separating the roles of decision-maker and fact-finder

One of the most useful and under-appreciated innovations in the US system of governance is the division of labor between the FAA, which makes regulatory decisions about aviation, and the National Transportation Safety Board, which is responsible for establishing the facts after any accident.

There is an inevitable tendency for an agency that has to make technical decisions to report to the public a version of the facts that supports its decisions. These agencies turn into advocates for specific positions. In the process, they lose their scientific objectivity. During this pandemic, we have seen several important instances where agencies that were responsible for difficult real-time decisions that were central to our pandemic response – the CDC, the FDA, and the WHO – justified their decisions by presenting the public with a biased or misleading summary of the facts.

There are now four important questions on which informed observers now question the factual answers provided by these agencies:

1. Do masks limit the spread of the virus? 2. Do people who are infected by the SARS-CoV-2 virus but do not currently show symptoms cause a substantial

fraction of its transmission? 3. Does the virus spread through the air mainly via large droplets or also through smaller aerosol particles? 4. Does a policy of testing for the virus and isolating those who test positive reduce the spread of the virus?

I will not try to summarize the evidence that bears on these four questions. Nor will I try second guess the policy decisions by these agencies that were tied to the answers they provided to these four questions.

However, as someone who has spent a great deal of time trying to understand the facts and the positions taken by the different parties in the debates about the facts, I will report that there are credible scientists who now have grave concerns about the scientific integrity of the messages that these agencies have conveyed to the public.

The risk here is not just that a biased reading of the facts will lead to bad decisions or hinder reconsideration of those decisions as new evidence comes in. The far more troubling concern is that well informed people have stopped believing the assertions of fact that these agencies make and that their skepticism is fueling a broader distrust of authority by those who are not as well informed.

To make progress, we have to base our decisions on the facts. If voters do not trust scientific authorities, our democracy will not be able to base our public policy decisions on the facts.

There are many possible ways to divide up the responsibility for making decisions about public health between states and the Federal government. And many possible places where we could assign to some other, independent agency the role of finder of fact. What would not make sense is to create many different copies of the CDC or the FDA at the state level, each of which continues to have conflicting responsibilities to make decisions and provide a clear statement of an evolving pattern of facts to the public.

5 Appendix A

Implementing a National Technology Strategy with Self-Organizing Industry Investment Boards (Brookings Papers on Economic Activity, 1993, No. 2)

PAUL M. ROMER
University of California, Berkeley
National Bureau of Economic Research
Canadian Institute for Advanced Research

Implementing a National Technology Strategy with Self-Organizing Industry Investment Boards

THE MOST IMPORTANT LESSON from the study of research and development, economic growth, and the history of technology is that there are more ways to arrange the objects of the physical world than humans can possibly imagine. Ultimately, all increases in standards of living can be traced to discoveries of more valuable arrangements for the things in the earth's crust and atmosphere. The personal computer that I used to write this paper is made from almost exactly the same physical materials as the PC that I bought ten years ago—about thirty pounds of steel, copper, aluminum, plastic, and silicon, with bits of gold, iron oxide, and miscellaneous other elements mixed in. In my new PC these materials are arranged in a slightly different way that makes them about fifty times more useful than they were in the original configuration.

No amount of savings and investment, no policy of macroeconomic fine-tuning, no set of tax and spending incentives can generate sustained economic growth unless it is accompanied by the countless large and small discoveries that are required to create more value from a fixed set of natural resources. These discoveries are the product of a complicated set of market and nonmarket institutions that constitute what has been called a national innovation system. This paper considers both the economic

The author gratefully acknowledges comments from the discussants and the editors. This material is based on work supported by a National Science Foundation grant.

opportunities and the political risks inherent in attempts to strengthen this system. It proposes an institutional arrangement that could provide more financial support for innovative activity and direct it toward areas with large economic payoffs.

Even more impressive than the availability of a much faster PC to do word processing is the fact that my family drinks milk from a cow. If the milk cow did not exist, no one would ever believe that carbon, oxygen, hydrogen, and a few other types of atoms could possibly be assembled into a chemical refinery that automatically converts grass, water, and air into a nearly perfect liquid protein dietary supplement. This refinery operates with almost no human supervision, is mobile so it can search out its own inputs, can heal most mechanical failures, and can detect and neutralize any microscopic pathogens that enter the system. It even makes more than milk. It can build several replacement factories out of the same raw materials: grass, water, and air.

Compared with the cow, my PC has all the sophistication of a Tinkertoy. And if the particular arrangement of atoms that make a cow can come together through a blind process of mutation and selection—nature's version of trial and error—imagine how many other ways atoms can be arranged that are as astonishing and valuable as a microprocessor and a cow. The fundamental challenge in economic growth is to find these new arrangements.

This optimistic potential for exploiting what Vannevar Bush called "the endless frontier" of scientific and technological opportunity is limited only by the difficulty of organizing collective action.[1] It takes collective action to encourage discovery and sustain the free flow of ideas, and the political mechanisms used to undertake collective action suffer from serious inherent defects.

But people create new institutions, just as they discover new technologies. Policy innovators discover better ways to undertake collective action, just as scientists, engineers, and product designers discover better ways to arrange physical objects. Vannevar Bush did more than write a report about the endless frontier of science. He midwifed the birth of the National Science Foundation and peer-reviewed research grants for basic research at universities. He identified an important opportunity for collec-

1. Bush (1945).

tive action and then constructed an institutional arrangement to exploit this opportunity at minimal political cost.

Any contemporary discussion of a national technology strategy should be based on a balanced assessment of the potential benefits from collective action to spur growth and the risks inherent in undertaking collective action through the political process. Even die-hard free marketeers should concede that there is an opportunity for collective action to support some forms of basic research. On the other side, even the most ardent advocate of an activist government must admit that the nation's fifty-year experiment with an explicit industrial policy for the savings and loan industry shows how badly awry well-intentioned collective action can go when it is implemented through the existing political institutions.

Once the tension between the large opportunities for benefiting from collective action and limited political and institutional capabilities for undertaking effective collective action is recognized, it becomes clear that innovation in institutional mechanisms could be very valuable. Surely, the opportunity to experiment with new institutions is as great as the opportunity to experiment with new arrangements of physical objects. The standard dichotomy in economic policy debates—market exchange versus government intervention—does not capture the complexity of the kinds of social institutions already used to achieve common goals. It also fails to suggest the broad range of new institutional arrangements that could be tried. In its purest form the market is a mechanism for allowing independent action by all individuals, with no explicit coordination. The government is a mechanism for explicitly coordinating the actions of all people. Most economic activity is supported by institutional arrangements that are intermediate between these extremes.

After all, the modern capitalist economy is organized neither as a market nor as a government. It lets large numbers of people exploit the benefits of collective action through explicit, hierarchical coordination as members of a corporation, and in this sense large parts of a capitalist economy function like a government. But an economy with many firms subjects each to the discipline of competition, so in this sense the economy is organized more like a market. The diverse population of corporations is constantly changing, constantly finding new and different ways for large groups of people to work toward some collective end. Some corporations find new ways to structure themselves. Others die off and are replaced by new and better arrangements. What a national innovation system needs is something like this kind of dynamic, something that lets us take advantage of mutually beneficial coordinated action and that uses the pressures of competition and a market test to shut down ineffective institutional arrangements and to reward promising ones.

To show how such a process might work, I describe a specific policy proposal conceived with this end in mind. Whatever the final judgment is on this particular proposal, the general analytical point will, I hope, be clear. To encourage technical change, it is not enough to call on business leaders to be more innovative. Policymakers must themselves be willing to experiment with new institutional arrangements.

Self-Organizing Industry Investment Boards

This proposal specifies a process, not a specific policy. To see how it would work, consider a hypothetical example. Suppose you run the Acme Widget company and I run Consolidated Widgets. You and I and many other widget makers come together and decide that there are industry-wide opportunities and problems that independent action by individual firms cannot address. You think, for example, that university professors could be doing useful research on questions about the basic principles of widget design and manufacturing if they just had the funds and the incentive to do so. I believe that the upstream industry that manufactures widget-making equipment could be designing more useful specialized equipment.

Under the proposal outlined in this paper, widget makers could take collective action on these matters by following the steps specified in a general piece of enabling legislation. They would start by petitioning the secretary of commerce, giving the argument for provision of an industry-specific public good. The secretary would then hold a hearing to certify that the proposal for collective action addresses a genuine public need. If the proposal passes this test, an election would be held in which all widget manufacturers vote whether to levy, say, a 1 percent tax on widget sales. If a large enough fraction of the industry (measured as a fraction of total sales, as a fraction of the total number of firms, or some combination of the two) votes in favor of this initiative, a tax backed by the full force of law would be imposed on the entire industry. The proceeds from the tax would not, however, go to the government.

Part of the presentation to the secretary of commerce and then to the industry would outline plans for creating industry investment boards. You will take the initiative to organize an investment board that will fund university-based widget research. I will organize a board that supports the development of widget-making machinery in the upstream industry. Both of these boards will function as pass-throughs, accepting tax obligations from contributing firms and using these funds to support research in universities in the first case or development in upstream firms in the second. After the tax is passed, the two boards would solicit the tax funds from the firms in the industry. Suppose that your firm has sales of $100 million a year, so it has to decide where to allocate its $1 million in annual tax obligations. Because you are organizing the "University Research Board," you might decide to allocate all of the $1 million to it. My firm has sales of $200 million a year. I might decide to have my firm split its $2 million tax obligation, giving $1 million to the "Upstream Equipment Development Board" that I help organize and $1 million to the University Research Board. Each year the leaders of other firms would have to decide how to allocate their tax obligations between these two boards and any others that might be created. The amount they have to contribute is fixed by their sales and the tax rate. They are free, nevertheless, to decide which board receives their contributions. If they do not approve of the boards that exist, they are free to start a new one.

Each industry board would have a board of directors answerable to its contributing firms and would operate as a private, nonprofit foundation. The boards would be limited by the general terms of the enabling legislation but would otherwise have wide latitude to make decisions without any direct oversight or second-guessing by the executive, legislative, or judicial branches of government. A general limitation would require each board to invest only in common property resources that benefit the entire industry. For example, all specialized equipment developed by the Upstream Equipment Development Board in collaboration with the upstream industry would be available for sale to all widget-making firms on equal terms. All research funded by the University Research Board would be freely distributed in the tradition of open university-based science. Thus, neither board could fund research or equipment development that would be conducted in-house by a firm in the widget industry. All of the activities of the boards would have to be made public.

The enabling legislation should also specify that absolutely no tax funds could be used to support lobbying, public relations, or any kind of political activity. Also, no direct or indirect kickbacks or side payments to firms in the industry would be permitted. The tax rate used in each industry could vary, but a maximum tax, on the order of 2 percent of sales, would need to be specified in the enabling legislation. The legislation should also articulate the general principle that the tax should be a domestic consumption tax rather than a production tax. Units produced domestically for sale abroad would not be subject to the tax, but units produced abroad and sold domestically would. The legislation would also mandate equal treatment for all firms. Foreign firms would vote in the referendum, pay tax on their sales in the United States, allocate their tax obligations, and participate in the governance of the board or boards they support, in exactly the same fashion as domestic firms.

Suppose that other industry leaders come to believe that the most important problem facing the industry is an inadequate supply of postgraduate engineers with special training in widget-related design issues. Or suppose they decide that the most important investment would be to disseminate existing information about the principles of manufacturing to widget-making firms. Other firms might decide that junior college training of basic skills for current and future workers in the industry is the most important priority. An interested firm or group of firms could, at any time, petition the secretary of commerce, participate in public hearings, and then establish a Ph.D. Engineering Fellowship Board, an Extension and Diffusion Board, or a Worker Training Board. These new boards would compete with existing boards for the tax obligations of the different firms in the industry.

Over time, if contributing firms felt that the University Research Board was funding an incestuous network of scientists doing research of limited value to the industry, a competing research board could come into existence, or the funding firms could simply vote with their feet and take their tax dollars to other boards. In the worst case, if none of the boards were doing work that justified the cost, firms could use a periodically scheduled election to rescind the tax altogether.

The original proposal to the secretary of commerce would have to specify how the tax would be administered. Some boundary would have to be established, for example, between true widgets, which would be taxed, and near-widgets, which would not. An overall industry association, financed from an expressly limited share of total industry tax revenue, would have to be created to administer the tax system and perform the basic audit and information collection activities necessary to monitor compliance. Cases of explicit fraud or intentional noncompliance could be referred to an appropriate government agency with the power to impose fines, compel testimony, and, if necessary, undertake prosecution.

Precedents

This proposal may sound like political science fiction, but it is modeled on arrangements that already exist. The closest precedent is specified in the Agricultural Marketing Agreement Act of 1937, which outlines procedures for establishing arrangements called marketing orders.[2] Growers of a particular agricultural commodity can petition the secretary of agriculture to establish a market order on their behalf. Typically, the market order is approved if either two-thirds of the growers or growers representing two-thirds of output by volume support the proposal in the subsequent referendum. A new referendum is usually held every six years to gauge grower support. At any time, the secretary of agriculture can suspend any marketing order that is not operating in accordance with the aims of the enabling legislation.

Marketing orders designed to support research and development (R&D) and market promotion activities differ from the proposal in this paper only in that the funds raised by the tax automatically go to a single marketing board. There is no element of competition between boards and no free entry.

About three-quarters of the marketing orders in agriculture collect funds for research, development, and market promotion. (This is how the popular California raisin advertisements were financed, for example.) Marketing orders can also specify package and container regula-

2. Powers (1990).

tions, or size and grade regulations. These kinds of activities should be sharply distinguished from the notorious volume control regulations (that is, output restrictions) imposed in about half of all agricultural marketing orders. Quantity restrictions are clearly an activity that no economist would want to encourage, and they should not be tolerated, much less encouraged, in legislation that specifies the acceptable activities for the investment boards contemplated here.

This proposal has other, less obvious parallels with existing arrangements. For example, many (but not all) public electrical utilities contribute to the Electric Power Research Institute (EPRI), which finances research on a variety of industry-related matters. If local regulators let contributions to EPRI count as part of cost in figuring the utility's rate base, they are, in effect, using a government-sanctioned tax on electricity to support industry-related research. Just as in agriculture, however, there are no alternative research boards. Some observers maintain that EPRI may not be as effective in stimulating research as it once was or as it could be, perhaps because it faces no competition.

Before the breakup of the Bell System, Bell Labs was supported in a similar fashion. Each of the operating companies paid a few percent of total revenues to AT&T, which supported Bell Labs out of the proceeds. Again, to the extent that this contribution was built into the rate base that utility regulators allowed the operating companies, it amounted to a government-sanctioned tax used for industry-related research purposes. Because AT&T controlled the vast majority of telephone operating companies, free riding was not a problem.

Most observers judge Bell Labs to have been an extremely effective research organization. It made fundamental contributions to basic science. Information theory was created there. Radio astronomy was invented there. The background radiation that is the best evidence available for the "big bang" was discovered there. Bell Labs also produced high-quality scientific discoveries that have had enormous practical and commercial implications: the transistor, the laser, fiber optic transmission of information, and Unix, the first major computer operating system to run on computers made by many different manufacturers.

As evidence that firms can come together and take actions that are in the interest of the industry as a whole, one need look no further than the pharmaceutical industry. It recently persuaded the Food and Drug Administration (FDA) to raise the fees it levies when a company submits a drug for approval. The explicit understanding was that the FDA would use the additional revenue from this fee to hire more evaluators so that the agency could reduce the time it takes to reach a decision on drug approval.

The Economic Opportunities for Collective Action

It takes two arguments to make a case for this proposal. I must show why there are any important unexploited economic gains that require collective action. Then I must compare the potential gains with the political risks of implementing the system. This section outlines the theoretical argument for large potential gains from collective action. The next refers to evidence on their quantitative importance. Subsequent sections take up the political risks of this and other forms of technology policy.

If physical objects were the only economic goods, there would be little opportunity for collective action beyond the universally recognized need to establish a system of property rights. But as table 1 suggests, objects are not the only economic goods. They are not even the most important goods. The table presents a two-way classification of different kinds of economic goods. The left and right columns classify goods according to their costs of production. The horizontal dimension classifies them by the strength of the property rights available for each good.

The column on the left lists physical objects that are consumed directly or that provide services that are consumed. These kinds of things—land, fish, a worker's labor effort—are typically thought of as economic goods. The column on the right lists goods that can be represented as bit strings. They are all examples of information in the mathematical sense of the term. Whether it is literally a bit string (such as computer code or a digital musical recording) or is something represented in words or symbols that could be converted into bit strings (the design for a microprocessor, the operations manual for Wal-Mart stores, or the results of scientific investigation), any good in this column is a piece of pure information.

The technical terms from public finance for the types of goods in the two columns are rival goods and nonrival goods. The objects are rival

Table 1. Economic Attributes of Various Goods

Degree of control (percent)	*Rival goods (objects)*	*Nonrival goods (bit strings)*
100	Private goods: for example, a piece of unimproved land	An encoded satellite television broadcast
		A digital music recording
		The design for a microprocessor
	A car	Computer code
		The operations manual for Wal-Mart stores
	A worker's labor effort	General principles of chemical engineering
		Principles behind window-based graphical user interfaces
		The do-loop in computer programming
	Fish in the sea	
	Clean air	
0	Sterile insects used for pest control	Public goods: for example, basic research in physics

goods because you and I are rivals for their use. You can eat the fish or I can, but not both of us. A bit string is a nonrival good because once it has been produced, we are not rivals for its use. I can listen to the musical recording or take advantage of the software code without in any way diminishing its usefulness to you or anyone else.

The most important result of the work on "endogenous growth" during the last ten years has been the renewed attention devoted to these attributes of ideas as economic goods. The first conclusion that emerges from the theory of growth was clear even in neoclassical models: it is the production of nonrival goods that makes growth possible. The second conclusion, which has emerged only after a great deal of work, is that the usual invisible hand result applies only to an artificial economy in which nonrival goods are provided exogenously by nature. In a real economy, an inherent, unavoidable conflict exists between the incentives necessary to encourage the production of these goods and the incentives that lead to the optimal distribution of these goods, both to users and to the developers of other related nonrival goods. This means that private property rights and market exchange are not the perfect institutions for supporting growth. In fact, no simple description of the perfect institutional arrangement can exist. In any particular context, one must explicitly address the trade-offs both between the incentives for discovery and those for diffusion and between the limitations of market mechanisms and those of political mechanisms.

The two columns in the table correspond roughly to the distinction emphasized in the introduction between physical objects and intangible discoveries about new ways to arrange those objects. Farming, resource extraction, manufacturing, and distribution are all examples of actions that transform physical objects. For most people, these are the kinds of things that first come to mind when they think of economic activity. Discovery, in contrast, is how new, nonrival instructions are found for using rival goods more creatively. As the introduction suggests, discovery is where the real action is in economic life. Because nonrival goods are intangible, they are hard to measure. Moreover, they typically affect the economy in small increments that are hard to perceive from one year to the next. (The computer industry, where dramatic change can be seen in real time, is a notable exception.) But one needs to read only a little history to appreciate the profound cumulative effect of these nonrival goods—these discoveries—in almost all areas of economic activity. If the Earth were returned to the physical state that existed ten thousand years ago, wiping out all structures, physical capital, and civil engineering projects, but the total stock of accumulated knowledge were retained (in an exempted library where books and other records were kept), current standards of living would be recovered within a few generations. If the experiment were reversed, with the physical state of the world retained but the state of knowledge returned to what it was ten thousand years ago, our economic prospects would be much bleaker.

The sharp distinction that the table draws between rival and nonrival goods is rarely apparent in practice because real goods are almost always a mixture of the two. If you buy shrink-wrapped software, for example, you purchase a bundle that consists of rival objects (some floppy disks and a book) and a nonrival good (the legal right to use the bit string that encodes the computer program). The distinction between an object such as a floppy disk and an intangible such as the computer code stored on the disk is important because it points to a fundamental difference in the associated costs of production. The cost of producing the computer code is virtually all fixed cost, or "first copy" cost. Firms spend millions (sometimes hundreds of millions) of dollars finding just the right bit string, but once they have it, they can replicate it at essentially zero cost. Once a nonrival good is produced, it is a good with no opportunity cost. The cost of the floppy disk, in contrast, is almost entirely the constant cost of producing additional disks; almost, but not quite, because, of course, the floppy disk itself reflects a sophisticated design that required an important fixed-cost investment. The floppy disk is not a pure rival good, but unimproved land is, because it has no underlying design cost.

In contrast to the horizontal distinction between rival and nonrival goods, the vertical dimension in the table reflects the more familiar issue of property rights, appropriability, excludability, or, simply, control. It is possible to control different kinds of goods by maintaining physical possession, perhaps with the assistance of the legal system, or by keeping some valuable piece of information secret. Land is placed relatively high up in its column because it is rarely stolen and because the cost of maintaining control over land is small compared with its market value. An automobile is lower down the column, because cars are more frequently stolen and the total resources spent by society on maintaining control over cars is higher. Goods that are both object-like (that is, rival) and over which almost perfect control can be maintained are called private goods. Land is a private good, and a car is close enough for most analytical purposes.

Farther down the rival goods column are examples of objects (or services from objects) for which property rights are weaker and control is less complete. When a firm hires a worker, it purchases labor services during certain periods of time. Because labor effort is difficult to observe, the firm does not always get the good for which it has paid. The existing legal system cannot enforce this kind of contract at all well. It would be absurd to propose that the firm go to court and sue for compensation whenever the employee leaves work early. Firms therefore find ways to execute transactions with their workers that are cheaper than writing explicit legal contracts that list all contingencies and then litigating every dispute. This is what the transactions cost theory of the firm is all about.[3] It emphasizes that firms are institutions that provide alternative systems for establishing property rights and enforcing contracts. These institutions make possible investments and gains from trade that would otherwise be impossible to exploit.

At the bottom of the column are examples such as fish in the sea or the sterile insects that are released in agricultural settings to control pests. Fish and insects are objects. They are rival goods. The insects can neutralize fertile pests in my valley or yours, but not both. They unambiguously belong in the left-hand column. Control over these goods, however, is especially weak. If one farmer paid for and released sterile insects, the benefits would spill over onto the farms of his neighbors. Much of the economic analysis of policy is framed in terms of these spillovers—these instances of incomplete property rights. Control can be weak, so spillovers can be present, both for rival and nonrival goods.

The right column lists examples of pieces of pure information with various different degrees of control. An encrypted satellite television broadcast, the kind used to distribute movie channels to cable television systems across the country, is a pure nonrival good with very strong property rights. Musical recordings, microprocessor designs, and computer code are examples of goods for which control is less than perfect but which nevertheless are supplied by commercial firms able to sell their goods in the market at a significant markup over marginal cost. The kind of knowledge possessed by the Wal-Mart corporation about managing retail stores is only partly controlled—other firms copy what it does. That knowledge nevertheless represents an asset in which a firm can invest and on which it can earn a sizable return.

Next come goods over which control is even weaker—these include goods such as the general principles that are the basis for chemical engineering, the insights behind the notion of a window-based graphical user interface for computer programs, or basic ideas such as the do-loop in computer programming. Finally, at the bottom, are goods, such as results from research in physics, that are nonrival and whose use is virtually impossible to control. Economists call these pure public goods. (The term is somewhat misleading because not all public goods are provided by the government—think of charitable contributions to

3. See, for example, Coase (1988); and Williamson (1975).

support public television. Moreover, not all goods provided by the government are what economists call public goods—think of the sterile insects.)

One of the most important insights in economics is that if people have strong control over ordinary objects (those in the upper left corner of the table where the private goods are) and if there are many potential buyers and sellers, decentralized exchange between self-interested traders leads to efficient outcomes. This is the lesson of laissez-faire, or the invisible hand. If control over objects is weak, outcomes may be inefficient—as in the proverbial tragedy of the commons. Everyone will be a free rider. These results are the basis for the strong, almost unthinking reaction among some economists that moving up the columns of the table and increasing the degree of control will always enhance efficiency. If others cannot be excluded from enjoying the benefits of some service, the government may collect taxes and pay for the service, as it does for the release of sterile insects, but this kind of collective action is used only as a last resort. Stronger property rights, this line of thinking suggests, would always be preferable.

This intuition is correct for rival goods but simply does not apply to nonrival goods, where strong property rights are inherently associated with monopoly power. If there are strong property rights, there cannot be many sellers. If firms that produce nonrival goods are to avoid large losses, these goods must sell for a price that is higher than marginal cost. Marginal cost on bit strings is zero, but the initial fixed costs of producing them can be very large.

To see why extremely strong property rights might be a problem, imagine that Bell Labs had been given a nonexpiring, ironclad patent on the discovery of the transistor. Or even worse, imagine that such a patent had gone to an organization such as IBM or General Motors. Think of how different the digital electronics and consumer electronics industries would be if every inventor who improved on the design of the transistor and every person who applied the transistor in a new setting had to negotiate with one of these large, bureaucratic organizations for permission to proceed. Not only would the prices have been higher, but the rate of discovery of all the inventions that reduced the cost of transistors would have been lower. These kinds of discoveries, which were made by many different individuals in many different firms, have in just a few decades caused a millionfold reduction in the cost of producing a transistor.

A standard response to concern about strong patent rights is the claim that if GM had owned the patent on the transistor, it would have had an incentive to do everything right—to make all the right inventions and to agree to efficient contracts with outsiders. The relevant retort is that GM has had an incentive to find innovative ways to design and build high-quality automobiles and has not always succeeded in doing so. In the car business consumers are protected by the presence of a diverse set of competing manufacturers. But with an effective patent, all of the eggs would necessarily be placed in a single corporate basket.[4]

Or consider the computer software industry, an area of economic activity where producers are exquisitely aware of the need for some degree of intellectual property rights and monopoly power. Many thoughtful participants have also recognized that property rights can be too strong. Monopoly power can impose serious distortions and negotiation costs. Imagine, they say, if someone had been able to obtain a long-lived patent on the do-loop or the blinking cursor.

If people were pathologically honest and compulsively followed instructions, the economic problem of producing nonrival goods would be easy to solve. (This definitely is political science fiction.) In this kind of world, undertaking collective action would pose no problem. Everyone would diligently search out new opportunities for discoveries as they went about their other activities and would report all they learned to a central, coordinating agency. The government would direct a subset of these people to do the R&D necessary to take advantage of the most promising opportunities. The government could request that everyone else contribute a share of their income to the researchers who produced the desired nonrival goods—the software, the movies, the music, the books, the microprocessor designs, the innovative ways to organize a retail chain, the technological and scientific discoveries. These producers of nonrival goods could then give away all of the underlying discoveries at marginal cost, zero.

A world populated by real people instead of these science fiction

4. See Merges and Nelson (1992) for an extended evaluation of these kinds of costs from strong property rights.

automatons faces two distinct problems in providing nonrival goods: how to share costs and how to select the most promising opportunities for investment. Real people will choose to be free riders if they can. They will not share the fixed costs of goods that are freely disseminated if they do not have to. In addition, assembling all the information necessary to decide which of the extremely large number of possible nonrival goods to produce is difficult. The kind of calculation showing that there are many wonderful things to discover also implies that there are an almost infinite number of ways to waste effort on interesting but socially useless nonrival goods. Think about software. About $10^{1,000,000}$ different bit strings can fit on a 360K floppy disk. For comparison, a year has about 10^7 seconds, and about 10^{18} seconds have passed since the big bang. Out of $10^{1,000,000}$ possibilities only a very small fraction need to be useful for there to be many useful software programs still to be discovered. But so many other possibilities also are available that all human ingenuity for the rest of time could be devoted to producing useless computer code.

The government's powers of coercion make it uniquely capable of solving the cost-sharing problem. Unfortunately, these powers also make the government uniquely capable of wasting large amounts of resources on socially useless purposes. (Recall the experience in the savings and loan industry.) Markets, conversely, can solve the sharing problem only by introducing monopoly distortions, but they are better than governments at selecting the opportunities to pursue and avoiding wasteful spending. Because people operating in the market are motivated by the potential for profit, they seek out only those nonrival goods that have real value. The parallel or simultaneous search by large numbers of market participants can efficiently evaluate many possibilities. Bankruptcy constraints quickly cut off the flow of resources to projects that turn out to be unpromising.

Under the existing institutional arrangements for producing nonrival goods, one or the other of these extreme mechanisms is typically selected as being most appropriate for a given type of good. In the public good portion (the bottom) of the nonrival column, the government pays for basic research and gives away the results. This arrangement is chosen partly because dissemination of these goods is so important. (Think, for example, of the polio vaccine.) In addition, because of the efforts of Vannevar Bush and people like him, the institution of peer review of competitive research grants is now available. It offers a reasonably good solution to the problem of selecting which projects to fund.

For nonrival goods at the top of the column, there is little prospect of setting up a government body that could make the right decisions about what to provide. No one would take seriously the suggestion that the government should extensively subsidize the production of popular music recordings, movies, or the design of new kinds of microprocessors so that these goods can be sold at marginal cost. Instead, society relies on market mechanisms to make those decisions and accepts the limits on dissemination and the monpoly distortions that the use of the market entails. The monopoly markups on compact disks, movies, and microprocessors cause relatively small welfare losses, and selecting which goods to produce poses an institutional design problem that dwarfs the problem faced in basic research.

The existing arrangement with government provision of basic research and market provision of final goods seems to work reasonably well for nonrival goods at the top and bottom of the column. It is the intermediate zone where the most important opportunities may now be missed. This region includes what Richard Nelson has called generic research. As he argues on the basis of case studies in different industries, this area may offer particularly large returns from investment in research.[5] It includes goods such as the principles of chemical engineering, the insights behind the design of computer interfaces, and the fundamentals of program design. It is this region that my proposal tries to address through its mixture of government and private sector mechanisms. Without trying to identify in advance what these areas are and what the specific opportunity for collective action is, the proposal seeks to create a mechanism that combines the government's efficiency at solving free-rider problems with the market's effectiveness in selecting practical problems that offer the highest rates of return. Market participants can then make the right decisions about where the returns on investment are highest for the industry.

More is at stake here than just the rate at which knowledge is transferred from basic research to commercial application. These areas play a special role as the intermediaries between the basic research com-

5. Nelson (1983).

munity and the final users of technological knowledge in an economy. Better communication here is important not just for commercial outcomes, but for the vitality of the basic research endeavor. Without some point of contact with the practical opportunities and challenges of the world, basic science risks drifting into irrelevancy. The recent trend toward closer contacts between individual firms and universities avoids the problem, but firms pay for research only if they get proprietary control over the results. These arrangements therefore tend to undermine the traditions of open dissemination of ideas that have made our universities so successful. The proposal outlined here could provide a mechanism for connecting industry with universities, without jeopardizing the traditional role of the university.

Economic Magnitudes

Economists have uncovered a great deal of evidence suggesting the economic importance of nonrival goods. In his very useful survey of the econometric work on spillovers from R&D, Zvi Griliches describes one of the first attempts to compute a social rate of return on investments in R&D.[6] T. W. Schultz computed the total resources saved by technological change in agriculture, compared the savings to total expenditure on R&D, and found a high ratio of benefit to cost.[7] In his refinement of this calculation, Griliches himself found that the social rate of return on public investment in research on hybrid corn had a rate of return of about 40 percent, a number illustrative of the magnitude of returns on investment in R&D that has been found in many subsequent investigations.[8] A sample of discoveries has been used to make this kind of calculation both in agriculture and in manufacturing.[9] It has also been made in agriculture, in manufacturing, and at the national level by means of regression analysis using data on total factor productivity.

Three conclusions emerge from this large body of work. The first is that the social rate of return to investment in the broad class of nonrival goods is quite high, on the order of 30–50 percent. This level of return

6. Griliches (1992).
7. Schultz (1953).
8. Griliches (1958, p. 425).
9. Mansfield and others (1977).

confirms the claim made above that these are in some sense the most important kinds of investments that can be made. Second, the research demonstrates that the social rates of return are significantly higher than the private rates of return. This gap can arise because firms that produce new nonrival goods have only weak control or property rights over them. Alternatively, it can arise because property rights are strong; monopoly pricing then induces its own wedge between private and social values. The third conclusion is that economists cannot estimate the *ex post* rate of return in any one industry or area of economic activity with anything like the precision required if econometric estimates alone were to be used to make decisions about where to direct research dollars. This literature offers no support for the idea that academics or bureaucrats will be able to read the numbers and pick winners.

The calculation by Griliches suggests that the difference between private and social benefits from research is important for more than the microeconomic details. The macroeconomic effects of this difference can be quite large. Using the calculated social rates of return to investments in R&D, he can explain the majority of the total factor productivity growth at the national level as the result of measured spending on R&D. And this kind of calculation must lead to an underestimate of the importance of nonrival goods because it cannot capture the kind of innovation that led, for example, to discount retailing in the United States. Wal-Mart no doubt does not show up as a big player in the R&D statistics but has nevertheless helped transform an extremely important sector of the economy, significantly lowering costs in the retail sector. Griliches' calculation is also consistent with the recent cross-country regression estimates by Lichtenberg and by Coe and Helpman suggesting that social returns to R&D, measured at the national and international level, are still very high.[10]

Principles of Political Action

Having made the case that there are important opportunities for collective action to encourage the production of nonrival goods, especially in areas of practical importance, I now evaluate the potential costs of

10. Lichtenberg (1992); Coe and Helpman (1993).

trying to undertake collective action. In doing so several general principles suggest themselves. Other observers would no doubt add to the list and might change the emphasis, but the generalizations listed below should not be particularly controversial.

Reaction

Every policymaker operates in an environment characterized by competition among many different independent policymakers. A government of divided powers, such as ours, has a number of policymakers, and every government in the world must take account of the actions of other governments.

In this setting any change in policy can induce a reaction from other policymakers. A good working hypothesis is that policymakers operate in a strategic environment characterized by tit-for-tat and in which the details of what others are doing are sometimes hard to discern. This suggests a sequence of tests in evaluating any policy initiative:

- *Self-interest.* Would a policy be worth adopting if no other policymaker changed policy in response?
- *Reflection.* Would a policy be worth adopting if other policymakers responded by adopting the same policy?
- *Robust reflection.* Would a policy be worth adopting if other policymakers responded by adopting similar but more pernicious policies?

These tests reflect increasingly sophisticated views of strategic behavior. The self-interest test applies when there are no other players or in a one-time strategic interaction. The reflection test applies in cases of repeated interaction—what economists call a repeated game. The robust reflection test applies in cases of repeated strategic interaction if the actions or motivations of other players are difficult to observe—that is, in a repeated game with asymmetric information. In both of the reflection tests, the implicit model of the policy equilibrium is based on trigger strategies in which all parties defect as soon as one does.

One example that bears directly on technology policy and that suggests the importance of the robust reflection test can be cited.[11] In the

11. This account is based on Crease (1991) and Office of Technology Assessment (1991).

early 1980s George Keyworth, then the science advisor to President Reagan, invented the concept of a "Presidential Initiative" and simply put a $140 million item in the Department of Energy budget to support funding for a National Center for Advanced Materials (NCAM), bypassing all of the usual mechanisms of peer review.

Suppose for the purposes of argument that this project represents a good investment in science and technology and thus passes the self-interest test. Suppose also that it passes the reflection test in a policy game between Congress and the executive branch. If both sides begin to use the special initiative process to fast-track high priority science and technology projects that are truly in the national interest, the net effect would be positive.

What Keyworth did not anticipate was that Congress would behave differently from the way he did because members of Congress face incentives that are different from his. They used the freedom from peer review to fund projects of dubious scientific value. At the same time that Congress rejected Keyworth's attempt to bypass the traditional procedures with his presidential initiative (which was eventually funded after undergoing peer review), it approved two congressional initiatives that also bypassed all peer or agency review and that had no apparent justification in terms of national interests. For example, one initiative gave Columbia University $24 million to renovate its chemistry building. (In a play on Keyworth's words, this was referred to as the National Center for Chemical Research.)

These actions marked the end of the traditional consensus that basic research funding had to pass peer review and would not be eligible for congressional earmarking. Universities began to hire their own lobbying firms and compete actively for earmarked funding. Senators and representatives now compete to provide resources for their constituents. Since 1982 earmarked research funds have grown from $9 million a year to $470 million in 1991 and $707 million in 1992.[12] For comparison, the annual budget for the entire National Science Foundation is about $2 billion.

Almost everyone concedes that much of the congressionally earmarked funding represents pure pork-barrel spending. The academic community, the executive branch, and the defenders of traditional sci-

12. Savage (1992, p. 8, table 1).

ence in Congress have tried to put the genie back in the bottle but to no avail. For example, when the Bush administration tried to rescind funding for specific examples of pork-barrel science, Congress threatened an extended battle of tit-for-tat by canceling funding for thirty-one specific research grants for the National Science Foundation and the National Institutes of Health. The Clinton administration initially said the right things but has subsequently backed down and gone along with the intentions of Congress.

One can always argue that Congress would eventually have developed a high level of earmarked pork-barrel funding on its own, but the specific sequence of events and the testimony of participants suggests that Keyworth's initiative played a crucial role in the shift from the old equilibrium to the new one. If so, it did not pass the robust reflection test.

Delegation

The dissatisfaction with congressional earmarking as a mechanism for allocating funds to universities reflects a general principle. Successful decisionmaking on matters of science and technology policy requires that responsibility for specific spending decisions be delegated to a body that is not under the direct control of members of Congress. For example, delegation is the key element in the success of the peer review system. It is the mechanism whereby members of Congress commit to each other that they will refrain from pursuing special advantage for their constituents as long as everyone else refrains as well. The lack of delegation is at the heart of the failure of many science and technology projects that must, because of their large dollar cost, be directly funded by Congress. The summary by Cohen and Noll of the results from six case studies of large commercialization projects (the supersonic transport, satellite technology, the space shuttle, the Clinch River breeder reactor, the synfuels project, and the photovoltaics commercialization program) is illustrative. "The overriding lesson from the case studies is that the goal of economic efficiency—to cure market failures in privately sponsored commercial innovation—is so severely constrained by political forces that an effective, coherent national commercial R&D program has never been put in place," they wrote.[13]

13. Cohen and Noll (1991, p. 378).

The problem with direct political control of large discretionary spending projects is one of incentives, not ethics. Members of Congress are rewarded in large part for providing services and benefits to constituents. This creates strong incentives to select projects for reasons that do not have much to do with economic efficiency or even institutional missions. For example, the only possible explanation for the fact that the National Aeronautics and Space Administration (NASA) has been authorized to build the Christopher Columbus Center of Marine Research and Exploration in Maryland, a project the agency did not request, is that the senator who chairs the panel that writes the NASA appropriation is from Maryland. Even if the ultimate goal of this or any other senator is to do the very best job possible of promoting national science and technology objectives, the first priority is to build a strong local power base to ensure reelection.

From an institutional point of view, the problem is further compounded, because what Congress can delegate, it can also take back. A consensus that everyone should refrain from seeking narrow advantage in some area of government activity can easily break down if the opportunity presents itself and the temptation becomes too large. Some proponents of direct government funding for specific kinds of commercial innovation point to the former Defense Advanced Research Projects Agency (DARPA) as an agency comparable to the National Science Foundation, able to make independent decisions free from congressional interference. This may have been true in the past, when its activities were clothed in secrecy and were not widely understood. But lawmakers from Michigan inserted $25 million into the fiscal 1993 defense budget to fund purchases by what is now called ARPA of flat panel display screens. (To emphasize the increased importance that this agency is to place on technologies that have both civilian and military applications, the new administration has taken "defense" out of the name.) It will come as no surprise that the legislation was written so that the company best positioned to fill the request happened to be located in Michigan.[14] Officials at ARPA reportedly disapproved, but they know where their funding comes from. They wrote a request for proposals that favored the Michigan firm and are now negotiating a final purchase contract.

14. Carey (1993).

Inertia and Competition

A third general principle is that large organizations are very difficult to change. They seem to adapt to changing circumstances only when confronted with serious competitive threats. This generalization seems to apply equally well to private and public sector organizations. Large companies such as Ford and Chrysler, which are now two of the low-cost, high-quality automobile manufacturers in the United States, began the major reorganizations that transformed their operations only after being faced with the serious prospect of bankruptcy. (Some economists point to the government's intervention to save Chrysler as a success. Perhaps, however, the government should have demonstrated its willingness to let a big-three automaker fail. If it had, Ford and especially GM might have begun to change much sooner.) Similarly, while the rest of the Soviet economy fell further and further behind Western levels of output and technology, the Soviet arms industry—propelled by the intense pressures generated by military competition—managed to manufacture fighter aircraft that were roughly on a par with those made in the United States.

This suggests a role for competition that is quite different from the one economists teach to students. Competition in this sense has nothing to do with price-taking and tangency conditions. Instead, it affects the power that the leaders of a large hierarchical organization have to overcome the incentive and monitoring problems that tend to plague these institutions under normal circumstances.

Reaction, delegation, and organizational inertia are general considerations that policy analysis in any country of the world should address. The next two issues are somewhat more specific to the current political and institutional context in the United States.

Pay as You Go

In the current budget climate in the United States, no large new spending program can be undertaken unless it includes a funding source. The magnitude of the change in spending priorities that an aggressive technology policy might contemplate makes the importance of this observation clear. The substantial difference between the social and private rates of return to investment in nonrival goods and their relatively small share of total investment suggests that it would be reasonable to consider an increase in annual spending on R&D of as much as 1 percent of gross domestic product, or roughly $60 billion a year. This is not an impossible amount to contemplate in a $6 trillion economy, but it is a very large amount to raise in new taxes in the midst of the current budget climate in Washington.

Divided Government

The American system of government was intentionally designed to limit the ability of any one part to take decisive, unilateral action. Subsequent developments, such as the recent erosion of the seniority system and party discipline in Congress and the expansion of the powers of the judiciary have probably pushed the system even further in this direction. The government is not an actor. It is a game with a very large number of players, many of whom can veto, or at least delay, any proposal. The equilibria that result are quite different from those that emerge, for example, under a parliamentary system.

To see the relevance of divided government for technology policy, consider two examples. First, ARPA (and before it, DARPA) has explicitly encouraged technological development of massively parallel supercomputers. It picked particular firms that it aided during the development of prototype machines. Firms that were not favored have objected to the perceived favoritism and have managed to instigate, no doubt through their local congressional representatives, a General Accounting Office investigation of ARPA that was critical of its supposed favoritism.

Set aside the question of whether ARPA is right that massively parallel computing is an important area of technology. (It almost surely is.) Also set aside the more problematic question of whether ARPA is able to pick the most promising firms in this area. The relevant point is that any agency directly answerable to Congress is increasingly unable to exercise any independent judgment about important technical issues. Soon, the selection of technology projects at ARPA will look like the rest of military procurement.

Military procurement is, of course, a familiar whipping boy. The Air Force, for example, has been trying to buy a large quantity of desktop computers for years now under its new "expedited" bidding process. The winner has been announced twice, only to be overruled when losing bidders protested to the General Services Administration board of contract appeals. As James Q. Wilson has emphasized, the inflexible, rule-driven, excessively bureaucratic nature of procurement is an inevitable outcome of a process in which a losing party has access to multiple bodies—members of Congress, other agencies, or the courts—to question the integrity and professionalism of the person who made the decision to give the bid to someone else.[15]

Evaluating the Self-Organizing Industry Boards

The economic case for creating the industry investment boards has already been made. The opportunity for collective action is a logical consequence of the existence of nonrival goods. (So, for that matter, is the potential for economic growth.) The private sector-government hybrid proposed here would use the tax system to share the cost of producing new nonrival goods, and it would use market forces to solve the selection problem of deciding how the funds could best be invested. Private, for-profit firms would make the decisions about the relative importance of various industry-specific public goods and whether these interests are sufficient to justify any funds at all. In addition, both the industry boards and the organizations, such as research universities, research labs, and training schools, that they might support would be forced to compete for funds. Thus, both private sector incentives and competitive pressures can be brought to bear on the selection process.

If it succeeds, this mechanism could support investments that span the gap between the most general forms of basic research and the product development activities of individual firms. During the past few decades new discoveries in this gap have been funded only if they bear on the health care or military missions of the federal government. Because other practical areas have been neglected and because support for health- and defense-related research is likely to contract in the coming years, it is particularly important to find a new way to select the practical problems that motivate good fundamental research and create economic value. A new source of funds to support work on these problems must also be found.

15. Wilson (1989).

Another, more subtle advantage of this kind of proposal is that it addresses broader issues than those typically considered in the technology policy debate. As Henry Ergas has emphasized, explicit government policies affecting procurement, subsidies for firms, diffusion, or the educational system take place within a broader economic and institutional context that crucially determines their effectiveness.[16] Implementation of this proposal can empower firms in various industries to change some of the basic features of this larger context. They can change, for example, the patterns of interaction between university researchers and those in industry. They can change the system of training at the postgraduate or vocational level. They can create new institutions for setting standards. In so doing, they can enhance the effectiveness of the policies that governments now take (too often with little success) to encourage the commercial application of new technology.

Some of the details of the proposal could be debated but are of secondary importance. For example, the primary effects of selecting a specific tax will come from investing the revenue in an area with large disparities between private and social returns. The distortions associated with different kinds of taxes will be second-order small. The decisive issue in selecting the tax will come not from the conventional analysis of elasticities and deadweight losses, but from the costs of collection and enforcement. Industry participants would have the right incentives to minimize these costs and should therefore be given wide latitude to propose their own enforcement mechanisms.

The political case for this particular proposal is somewhat more complicated. This system was explicitly designed to survive the robust reflection test. If every country decided to compete internationally by creating private sector initiatives that support industry-specific versions of Bell Labs, or open versions of SEMATECH, or new schools of biotechnical engineering, everyone would benefit from the discoveries that result. The federal government would certainly want to insist that other countries offer the same treatment to U.S. firms that they offer to domestic firms. If such a system were adopted in another country, the U.S. government would also insist that the spending must be transparent and that it not be used to offer continuing subsidies to domestic firms in an industry. If differential treatment or direct subsidies were possible,

16. Ergas (1987).

there is a real risk that even if the United States did not abuse this kind of arrangement, other countries could and would.

This proposal is also designed to ensure an adequate degree of delegation. The only effective way to do this is to make sure that the decisionmakers can operate without having to go through the congressional appropriations process each year. (This is why the Federal Reserve Board is independent in a way that no other quasi-governmental entity is.) Absent this freedom, the kind of interference that is beginning to occur at ARPA is sure to arise here as well. Finally, the provisions specifying free entry of new industry boards are designed to ensure that the inertia that typically saps the effectiveness of any complacent organization will not affect prospects for an entire industry—or for the nation as a whole.

In more parochial terms, the proposal is obviously designed to be self-financing and is tailored to keep any new financing arranged under this proposal isolated from the ongoing and protracted debates about federal budget priorities and deficit policy. As the example of the FDA fees suggests, it is not entirely unrealistic to believe that firms in some industries would be willing to back the required taxes if they could control how the proceeds are used. Finally, this kind of system, unlike any government agency, would be capable of decisive action because it would be free from interference from the legislative and executive branches of government.

One of the most serious (and certainly the most frequent) complaint about this proposal is that it may not work for all industries, perhaps not even for very many of them. Compared with a multibillion dollar program that puts control of innovation into the hands of some technology bureaucrats, this proposal is indeed less aggressive. But if one believes that the risks from careless intervention by the government are potentially very large, this built-in conservatism is an advantage, not a weakness. The imperative in policy design, as in medicine, should be to do no harm. In its current form, this proposal will be exploited only in industries whose participants perceive benefits that are large enough to justify the costs to them of the tax (which may in fact be small) and the costs of setting up and administering the system (which may be large). Where it is implemented, society will very likely derive important gains. Where it is not, the case for other, stronger measures must be viewed as open in principle, but with a very strong burden of proof on those who call for more explicit government intervention. Surely, the right way to proceed in formulating policy is to try the less risky, less intrusive measures first and save the measures favored by advocates of aggressive government intervention for later.

If this criticism is right, this mechanism will be implemented in few places. If so, there is less to be gained by creating this process, but little risk in trying. And the criticism may well not be correct. For many areas the opportunities seem to be large, and the mechanics of collecting the tax straightforward. For example, the telephone and cable TV industries could agree to a common, 1 percent revenue surcharge that would be used to support research and training on new digital communications and the interface between computing and imaging.[17] In effect, these industries could recreate the source of funds that once supported the rich interaction between practical problem-solving and fundamental research that took place at Bell Labs. Given the increasing likelihood that the "Baby Bells" will end up directly competing with each other (for example, through links with cable firms that operate in the service area of another phone company), the existing arrangements that support BellCore may erode. In any case, BellCore, the part of the original Bell Labs now controlled by the telephone operating companies, probably could benefit from being exposed to a more competitive research environment.

One could imagine an industry initiative by semiconductor manufacturers, with a tax of, say, 20 cents per million transistors on all domestic chip sales. If a firm in the semiconductor manufacturing industry believes that the SEMATECH consortium deserves support, it could use

17. After this paper was written, I learned of a research organization called CableLabs in Boulder, Colorado, that collects a voluntary charge of 2 cents per subscriber per month from the operators of cable television systems. Contributing firms cover 85 percent of the subscribers in the United States and about 70 percent of subscribers in Canada. CableLabs employs a small number of scientists and contracts out much of its research to universities, public and private research laboratories, and some equipment suppliers to the industry. It is currently working on problems such as digital compression and digital transmission of cable signals.

The fact that an organization very close to the one described here could come into existence even without the support of legislation that solves free-rider problems suggests the existence of a much larger unmet demand for the kind of collective industry-specific investment activity than skeptics have realized.

its funds to ensure support once ARPA withdraws its funding, as it is now scheduled to do. SEMATECH could compete with universities for funding from an industry-controlled research board, provided it operated according to the open disclosure and equal access rules required under the new system. If the research board did not believe that SEMATECH was a good investment, it could use its funds to support basic research at universities or the national labs. Other boards could support graduate training fellowships, worker training, or whatever other need they perceived to be most pressing.

One could also imagine initiatives in machine tools or in automobile assembly and design. More prosaic economic areas such as construction might also be covered, perhaps in subsegments such as single-family home construction, an area that seems to have been conspicuously lacking in technological progress.[18] A new initiative for electrical power utilities might compete with EPRI in areas such as electricity storage. Biotech firms could use their funds to create a school of biotechnical engineering that could do for them what chemical engineering at MIT did for the petroleum and chemical industries in the United States.[19] Software firms could tax shrink-wrapped software to help pay for basic research in computer science or to establish better curriculum standards, testing, and teaching at the junior college and undergraduate level. Software firms might also benefit from the creation of a software engineering discipline separate from the pure research activities of existing computer science programs. These schools could support the systematization of private sector knowledge about principles of software engineering and could train skilled professionals for software production rather than for university research, just as chemical engineering serves these functions for the petroleum and chemical industries. This kind of change in the institutional infrastructure could prove to be far more important for an industry than any particular government program of subsidies or attempts to pick winners.

In any industry complicated details of administration and tax collec-

18. See Nelson (1983) for a discussion of previous attempts to transfer the successes of the extension service in agriculture to the homebuilding industry. At least part of the problem in previous initiatives has been the attempt to implement the program from the top down, rather than meeting needs that were perceived by industry participants.

19. Rosenberg and Nelson (1993).

tion would have to be worked out. For example, if a tax of 20 cents per million transistors were levied on semiconductor chips, some decision would have to be made concerning the level at which tax obligations would be calculated and monitored. Typically, firms try to avoid responsibility for taxes, but under this proposal, control of how the funds are spent comes along with responsibility for collecting the tax. A tax per transistor could be levied when a semiconductor manufacturer such as Intel sells one of its chips. Intel would be responsible for reporting its total sales of transistors and would be able to decide which industry research board would get its taxes. When Toshiba exports memory chips to the United States, it could be responsible for reporting sales and could allocate its tax obligations among different boards. Alternatively, a firm such as Dell Computers that buys semiconductor chips from Toshiba would be responsible for reporting its purchases of transistors not already covered by a tax payment and could decide how these tax obligations would be allocated. One could even leave the choice of who pays the tax and controls the revenue up to the firms involved. Toshiba could sell some chips on which the tax has already been paid if it wanted to control some of the tax revenue. It could also sell some chips on which taxes have not been paid if Dell or another purchaser wanted to pay the tax and control the proceeds. All that would be required is that someone pay the tax before the goods move into the hands of the final purchaser.

A little thought would be needed to resolve the practical details, such as implementing a system for keeping track of the chips on which tax payments have already been made and those on which they have not. This problem is the kind that a government agency might find very difficult to solve or might solve by imposing large costs on the private firms that have to comply with its rules. But private firms routinely solve all kinds of difficult logistical, contractual, and monitoring problems. Dairy cooperatives make sure that fresh milk is always on the shelf in the grocery store. Railroads keep track of freight cars that are shuttled among different trains operated by different companies. Employers make sure that workers actually do their jobs in the absence of effective legal enforcement of employment contracts. Policy analysts should not underestimate the ingenuity of private firms in solving monitoring, enforcement, and tax collection problems in their industry if they perceive it to be in their interest to do so. And if firms do not perceive this program to be in their interest, it will not even be attempted.

Perhaps the biggest risk inherent in this proposal is that existing firms in an industry might use it to protect themselves from new competition. For this to take place, some degree of collusion between existing firms and the government would almost certainly be necessary. The requirements for openness and the oversight powers of the secretary of commerce would make it very difficult for existing firms to limit competition without at least tacit government approval. Unfortunately, cases in which the government would collude with the threatened firms in an industry are not hard to imagine. For example, the United Steelworkers have proposed a surcharge of $5 a ton on all imported and domestic steel to cover unfunded pension and health care obligations at large steel firms. This is a clear attempt to tax the new, innovative, nonunion minimill steel producers and subsidize the existing firms and their workers. The United Auto Workers are also considering this kind of arrangement for the automobile industry. Because the government, through its Pension Benefit Guarantee Corporation, is faced with large potential losses on the pension obligations of the largest steel and auto firms, it might be willing to go along with this kind of scheme.

Nothing prevents Congress from passing this kind of surcharge as a separate measure, but it could not be imposed surreptitiously under the structure of the industry investment boards. The enabling legislation would specifically prohibit any kind of spending activities by boards that would shift costs from or transfer resources among existing firms. The ultimate protection, however, comes from the freedom that firms in an industry have to create alternative boards and to control how funds are spent. Even if the established steel producers, because of their size, could impose an industry-wide tax that minimill operators do not want, the established producers have no way to coerce the minimill operators to fund an Industry Pension Board. The minimill operators could merely establish a separate board devoted to research on new technologies for minimills and refuse to contribute to any board dominated by the old, integrated producers.

The freedom to create new boards and to choose among competing boards is so important that any proposal that did not include these provisions would have an entirely different character. Without these protections, the political risks would loom very large compared with the potential economic benefits.

Alternative Technology Policy Proposals

The criteria listed above can be used to evaluate other technology policy proposals.

Increased Funding for Basic Research at Universities

Funding for basic university research has until recently received broad support from economists and politicians, partly because the economic advantages of free dissemination of basic scientific discoveries are so obvious. In addition, the delegation problem has largely been solved. Peer review delegates resources to the scientific community, which decides what issues to pursue within broad guidelines established by the funding agencies. (But as noted above, this delegation function is increasingly threatened by direct congressional earmarks that bypass peer review.)

More and more policymakers, however, are realizing that the existing arrangement solves only one part of the broader selection problem of deciding which areas deserve additional research support. Spending is concentrated on pure scientific research and on practical problems in the areas of health and defense, but too few resources are devoted to areas of basic research and training that are motivated by the commercial opportunities faced by private firms. Because policymakers are beginning to emphasize the commercial relevance of research and because budgetary and geopolitical changes may reduce mission-oriented basic research in health and defense, the amount of support universities receive for basic research may soon decline substantially. From a practical point of view, the relevant question is not how much government support for basic science at universities can be increased, but whether the current level can be preserved.

To understand how a system of industry investment boards could protect the existing level of support for pure research at universities and at the same time address the reasonable concerns of policymakers for a commercial payoff from some kinds of research, it is useful to look at the history of universities in the United States. As Nathan Rosenberg and Richard Nelson demonstrate, the university system in the United States has always paid a great deal of attention to practical problems and has contributed enormously to their resolution.[20] But the nature of those practical problems has changed dramatically over the years.

Before World War II universities primarily addressed practical problems faced by commercial interests. The land grant universities were devoted to agriculture and "the mechanic arts." Farming, railroad transportation, electrical generation and transmission, and chemical processing all required their own specific kinds of technical research and training. Europeans sneered at the vocational orientation of higher education in the United States during the nineteenth century and the first part of the twentieth century, but Americans derived important national economic benefits from a responsive university system. To illustrate that responsiveness, Rosenberg and Nelson describe the emergence of programs of electrical engineering. In 1882—the same year that Edison opened the first electrical station in New York—MIT introduced its first electrical engineering courses. Cornell followed in 1883. By the 1890s schools of electrical engineering were the primary suppliers of skilled professionals for the electrical industry, having supplanted unsuccessful attempts by firms to train engineers in house.

This experience was not unique. Universities trained people for work in many areas of commercial importance and developed bodies of knowledge that could be used there. They did not do this, one can be sure, entirely out of patriotism or dedication to national economic success. Universities were attentive to private sector opportunities because the private sector paid the bills.

During and immediately after World War II, the federal government became by far the largest patron of research and advanced training at universities. This shift in funding accompanied a shift toward solving practical problems in health and defense and pursuing more abstract problems. The Department of Defense and departments with related

20. Rosenberg and Nelson (1993).

missions (Energy and NASA) provide the majority of all research support in the engineering disciplines. The total expenditure at the National Institutes of Health is five or six times larger than the budget at the National Science Foundation.

There is a continuing debate about whether this postwar shift in emphasis has been beneficial for the business sector. Many observers say that the commercial spin-offs from military research have decreased in recent years. The typical interpretation of this finding is that civilian and military technologies are diverging for exogenous technological reasons. An alternative explanation is that the initial surge in funding from the government in the immediate postwar years stimulated a system that for historical reasons was still attentive to commercial applications and opportunities. As the system evolved in the 1960s and 1970s, however, its connections to the civilian business sector atrophied. The increasing divergence may therefore reflect the delayed effects of incentives and funding rather than exogenous technological developments.

However this debate is resolved, the end of the Cold War clearly means that defense-related support for research will be substantially reduced in the next decade. Moreover, the enthusiastic support for technological advance in medicine and health may increasingly come under attack as the nation tries to come to terms with the rapidly escalating cost of medical care. Some policy analysts are already suggesting that because new health technologies are not being priced and allocated correctly, slowing the rate of technological advance would be the best way to control health care expenditures.

Stagnation, or even real cuts in government funding for university research, is thus a distinct possibility. Moreover, demands will grow for more direct commercial payoffs from research. This pressure is reflected, for example, in the controversial planning process that Director Bernadette Healy tried to impose on biological research funded by the National Institutes of Health in 1992. Responding to congressional pressure, she tried to adjust the decision process and priorities of the traditional peer-reviewed, basic research program to give greater emphasis to commercial objectives. In the current institutional environment, this pressure could undermine the strengths of the existing basic research system without generating the hoped-for commercial benefits.

The National Institutes of Health, the National Science Foundation, and the Department of Defense are unlikely to be the best institutions for assessing commercial opportunities.

A more productive response to politicians and voters who ask for research and training that have closer economic payoffs would be to agree to a division of labor. Basic science agencies such as the National Institutes of Health and the National Science Foundation could continue to fund the most adventurous and forward-looking research at their existing levels, and an independent source of funds could be created for the commercially relevant research that should be under the control of people in the private sector who are knowledgeable about the opportunities. These funds could fill in part of the gap that would open if mission-oriented funding in defense and health is cut back. Such a dual system could both restore the strengths of the pre–World War II connections between universities and the private sector and preserve the advantages of the more abstract basic research system that developed after the war. It might also provide more total funding than is currently available. The system of industry investment boards would establish this more balanced arrangement.

It is important to recognize (as Rosenberg and Nelson do) that an increased emphasis on practical problems is completely consistent with a division of labor in which universities concentrate on basic research, where free dissemination of knowledge is most important, and firms concentrate on R&D activities over which property rights should be strong. It would be very unwise for university researchers to perform proprietary research for private firms, yet many collaborations between business firms and universities or teaching hospitals are now taking precisely that direction. Universities in search of additional funding are increasingly seeking out arrangements under which they give up some of the traditions of open science and in effect become research subcontractors employed by private firms. Because the industry investment boards can solve, or at least mitigate, the free-rider problem, they can support universities and help set the research agenda without endangering the free exchange of ideas.

Closer interaction between firms confronted with practical problems and researchers pursuing fundamental questions may lead not just to bigger economic benefits but also to better basic science. For example, solid state physics was enormously stimulated by an attempt at Bell

Labs to solve a reliability problem with vacuum tube amplifiers. The labs had scientists on staff who knew some of the basic properties of semiconductors from solid state physics. They used this knowledge as a basis for their experiments leading up to discovery of the transistor. This discovery fed back into the basic research and training activities of solid state physicists and completely transformed the field. High temperature superconductors and opto-electronic devices are just two of the many underexploited opportunities that pose as many challenges for basic scientists as they do for practical problem solvers. These areas are at least as promising for the pursuit of basic knowledge as is the search in high energy physics for "the God particle," using ever more expensive particle accelerators.

In summary, issuing the perennial call for a massive increase in federal support for existing basic research programs at universities is a totally unrealistic approach to a technology strategy. It is politically unrealistic and would not by itself generate much in the way of economic benefits. It might not even produce better basic science. The real challenge will be to keep government support for basic science from falling. At the same time, universities will feel growing pressure to contribute more directly to economic performance. Unless these pressures are channeled in a constructive way, they could undermine much of the strength of the basic research system, without achieving the desired practical benefits.

R&D Tax Credit

On efficiency grounds the optimal R&D tax credit would be so generous in its definition of R&D activity that no corporation would pay any income tax. Because of the widely acknowledged distortions that arise from the double taxation of capital income (once as corporate income, then as individual income), the efficiency gains from abolishing the corporate income tax would be large.

The United States has a corporate income tax for political reasons, not on grounds of economic efficiency. The potential for substantial increases in the generosity of the R&D tax credit is therefore limited by the lack of substitute funding sources and sensitivity to the political risks of giving "tax preferences to corporations instead of people." If politicians continue to insist on raising revenue from the corporate income tax, an argument can be made, in principle, for taxing R&D capital less heavily than physical capital because R&D has a higher social rate of return. In practice, the crucial issue is whether the tax authorities are able to distinguish R&D spending from other kinds of spending. Corporations will surely get better at redefining what they call R&D in response to bigger tax incentives for doing so. Over time, the revenue cost of each additional dollar of new R&D generated through this mechanism will grow. As a result, serious practical limits will be placed on the size of the subsidy that can be provided in this way as long as substantial revenue is collected through the corporate income tax.

Compared with other, more specific technology policies, the R&D tax credit is relatively safe politically, which no doubt explains why it has attracted broader political support than direct subsidies to firms. In effect, the tax credit specifies an aggregate level of subsidies for R&D, and then it solves the delegation problem by letting individual firms select their own R&D projects.

Whatever the decision on the constrained, second-best level of the corporate tax rate and the R&D tax credit, it should be clear that R&D spending by private firms (like government-funded pure basic scientific research) is a complement to the kind of industry-specific public goods that the industry investment boards would proivide. Taken together, these three mechanisms (private R&D in firms, government support for basic research, and industry support for industry-specific public goods) would cover most of the spectrum of new ideas illustrated in the nonrival column in table 1.

Infrastructure

Increased funding for infrastructure as a technology policy has at least two serious strikes against it. First, no economic case can be made for infrastructure as a good that requires collective action. Very little infrastructure is a nonrival good. At best infrastructure is what is called a club good in public finance, and these goods can be efficiently provided by decentralized action in markets. Except for the bargaining problem over rights-of-way that is readily solved by the powers of eminent domain, no collective action is needed to provide infrastructure. (The state could use its powers of eminent domain to assist in land acquisition where necessary but still rely on private sector decision processes in all other areas.) If anything, governments should be moving away from government-subsidized infrastructure and toward greater *reliance on market-like mechanisms such as congestion charges on* roads and competition among providers of basic telephone service. Localities or even private firms should be allowed to select and fund infrastructure projects.

The second problem is that federal funding for infrastructure is rife with the kind of waste inherent in a system that lacks any degree of delegation in project selection and funding. The government continues to provide support for ''mass transit'' that defies all economic logic. (My favorite example is the subsidy for the buses in Vail, Colorado, that take skiers through town and between ski lifts.)

The dollars at stake in decisions on popular infrastructure projects are truly frightening. A plausible estimate for the cost of roadbed construction for magnetic levitation trains is $60 million a mile. Just putting in the track for a train that runs from Boston to Chicago could cost $60 billion, or about thirty years worth of funding for the National Science Foundation.

Dual-Use Technology

Many advocates of a more aggressive technology policy have pointed to the apparent successes at DARPA as a model for a more activist role for the government. This approach should set off all kinds of warning bells about serious political risks. The most important problems pertain to funding, delegation, divided government, selection, inertia, and the robust reflection test.

The funding problem is self-explanatory. Budget constraints mean that the federal government is highly unlikely to make available enough funds to undertake a technology policy initiative that could make much difference on a national scale. The delegation problem has already been noted in the discussion of the contract for research on flat panel display screens. ARPA will be under increasing pressure to conform to congressional demands, especially if its budget grows. ARPA projects would then offer the best opportunity for every member of Congress to provide an identifiable local project. ARPA projects would soon look like infrastructure, earmarked pork-barrel science projects, and military procurement. Asymptotically, every technology project would have part of its spending located in every congressional district, and the selection process for projects would become increasingly rule-bound and bureaucratic.

In its proposal on technology policy, the National Academy of Sciences Panel on the Government Role in Civilian Technology proposed that an independent civilian technology corporation be created with a one-time grant from the government of $5 billion.[21] This proposal reflects the panel's appreciation of the importance of the delegation and divided government problems, but it cannot be regarded as a true solution. A private corporation would in principle be free to make independent decisions and would not be subject to vetoes and second-guessing by multiple actors in the government. But the proposed $5 billion initial grant is far too small to make a permanent contribution to a national technology policy. (This relatively low figure was presumably selected in the hope that Congress would fund it. Even at this low level, funding now seems extremely unlikely.) The interest income on $5 billion would be about $250 million a year. Compare this with the $5 billion or more that IBM and GM each spend on R&D annually. Surely, the proponents of the Civilian Technology Corporation are assuming that if the corporation succeeds, it could go back for more funds. But that means the corporation would have to be highly responsive to the wishes of members of Congress from the start.

The comparison with IBM and GM also raises a caution about the selection problem and institutional inertia. Some proponents of the ARPA, or government planning, approach argue that industry leaders are not competent to make the right decisions about technology in the way a few key people at DARPA are alleged to have done. The countervailing argument is that putting all of the public's technology resources into a single hierarchical, bureaucratic structure leaves no protection if that organization stops producing valuable results. None of the existing proposals put any competitive pressure on ARPA or an independent technology corporation. If the entire auto and computer industries had depended on GM and IBM for all of their innovation, cars and computers would now be much less useful than they are. That

21. Panel on the Government Role in Civilian Technology (1992, p. 94).

is true even though the research staffs in both corporations are highly talented and even though some discoveries of real value have emerged from both institutions (for example, catalytic converters from GM and high temperature superconductors at IBM.) There is no substitute for competitive pressure on any such organization.

Even if all of the other objections could be met, the technology agency approach is probably not the kind of policy that passes the reflection or robust reflection tests. It is all too easy to imagine that an ARPA or a well-funded civilian technology program could be used to subsidize something like the European Airbus program, which Americans now find so objectionable. Mechanisms for direct government subsidies to firms may not be something that the United States wants to encourage in other countries.

Research Joint Ventures

Interest in the United States in research joint ventures seems to have been motivated primarily by the success of the VLSI (very large scale integration) joint venture among semiconductor firms in Japan, an example that is now understood to have been unusually successful and not representative of the typical outcome. Research joint ventures suffer from the obvious economic problem that they cannot solve the free-rider problem and achieve free dissemination of the results at the same time. What has been attempted in the United States, for example at SEMATECH, is a mixture of some degree of insider advantage that is promised to participants in the joint venture together with federal matching funding to sweeten the pot. One half of SEMATECH funding was provided by DARPA, so if this model were to be adopted more widely, all of the concerns about delegation and divided government that apply to DARPA would extend to the government-assisted joint venture.

But if government funds are not used, in any instance where several firms in an industry find it in their interest to work together on a research project, they presumably will be motivated to do so by the availability of important special advantages that accrue only to members of the joint venture. Such cases should raise serious concern that collusion in research will lead to product market collusion. A research joint venture could, for example, become the vehicle for preventing new entry into an industry or for reducing total industry expenditures on research.[22] Compared with the proposal outlined above, in which all joint effort has to be directed at activities that are freely available to any current or future participant in the industry, the research joint venture seems to hold out much greater risk of reduced total research and product market competition than do industry investment boards.

Managed Trade: Import Restrictions

Managed trade comes in two flavors, import restrictions and market access requirements. Import restrictions have the appeal that they are, in effect, self-financing. Import quotas, for example, can generate additional revenue for protected firms without imposing an explicit tax. Import restrictions suffer from almost every other conceivable problem. Some firms will support this option out of naked self-interest, but virtually no economist is willing to try to make a serious intellectual case for these measures.

Managed Trade: Market Access Requirements

Proposals that the U.S. government demand adequate access to foreign markets are more problematic. The purely economic case here is stronger than many economists who support free trade would like to admit. Market access is a necessary condition for effective protection of intellectual property rights. Suppose, for example, that a foreign country enforces a ten-year quarantine on all imported foreign software to check for software viruses. (Sounds plausible, right?) Domestically produced versions of word-processing software, spreadsheet programs, and other kinds of applications that are close copies of the versions available on world markets would no doubt be written and sold in this country. The effect for U.S. software firms would be as if the foreign firms simply sold bootleg versions of their programs. Formal copyright protection is worth nothing if a firm is deprived of the right to sell its goods.

It is not clear that market access restrictions would be in the self-interest of the foreign country. It depends on how costly the copying

22. For a discussion of these issues, see Katz and Ordover (1990).

process is and on what effects the protection has on domestic incentives. The experiences in Brazil, with its failed computer industry, and in Japan, with its (allegedly) protected personal computer industry, do not speak well for the long-run success of this strategy. If a protected domestic market were truly the key to export success, NEC and the Brazilians should have taken over the market for personal computers in the United States, but they have not. For internal political reasons, many countries may nevertheless restrict market access. The United States must therefore decide how it should respond to these restrictions.

Some advocate unilateral punitive measures, but this approach fails the robust reflection test. If market access requirements were given the same status as intellectual property rights and were monitored and enforced through a multilateral body, they would be beneficial. But a unilateral decision by the United States to adopt punitive measures as it sees fit could ultimately have very serious negative affects on the institutions that now support free trade. This would effectively condone the principle that countries can take unilateral action when they are unhappy with trade outcomes. Many other countries would soon start instituting punitive measures that are supposedly intended to punish market restrictions elsewhere but that are, in fact, pure income transfers to powerful domestic groups.

Moving ahead with unilateral punishments would be like announcing an intent to drive through red lights whenever no cars are coming on the cross street. We may believe that no harm will come from our actions if we use our new powers judiciously, and we may even be willing to let others do what we do if they promise to be as responsible as we are. The ultimate outcome is nevertheless likely to be one that we regret.

A useful comparison would be to contemplate the consequences of repealing the interstate commerce clause in the U.S. Constitution and letting individual states take punitive action whenever another state restricts market access. (Many state governments in the United States currently do try to give the same preference to local firms that the Japanese government apparently gives in procurement.)

Economists would almost universally agree that giving each state the power to interfere with interstate commerce would not be the way to deal with market access problems within the United States. The same argument presumably suggests that the United States should work to strengthen the General Agreement on Tariffs and Trade and the available multilateral institutions in the hope of moving trade among countries closer to the substantial, but not quite perfect, freedom that now prevails among the states of the United States.

Nevertheless, attacks on restricted market access will continue to be a very attractive political strategy for the federal government. Such attacks can be used to aid a small number of firms with high visibility. They do not require budget expenditures, they typically raise little domestic political opposition, and a confrontation with bad guys from another country always plays well in politics. (The war over the Falkland Islands may indeed have been like two bald men fighting over a comb, but it did wonders for the popularity of the Thatcher government.) Of all the policies described here, saber rattling and punitive measures designed to open foreign markets are the most likely candidates for a technology strategy in coming years.

Conclusions

In one of the last things he wrote, the late George Stigler concluded an essay on monopoly with these words: "The merits of laissez-faire rest less upon its famous theoretical foundations than upon its advantages over the actual performance of rival forms of economic organization."[23] Any discussion of technology policy should take these words as both a warning and a challenge. The warning is that even after admitting all of the deficiencies of economic markets, one must acknowledge what experience has so clearly demonstrated—that most of the familiar political alternatives are far less efficient mechanisms for allocating resources than the market is. The challenge for economists is to understand why markets perform well and then to build upon their strengths. The problem with the classical description of laissez-faire is its suggestion that the best of all possible arrangements for economic affairs has already been discovered and that it requires no collective action. The lesson from economic growth is that collective action is very important and that everything, including institutions, can always be improved.

23. Stigler (1993, p. 402).

The reality of economic life is that the pure individualistic market exchange invoked in the usual defense of laissez-faire has been supplanted by many institutions that allow collective action and improve on pure individualistic market exchange. Perhaps the most important such institution is the limited liability corporation, an invention that would not have been viable without the emergence of supporting regulatory and legal institutions. Other examples of institutional innovation include private universities, copyrights and patents, corporate research laboratories, and peer-reviewed competitive grants for research.

Each of these institutions represents an attempt to take advantage of the opportunities for mutually beneficial coordinated action that are presented by the world in which we live. Each is consistent with the operation of markets and supports success in the market as the final test of economic activity. Each builds on or emulates at least some of the strengths of the market: the push to do better that comes from competing against rivals, the pull that comes from opportunities for individual gain, the diversity that comes from having many different individuals and organizations working in parallel to achieve a given end, and the discipline that comes from clearly enforced criteria for success and failure. From this broader perspective, the "theoretical foundations" to which Stigler attached relatively little importance—the notions of price taking and tangency conditions—are only one small part of the picture.

This paper has tried to place the discussion of technology policy in this larger context. It has tried to shift the discussion away from narrow questions about whether the market or the government is better. It has avoided conjectures about which sectors should be given government support or about what are the most important "critical technologies." Instead, it has suggested that the debate focus on the underlying processes that lead to effective institutional arrangements.

To a large extent, market competition has become a process for selecting ever better corporate institutions that can channel the energies of large numbers of people toward the production of ordinary private goods. This paper describes a parallel process for selecting ever better institutions that can channel the energies of large numbers of people toward the production of the new discoveries that drive economic growth.

The process outlined above combines government-like aspects (mandatory taxes that solve free-rider problems) with market-like mechanisms (free entry, competition among different institutions, and decisions that are ultimately grounded in opportunities for profit). It is designed to fill the gap between private and public support for R&D. The essence of the proposal is to empower the firms in different industries, giving them the tools needed to solve the collective action problems inherent in providing industry-specific public goods. The proposal protects product market competition among firms and encourages competition among the organizations that would provide industry-specific public goods such as research or training. The argument is premised on the idea that long-run growth depends on our ability to discover innovative ways to arrange the limited stocks of objects. It concludes that the prospects for growth could be enhanced by searching for equally innovative ways to arrange our institutions.

Comments and Discussion

Comment by Zvi Griliches: This is an exciting paper. It is nice to see somebody young tackling big policy issues. The paper builds on an analysis of the past, but it is basically about the future. It is suggesting new mechanisms and new institutions to solve pressing economic problems. It takes as its historical example the "invention" of land grant colleges and the development of Bell Labs. It sees unlimited opportunities for investment in both basic and applied research and is looking for mechanisms that could provide the financing that would launch the next expeditions into the wide unknown.

The two basic assumptions of the paper are, first, that the research opportunities are out there and, second, that without new financing institutions, we may not get there. I believe that Romer is right on both counts but that the evidence for the first is not as overwhelming as he assumes it to be and that the particular institutional innovation may not work as well as advertised—but nothing ventured, nothing gained.

I will focus on the second point first. I see two difficulties with Romer's proposal. The first deals with the incidence of benefits from and the costs of technical change in an industry. Implicitly, a high elasticity of demand is assumed for the products of the various industries. Because the benefits from this research will be public, they are likely to be competed away and not redound to the benefit of the original investors in these boards. Moreover, if demand were inelastic, as in agriculture, such technical change could reduce the rents to the scarce resources in these industries and might be counterproductive from the private point of view. Thus, industries might not vote for such a levy if they see through it.

Second, if there are several boards and I am a small producer, I may choose to contribute my share to a peculiar board, specializing, for example, in supporting economic research at the University of Chicago, on the assumption that the big guys will give to the important research purposes anyway and that my contribution to those boards is unlikely to affect the industry's fate by much. In other words, allowing for alternative boards and for freedom of choice among them may still leave free-rider problems, even within this framework. Nevertheless, the proposal may be worth trying. People may not be as selfish as economists say they should be.

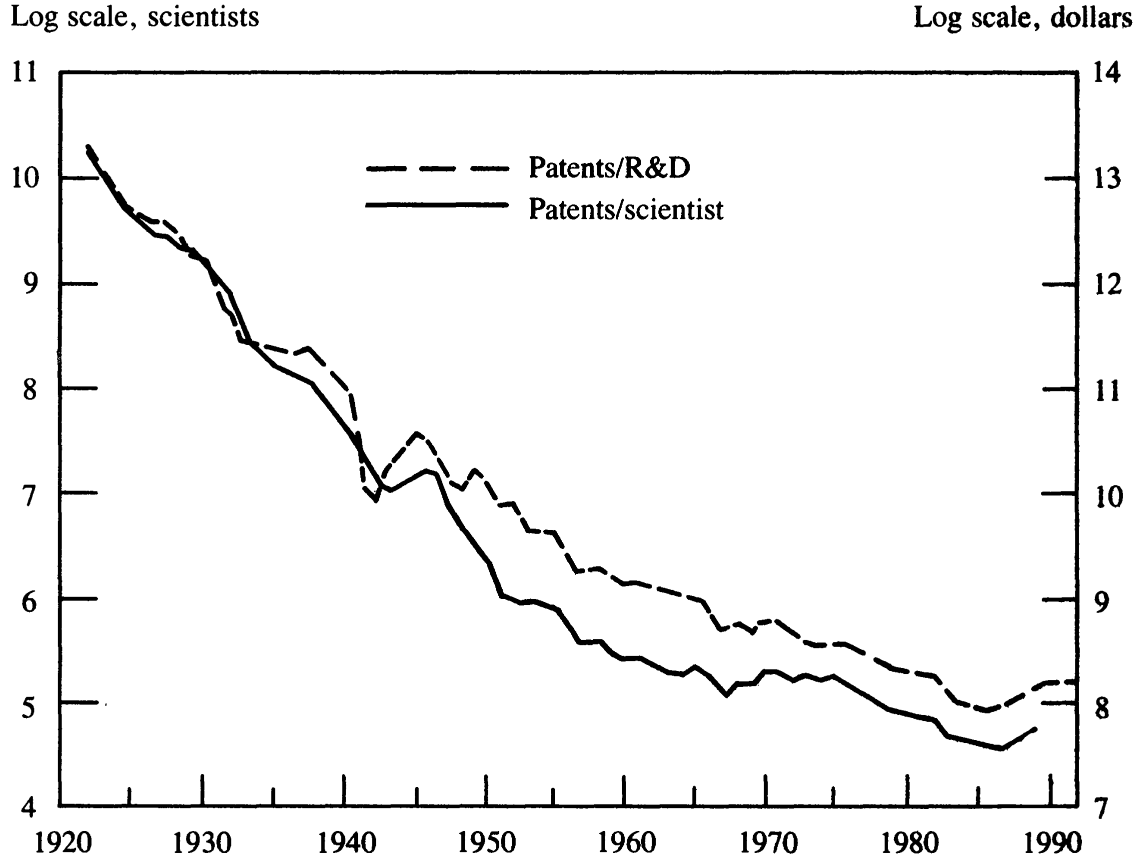
The more difficult notion is the one of the "endless frontier," the absence of diminishing returns to research. Because I have been worrying about this lately, I will focus the rest of my remarks on what we do not know about it.

One can ask two related questions about this. Are our recent "slow-growth" problems caused by a shortfall of research and development? And will encouraging more R&D be helpful here? I don't think that the sharp recession of the mid-1970s and the subsequent slow growth were the result either of the decline in the rate of aggregate R&D investment that occurred in the late 1960s and early 1970s or of a fall in the "potency" of these expenditures. The timing and magnitudes are all wrong.[1] It is probably a good thing to pursue more R&D in any case, for reasons of long-term growth, but not as a solution to our current problems. (Of course, Romer is not saying that his proposal will solve these problems, but it is the current economic situation that gives his proposal its resonance.) We are not currently constrained by our production possibilities frontier. Much existing knowledge is not being used yet. Much more diffusion of computers and learning about them is still in store. And the economy could be made much more efficient by eliminating a variety of barriers to its functioning. Nevertheless, R&D is one of the relevant margins, and we should be exploring it.

But before we encourage more R&D by creating new institutions, we have to face the crude evidence of diminishing returns to it. Figure 1 plots the ratio of patents to company-financed R&D for the last seventy years. If patents are taken as an indicator of inventive output, then the news is not all that good. Bronwyn Hall's paper in this volume

1. See Griliches (1988) for an elaboration of this argument.

Figure 1. Domestic Patent Applications per Company-Financed R&D in Industry (1972 dollars) and per Scientist and Engineer, 1921–92.



Source: Bureau of the Census, *Historical Statistics of the United States, Colonial Times to 1970*, pt. 2 (Department of Commerce, 1975), chap. W; Patent and Trademark Office, *Technology Assessment and Forecasts* (Department of Commerce, March 1977), app. A; National Science Board, *Science and Engineering Indicators*, 1991, NSB 91-1 (Washington); and Patent and Trademark Office releases. Number of applications by residence of inventor for 1940–59, unpublished memorandum of P. F. Fredrico, Patent and Trademark Office, January 18, 1961; for 1960, *Journal of the Patent Office Society*, 44 (February 1964), p. 168; for 1961–62, Commissioner of Patents and Trademarks, *Annual Report, 1966*, p. 26; for 1963–91, Patent and Trademark Office releases; extrapolated back, before 1940, by the number of total applications. Data on company R&D and the number of scientists and engineers in industry from the same sources, updated from NSF Reports 92-307, 92-330, and 92-331. Deflated by R&D price index from Griliches, "Comment on Mansfield," in *R&D, Patents, and Productivity*, (University of Chicago Press, 1984) p. 149, and Bronwyn H. Hall and others, "The R&D Master File Documentation," Technical Paper 72 (NBER, December 1988), updated.

Note: Left scale is the log of the number of domestic patent applications per 1,000 R&D scientists and engineers. Right scale is the log of the number of domestic patent applications per $1 million of company-financed R&D (in 1972 dollars).

shows a significant decline in the market's recent valuation of R&D. Evidence claiming to show exhaustion of inventive opportunity was presented by Baily and Chakrabarti at these same meetings some years ago, and Evenson looks at the international evidence on patenting and comes to similar conclusions.[2]

I myself do not think these facts need to be interpreted so pessimistically, but before we urge significant additional investments in R&D,

2. Baily and Chakrabarti (1988); Evenson (1993).

we need to convince ourselves that this would be a good investment. My own optimistic reading of these same data is based on scattered evidence that the economic "meaning" of patents has changed. This is the only reasonable way to interpret the fact that ratios of patents to R&D have been falling consistently, until quite recently, through good times and bad. If we did not worry about it in the 1950s and 1960s, when total factor productivity growth was rapid, why should we worry now? Moreover, even while—according to Hall's estimates—the market's valuation of new R&D was falling sharply, firms were expanding their R&D investments at rates not seen since the 1950s! Did companies know something that the market did not, or were they all lemmings? At the moment we face two contradictory facts: domestic patent applications started growing sharply in the late 1980s, indicating some revival of inventive activity, while the growth in companies' real R&D expenditures slowed to a crawl and may have actually turned negative. But overall, if the numbers are to be believed, we are not investing less in R&D today, relative to the size of the economy, than we did in the peak years of the 1960s. It is a puzzlement.

An argument can be made that precisely because we may have been facing the exhaustion of inventive opportunities, we need more R&D, but R&D of a special kind, the basic kind, the kind that would "recharge" the pool of knowledge and increase the effectiveness of the rest of it, which is spent on "D," rather than "R." We need additional investment in science and basic research to make the run-of-the-mill company's R&D more productive. In this sense, Romer's new institutions, specializing in basic and "generic" industrial R&D may be just what the doctor ordered. But what we want is basic research and basic results that will not be appropriated by individual firms, and I am not sure that we will succeed in convincing such firms to tax themselves for the public good.

I do want to dissent from Romer's view that the fact the molecules can be arranged in an infinity of ways implies an "endless" frontier for R&D exploration. It may not be feasible to discover the work of a new Shakespeare in the random typing of monkeys, even if it is there. More seriously, the notion that there may be no diminishing returns to research in the long run, and the associated notion of a permanent exponential growth of real income per capita, is a dream full of hubris, the notion that man is God-like and not subject to serious constraints or hard choices. Many years ago, in commenting on the "old" growth theory, Arrow remarked that "eternal exponential technological growth is just as unreasonable as eternal exponential population growth."[3]

The fact that the frontier may not be endless need not be taken pessimistically. First, we are still very far from being there. And, second, much is left to do to improve the situation as it is, including the diffusion of already-known technologies and the dismantling of the many manmade barriers to economic efficiency. Technological change is still endogenous, even if it is not endless, and although the outcomes of science and research are largely out of our hands, better economic understanding still has much to contribute. Schmookler, in one of the last things he wrote, complained about the state of this field: " . . . it may leave scientific and technological progress unexplained, which is unsatisfactory; alternatively, it explains scientific and technological progress as self-generated, which is dreadfully wrong. Such a view is wrong both scientifically and morally. It is wrong scientifically because . . . it is simply not true. It is wrong morally . . . because it deprives man generally of any sense of responsibility for the course of social and economic development. For if [technological change is endogenous], then all men must accept some measure of responsibility for what happens next."[4] It is admirable that Romer is willing to do just that.

General Discussion: In evaluating the potential of the author's plan, several participants raised examples of past experience with cooperative research efforts. Pointing out that few firms have formed research joint ventures in the decade since it first became permissible under antitrust laws, Timothy Bresnahan wondered if there really is an unfulfilled demand for more cooperative research. He also noted that existing cooperative institutions in the form of standards committees (such as ANSI) are abandoned by many in information technology industries when these committees attempt to coordinate the direction of technological progress through anticipatory standard setting. Bresnahan considered the rules imposed by standards committees to be quite modest and argued that the author's seemingly more restrictive institutional arrangements could prove even less popular.

3. Arrow (1969, p. 34).
4. Schmookler (1972, p. 84).

In contrast, Albert Link said that experience with SEMATECH, the cooperative research venture of the semiconductor industry, shows that cooperation can yield important benefits. Despite others' criticisms of that organization, Link said, it has ensured that American companies are competitive in their production of semiconductor manufacturing equipment, such as wafer polishing tools.

Link, however, questioned whether the author's proposal would succeed in increasing basic scientific research. He maintained that existing cooperative arrangements with universities are becoming more commercially oriented and now almost mirror research that is performed at private firms. Link nonetheless praised the proposal's concern with providing a forum for the more efficient evolution of standards, saying that this could lead to a more rapid diffusion of technology throughout a given industry.

Linda Cohen argued that the coercive part of the author's proposal might make it more viable than existing forms of noncoercive cooperative arrangements. She maintained that noncoercive cooperative agreements work well only in regulated and monopolistic industries (and, for unique reasons, in agriculture), and that they tend to break down under competitive conditions. She noted that when the electric industry became more competitive, Southern California Edison, one of the largest and most innovative utilities, dropped out of the Electric Power Research Institute because the utility no longer wanted to share technology. Using SEMATECH as an example, she argued that when noncoercive cooperation involves the public sector, the results of the research are forced to be more public, allowing many firms to opt out of the arrangement and get a free ride. In competitive markets, she concluded, coercion forcing 100 percent of participation may be the only feasible way to bring about sustained cooperative research.

Michael Katz argued that the government role in the author's proposed arrangement would end up being far more intrusive than the author assumes. He said that government would become integral in determining which firms fall within an industry and which do not. He also suggested that the author's proposal to make imported goods subject to the mandatory industry taxes could lead to abuse if other countries adopted this proposal as well. If his plan were effected in a country such as France, Katz said, nationalized French firms might be able to avoid paying their taxes through offsetting government subsidies, placing American and other foreign firms subject to the tax under an unfair burden.

Several participants discussed the author's presumption that almost limitless potential gains can be had from pushing out the knowledge frontier. Robert Gordon argued that diminishing returns present a serious constraint to the benefits of technical progress. He suggested that such diminishing returns can be observed by looking at the effects of the technological changes of the past century. Such changes have been transforming daily life over this period, but the rate of transformation has been slowing down. Martin Baily, in contrast, argued that the significant externalities connected with additional investments in research and development make a strong case for pushing out the knowledge frontier.

Robert Hall disagreed with the author's contention that providing extremely strong property rights over nonrival goods would lead to inefficiencies and lost opportunities for additional technological progress. Using the computer industry as an example, he noted that IBM has been liberally granted licenses for the use of its key patents. Bresnahan concurred with Hall's assessment and added that for the commercialization of information technology, the most important mechanism for coordination and reuse is the appointment of a monopoly vendor.

References

Arrow, Kenneth J. 1969. ''Classificatory Notes on the Production and Transmission of Technological Knowledge.'' *American Economic Review* 59 (May): 29–35.

Baily, Martin Neil, and Alok K. Chakrabarti. 1988. *Innovation and the Productivity Crisis*. Washington, D.C.: Brookings.

Bush, Vannevar. 1945. *Science: The Endless Frontier: A Report to the President on a Program for Postwar Scientific Research.* Washington, D.C.: National Science Foundation.

Carey, John. 1993. ''Industrial Policy or Industrial Folly?'' *Business Week* (May 17): 38.

Coase, R. H. 1988. *The Firm, the Market, and the Law.* Chicago: University of Chicago Press.

Coe, David T., and Elhanan Helpman. 1993. ''International R&D Spillovers.'' Working Paper 4444. Cambridge, Mass.: National Bureau of Economic Research. August.

Cohen, Linda R., and Roger G. Noll. 1991. *The Technology Pork Barrel.* Washington, D.C.: Brookings.

Crease, Robert P. 1991. ''Yesterday's Pork Projects: Where Are They Now?'' *Science* 254 (November 1): 641–43.

Ergas, Henry. 1987. ''Does Technology Policy Matter?'' In Guile, Bruce R., and Harvey Brooks, eds. *Technology and Global Industry: Companies and Nations in the World Economy*. Washington, D.C.: National Academy Press.

Evenson, R. E. 1993. ''Patents, R&D, and Invention Potential: International Evidence.'' *American Economic Review* 83 (May): 463–68.

Griliches, Zvi. 1958. ''Research Costs and Social Returns: Hybrid Corn and Related Innovations.'' *Journal of Political Economy* 66 (October): 419–31.

———. 1988. ''Productivity Puzzles and R&D: Another Nonexplanation.'' *Journal of Economic Perspectives* 2 (Fall): 9–21.

———. 1992. ''The Search for R&D Spillovers.'' *Scandinavian Journal of Economics* 94 (Supplement): S29–S47.

Katz, Michael L., and Janusz A. Ordover. 1990. ''R&D Cooperation and Competition.'' *Brookings Papers on Economic Activity, Microeconomics: 1990:* 137–203.

Lichtenberg, Frank R. 1992. ''R&D Investment and International Productivity Differences.'' Working Paper 4161. Cambridge, Mass.: National Bureau of Economic Research. September.

Mansfield, Edwin, J. Rapoport, A. Romeo, S. Wagner, and G. Beardsley. 1977. *The Production and Application of New Industrial Technology*. New York: Norton.

Merges, Robert P., and Richard R. Nelson. 1992. ''On Limiting or Encouraging Rivalry in Technical Progress: The Effect of Patent Scope Decisions." University of California (Berkeley) Center for Research and Management.

Nelson, Richard R. 1983. "Government Support of Technical Progess: Lessons from History." *Journal of Policy Analysis and Management* 2 (Summer): 499–514.

Office of Technology Assessment, U.S. Congress. 1991. *Federally Funded Research: Decisions for a Decade.* Washington, D.C.: GPO.

Panel on the Government Role in Civilian Technology, Committee on Science, Engineering, and Public Policy, National Academy of Sciences. 1992. *The Government Role in Civilian Technology: Building a New Alliance.* Washington, D.C.: National Academy Press.

Powers, Nicholas J. 1990. "Federal Marketing Orders for Fruits, Vegetables, Nuts, and Specialty Crops." Agricultural Economic Report 629. U.S. Department of Agriculture, Economic Research Center, Washington, D.C. March.

Rosenberg, Nathan, and Richard R. Nelson. 1993. "American Universities and Technical Advance in Industry." CEPR Publication 342. Center for Economic Policy Research, Stanford University. March.

Savage, James D. 1992. "Trends in the Distribution of Apparent Academic Earmarks in the Federal Government's FY 1980–92 Appropriations Bills." *CRS Report for Congress.* 92-726STR. GPO.

Schmookler, Jacob. 1972. *Patents, Invention, and Economic Change: Data and Selected Essays.* Cambridge, Mass.: Harvard University Press.

Schultz, Theodore W. 1953. *Economic Organization of Agriculture.* New York: McGraw-Hill.

Stigler, George. 1993. "Monopoly." In Henderson, David R., ed. *The Fortune Encyclopedia of Economics.* New York: Warner Books.

Williamson, Oliver E. 1975. *Markets and Hierarchies: Analysis and Antitrust Implications.* New York: Free Press.

Wilson, James Q. 1989. *Bureaucracy: What Government Agencies Do and Why They Do It.* New York: Basic Books.

6 Appendix B

Volume Title: Innovation Policy and the Economy, Volume 1 Volume Author/Editor: Adam B. Jaffe, Josh Lerner and Scott Stern, editors Volume Publisher: MIT Press Volume ISBN: 0-262-60041-2 Volume URL: http://www.nber.org/books/jaff01-1 Publication Date: January 2001

Chapter Title: Should the Government Subsidize Supply or Demand in the Market for Scientists and Engineers?

Volume Title: Innovation Policy and the Economy, Volume 1

Volume Author/Editor: Adam B. Jaffe, Josh Lerner and Scott Stern, editors

Volume Publisher: MIT Press

Volume ISBN: 0-262-60041-2

Volume URL: http://www.nber.org/books/jaff01-1

Publication Date: January 2001

Chapter Title: Should the Government Subsidize Supply or Demand in the Market for Scientists and Engineers?

Chapter Author: Paul M. Romer

Chapter URL: http://www.nber.org/chapters/c10781

Chapter pages in book: (p. 221 - 252)

7

Should the Government Subsidize Supply or Demand in the Market for Scientists and Engineers?

Paul M. Romer, *Graduate School of Business, Stanford University*

Executive Summary

This paper suggests that innovation policy in the United States has erred by subsidizing the private sector demand for scientists and engineers without asking whether the educational system provides the supply response necessary for these subsidies to work. It suggests that the existing institutional arrangements in higher education limit this supply response. To illustrate the path not taken, the paper considers specific programs that could increase the numbers of scientists and engineers available to the private sector.

Preface

My son attends an undergraduate institution that specializes in science and engineering. A degree from this school will cost more than $100,000 in tuition and 4 years of his time. For parents and students who contemplate an investment of this magnitude, the school provides information about labor market outcomes for its graduates. On its web site, the school provides the median salaries for students who accepted jobs after graduation and the Ph.D. completion rates for students who go on to graduate school. If you search, you can see the entire distribution of outcomes—a listing for each student of the starting salary or graduate school in which they enrolled.

If my son pursues a doctoral degree after he graduates, he will have to make an even larger investment. Net of the various sources of support that are available to graduate students, the direct tuition cost of a doctorate will probably be less than the cost of his undergraduate degree. He will, however, have to invest another 4 to 8 years of foregone earnings, which will be substantially higher once he completes his undergraduate degree.

His college is unusual. Most undergraduate institutions do not provide any useful information about labor market outcomes for degree recipients. Yet as *this example shows, it is perfectly feasible for a school to provide this kind of information. Given the stakes, it is even more surprising that many graduate programs in science and engineering also fail to provide this kind of information to prospective students. The paucity of information is obvious to anyone who is familiar with the graduate school application process, but to demonstrate it more formally, I asked a research assistant to begin the application process for the top 10 graduate departments of mathematics, physics, chemistry, biology, computer science, and electrical engineering in the United States. (The rankings were taken from US News and World Report.) For comparison, I also asked him to begin the application process to the top 10 business and law schools.*

In response to his 60 initial requests for information from the science and engineering programs, he received not one response giving information about the distribution of salaries for graduates, either in the initial information packet or in response to a follow-up inquiry from him. In contrast, he received salary information for 7 of the 10 business schools in the application packet, and in response to his second request, he was directed to a web page with salary information by one of the three nonrespondents from the first round. (It is possible that the information could have been found on the web page for the other two business schools, but to maintain consistency in the treatment of the different programs, he did not look for more information if a school did not respond with directions about where to get it.) Four out of the 10 law schools gave salary information in the application packet and three more of them directed him to this information in response to a second request.

I. Introduction

The most important economic policy question facing the advanced countries of the world is how to increase the trend rate of growth of output per capita. In the middle of the 20th century one might have argued that preventing depressions was the more urgent challenge, but at least in the advanced countries of the world, progress in macroeconomic stabilization policy has reduced the threat of a paralyzing economic collapse and even reduced the frequency of mild recessions. In this environment, the lure of better growth policy is compelling. If an economy can increase its trend rate of growth by even a small amount, the cumulative effect on standards of living is too big to ignore.

Many scholars and policymakers are convinced that during the 20th century, rapid technological progress in the United States drove the unprecedented growth in output and standards of living we enjoyed. In addition, they believe that this rapid rate of technological change was fostered by a publicly supported system of education that provided the essential input into the process of discovery and innovation—a steady flow of people trained in the scientific method and in the state of the art in their area of specialization.

If this interpretation of our recent past is correct, it follows that any proposal for achieving an even higher trend rate of growth in the United States should take full account of the detailed structure of our current system of higher education for natural scientists and engineers. Policymakers must recognize that these institutions exhibit puzzling features such as those described in the preface—an almost total lack of information on future market opportunities for students who enter their programs.

Unfortunately, in the last 20 years, innovation policy in the United States has almost entirely ignored the structure of our institutions of higher education. As a result, government programs that were intended to speed up the rate of technological progress may in fact have had little positive effect. We have undertaken major spending initiatives in the area of innovation policy, our most important area of economic policy, without subjecting their economic assumptions to even a cursory check for logical coherence or factual accuracy.

In what follows, I will point to the fundamental conceptual flaw behind the government programs that have been used in the last 20 years to encourage innovation in the private sector. These programs try to stimulate the demand for scientists and engineers in the private sector. To succeed, they depend on a positive supply response that the educational system seems incapable of providing. I will also describe a class of alternative policy programs that could be used to fill the gap created by an exclusive reliance on demand-side subsidies. These alternative programs return to an early style of government policy, one that works directly to increase the supply of scientific and engineering talent.

Section II below starts with a quick recapitulation of the reasons why decisions concerning innovation policy are so important for the economic well-being of future generations. Section III shows how a demand-side innovation policy such as a tax credit for research and development or a program of research grants affects the market for scientists and engineers. It shows why even a well-designed and extremely generous program of this kind will fail to induce more innovation and faster growth if the educational system does increase supply in response to changes in wages. Section IV provides an overview of trends in the supply of scientists and engineers. Sections V and VI look at the market for undergraduates and for Ph.D. recipients respectively. Section VII summarizes and interprets the evidence from these sections.

One of the surprising features of the political debate surrounding demand-subsidy policies is its narrow focus. Few participants in this debate seem to have considered the broad range of alternative programs. Section VIII tries to broaden the debate by suggesting feasible policies that could be considered. More specifically, it outlines a general process that policymakers could adopt for formulating growth policy. This process starts by distinguishing between goals and programs. To be specific, this section outlines four general goals that could guide the formulation of growth policy. The first possible goal that policymakers might adopt would be to target a specific increase in the number of students who receive undergraduate degrees in the natural sciences and engineering. A second would be to encourage more innovation in the graduate training programs that our universities offer to students who are interested in careers in science and engineering. A third would be to preserve the strength of our existing system of Ph.D. education. A fourth would balance amounts that the federal government spends on subsidies for supply and demand of scientists and engineers.

If policymakers in an economy were to adopt goals such as these, the next step would be to design specific programs that are intended to achieve these goals. In broadening the debate, this section also outlines three illustrative programs that policymakers could adopt to achieve these goals. The first is the introduction of training grants to universities that could be used to increase the fraction of undergraduates who receive degrees in natural science and engineering. The second is a system of exams that give objective measures of undergraduate achievement in natural science and engineering. The third is a new type of fellowship, backed by a substantial increase in funding, for students who continue their studies in graduate school.

The advantage of a process that separates goals from programs is that it establishes a natural way to evaluate any specific programs such as these. If the goals are precise and progress toward them can be quantified, then it should be easy to verify if any given program moves the economy closer to the goals. This makes it possible to experiment

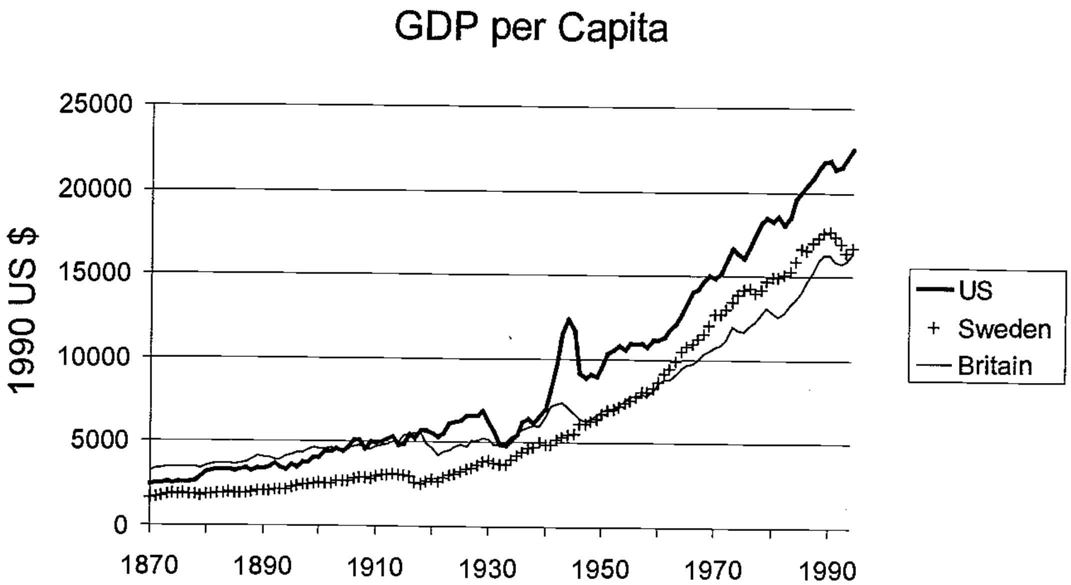


Figure 7.1
Income per capita from 1870 to 1992 for the U.S., Britain, and Sweden. Data are from Angus Maddison (1995).

with a variety of programs, to expand the ones that work, and to shut down the ones that do not.

II. The Importance of Technology Policy

A quick look at the data in figure 7.1 suggests that there must be policy choices, intentional or unintentional, that affect the trend rate of growth. Using data assembled by Angus Maddison (1995), this figure plots income per capita from 1870 to 1994 for the United States, Britain, and Sweden. Over the century-and-a-quarter of data presented there, income per capita grew in the United States at the rate of 1.8% per year. In Britain, it grew at 1.3% per year. In the beginning of the sample period, the United States was a technological laggard, so part of its more rapid growth could have come from a process of technological catch-up with Britain, which was at that time the worldwide technology leader. But even at the beginning of this period, it was clear that the United States was also capable of generating independent technological advances—for example, in the area of manufacturing based on the assembly of interchangeable parts. (See Rosenberg 1969 for an account of the reaction that the "American system of manufactures" caused in Britain by the middle of the 19th century.) Moreover, as the United States surged ahead of Britain in the 20th century, it maintained the

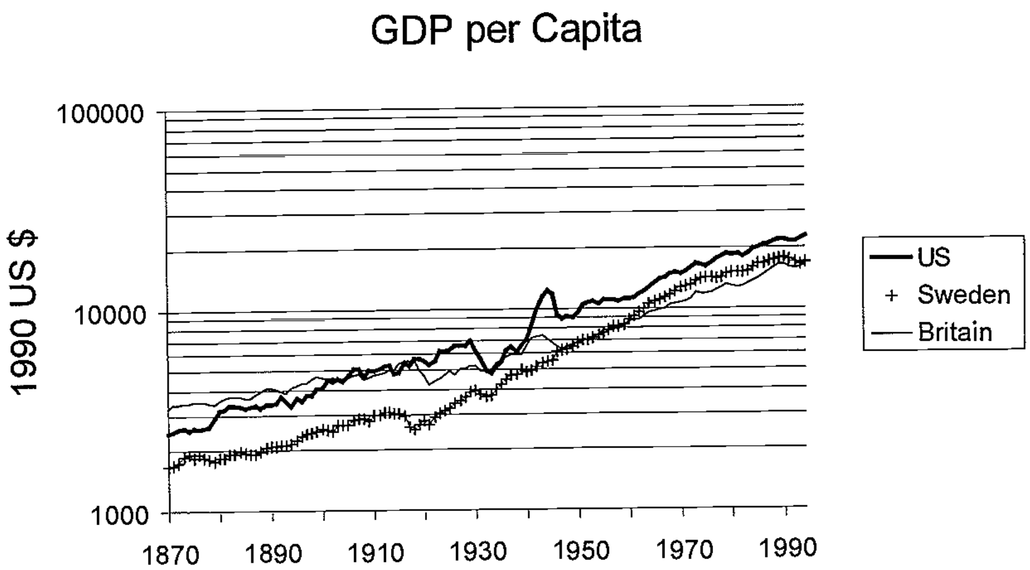


Figure 7.2
Same data as figure 7.1, on a logarithmic scale.
Data are from Angus Maddison (1995).

faster rate of growth that was apparent from the beginning. This is most clearly evident in figure 7.2. Because the data are plotted on a logarithmic or ratio scale, straight lines in the figure correspond to constant rates of growth. In the second half of the century, the rate of growth in Britain accelerated moderately. The rate of growth that had been initiated in the U.S. remained essentially unchanged. The policies and institutions in the United States made possible a trend rate of growth of income per person that was significantly faster than the trend that had pushed Britain into the position of worldwide leadership in the 19th century. Given the limited state of our knowledge of the process of technological change, we have no way to estimate what the upper bound on the feasible rate of growth for an economy might be. If economists had tried to make a judgment at the end of the 19th century, they would have been correct to argue that there was no historical precedent that could justify the possibility of an increase in the trend rate of growth of income per capita to 1.8% per year. Yet this increase is what we achieved in the 20th century.

The experience in Sweden suggests that even higher sustained rates of growth of income per capita can sometimes be possible. During the 50 years from 1920 to 1970, income per capita in Sweden grew at the much higher rate of 3% per year. Once again, this faster rate of growth could be due, at least in part, to the process of technological catch-up.

Moreover, growth in Sweden has slowed down considerably since 1970. Nevertheless, the experience in Sweden should at least force us to consider the possibility that if we arranged our institutions optimally, growth in the United States could take place at an even higher rate than that to which we have become accustomed. In the coming century, perhaps it will be possible to increase the rate of growth of income per capita by an additional 0.5% per year, to 2.3% per year.

The implications of a change of this magnitude would be staggering. For example, according to recent CBO estimates that were based on continuation of the historical trend rate of growth, in the year 2050 the three primary government entitlement programs—Social Security, Medicare, and Medicaid—will require an increase in government spending equal to about 9% of projected GDP. If the rate of growth over the next 50 years were to increase by 0.5% per year, GDP in 2050 would be 28 percentage points larger. By itself, faster growth could resolve all of the budget difficulties associated with the aging of the baby boom generation, and still leave ample resources for dealing with any number of other pressing social problems. And of course, the longer a higher growth rate can be sustained, the larger the effect it will have. By the year 2100, the additional 0.5% per year would translate into a GDP that is 1.65 times as large as it would otherwise have been.

Other types of evidence suggest that an increase in the rate of growth of 0.5% per year is not beyond the realm of possibility. In his survey of returns to investment in R&D, Zvi Griliches (1992) reports a wide range of estimates for the social return, with values that cluster in the range of 20% to 60%. Take 25% as a conservative estimate of the social return on additional investment in R&D. If we were to increase spending on R&D by 2% of GDP (and maintain the same rate of return on our investments in R&D—more on this in the next section) the rate of growth of output would increase by the hoped-for 0.5% per year. If the true social return is higher, say 50%, the extra investment in R&D needed to achieve this result would be correspondingly lower, just one additional percent of GDP. These estimates are also consistent with other estimates, which suggest that the level of resources currently devoted to research and development may be far below the efficient level. For example, after they calibrate a formal growth model to the results from micro level studies of the productivity of research and development, Chad Jones and John Williams (1998) calculate that the optimal quantity of resources to devote to research and development could be four times greater than the current level.

There is another way to look at estimates of this kind, one that is closer to the spirit of the analysis that follows below. GDP in the United States is about $10 trillion dollars. One percent of this would be $100 billion per year in additional spending on R&D. If it costs $200,000 per year to hire and equip the average worker in this sector, this means that we would need to increase the stock of workers employed in R&D by roughly 500,000. The question that policymakers must confront if they are serious about increasing the amount of R&D that is performed is where these additional high-skilled workers will come from.

There is no certainty that growth would necessarily speed up even if we did undertake all the right steps in an effort to do so. There is ambiguity in the historical record, and even if there were not, there is no guarantee that relationships that held in the past will continue to hold in the future. Moreover, even in the best case, we should recognize that there might be substantial lags between the initiation of better policy and the realization of faster output growth. For example, one highly successful example of a government policy that did increase the rate of technological change was the creation of the new academic discipline of computer science in the 1960s. (See Langlois and Mowery 1996 for a discussion of the episode.) Even now, with the passage of 40 years, our sense of the magnitude of the payoff from this investment is still growing.

Notwithstanding all these caveats, a possibility, even a remote possibility, of a change as profound as another permanent 0.5% increase in the trend rate of growth in the world's leading economy ought to excite the imagination. Compared to this, even landing a man on the moon would seem a minor achievement. One would think that this kind of possibility would inspire us to try new things, to make every effort to understand what will work and what will not as we strive for this goal. By this kind of standard, the efforts we have made in the last 2 decades have been remarkably timid and poorly conceived.

III. Demand Subsidies

Unless one is careful and makes use of some simple economic theory, it is easy to fall into an all-too-common trap in discussions about innovation policy. The key point was signaled above in the switch from a discussion of spending on R&D to a discussion of the number of workers engaged in this activity. To speed up growth, it is not enough to increase *spending* on research and development. Instead, an economy must increase the total *quantity of inputs* that go into the process of research and development. Spending is the product of a quantity and a price. To simplify the discussion, assume for now that people—scientists and engineers—are the key inputs in research and development. Formally, let E stand for spending on research and development and let N represent the number of scientists and engineers working in this area. Let w represent the average wage for these workers. Then trivially, $E = N \times w$. An increase in expenditure E will not necessarily translate into a corresponding increase in N, the number of scientists and engineers engaged in R&D. In principle, it is entirely possible for the entire increase in E to pass through as increases in w.

Continuing with the simplifying assumption that scientists and engineers are the only inputs in the production of research and development, we can illustrate how w is determined using the simple supply and demand framework presented in figure 7.3. The horizontal axis measures the number of scientists and engineers working in the private sector on R&D. The vertical axis measures their wage. The downward-sloping demand curve indicated by the solid line represents the

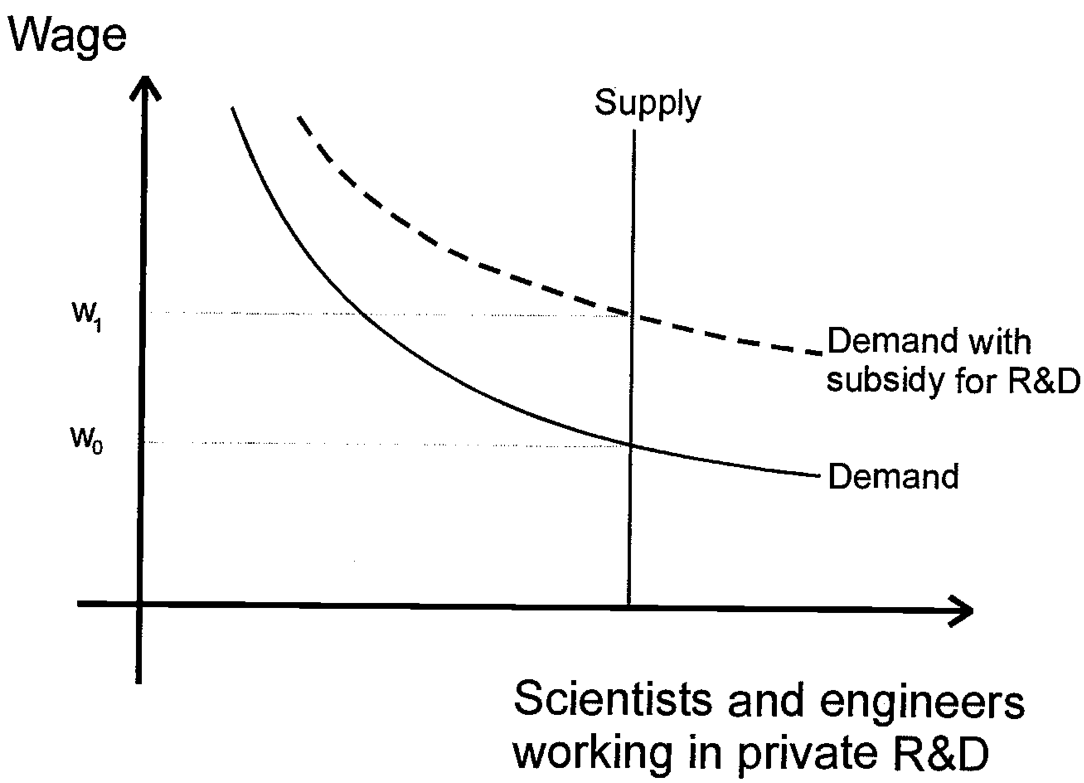


Figure 7.3
Wage w versus number of scientists and engineers working in private R&D

private return captured by a firm that hires some additional scientific workers and undertakes more research and development.

In the figure, the supply of scientists and engineers is represented by a vertical supply curve. The vertical slope of the curve reflects an assumption that the number of young people who become scientists and engineers and go to work in the private sector does not adjust in response to an increase in the wage that they receive for employment in the private sector. Section IV below outlines a much more complicated picture of the supply response of our educational system, but it is useful to start with the simple case of zero supply elasticity. To motivate this assumption, it is enough to keep the story from the preface in mind. The lack of information that is available to students who are making decisions about careers in science and technology suggests that our existing educational institutions may not lead to the kind of equilibration that we take for granted in many other contexts. If students do not have information about what wages will be, it will be much harder for them to adjust their career decisions in response to wage changes.

The downward-sloping dashed line in the figure represents the private demand for research workers when the government provides a subsidy for R&D. This subsidy could take the form of special tax advantages such as those afforded by the research and experimentation tax credit offered in the United States. Alternatively, the subsidy could take the form of cash payments to some firms as part of a cost-sharing agreement in which the government pays part of the cost of a research and development program. This is the kind of subsidy offered by partnership programs such as the Small Business Innovative Research (SBIR) grant program administered by the Small Business Administration or the Advanced Technology Program (ATP) administered by the Department of Commerce. Whether it comes in the form of tax credits or research grants, the effect of government spending is to shift up and to the right the demand for scientists and engineers who can perform the R&D.

From the perspective of a single firm, it seems obvious that a special tax incentive or a research grant will encourage the firm to hire more scientists and engineers and thereby to cause more inputs to be devoted to R&D. Yet one of the most basic insights in economics is that for the economy as a whole, things have to add up. If the total number of scientists and engineers is fixed, it is arithmetically impossible for employment of scientists and engineers to increase at all firms. As illustrated in the figure, if the supply curve of scientists and engineers is fixed, then the increase in demand induced by the subsidy will translate into a proportional increase in wages for scientists and engineers with no increase in the inputs that are devoted to R&D.

It is important to recognize that this argument is separate from the usual concerns about "additionality" that have been raised with respect to R&D demand subsidy programs. People who focus on this problem are worried about how much the demand curve shifts. That is, they are worried that an additional dollar in subsidies does not translate into much additional private spending on R&D. This is a nontrivial issue. The evidence does seem to suggest that more generous tax treatment for R&D leads to higher *reported* levels of spending on R&D at firms. (See for example Bronwyn Hall and John van Reenen 1999.) An additional dollar in tax benefits seems to lead to about one additional dollar in reported R&D expenditure by firms. However, there is much less evidence about the extent to which this increase in reported R&D spending represents a true increase in spending relative to that which would have taken place in the absence of the credit. It is quite possible that some of this spending comes from relabeling of spending that would have taken place anyway. Deciding what qualifies for this credit is apparently a nontrivial problem for the tax authorities. Between 20% and 30% of claimed expenditures by firms are disallowed each year (National Science Board 1998:4-48).

For the SBIR program, Josh Lerner (1999) finds that firms that receive grants from the government experience more rapid sales and employment growth than a comparison group of firms selected to be similar to the recipient firms. This could be an indication that firms that receive grants do devote more inputs to R&D. But it could also reflect unobserved, intrinsic differences between the control group, which was constructed ex post by the researcher, and the recipient group, which was selected on the basis of a detailed application process that was designed to select particularly promising firms. In related work, Scott Wallsten (2000) finds that firms that receive a research grant from the government under the SBIR program seem to substitute these grant funds for other sources of funds, with little or no net increase in spending on R&D.

For both the tax credit and direct grant programs, we can identify a coefficient m which measures the true increase in private spending on R&D associated with each additional subsidy dollar from the government. In each case, there is some uncertainty and debate about how large this coefficient is. But for any positive value of m, the argument outlined above shows that the entire increase in spending may show up as higher wages for the existing stock of workers, with no increase in the actual quantity of research and development that is performed. As a result, even a well-designed and carefully implemented subsidy could end up having no positive effect on the trend rate of growth for the nation as a whole.

Recent work by Austan Goolsbee (1998) suggests that, at least in the short run, the wage changes implied by a weak supply response are apparent in the data. He compares census data on wages for research workers with time series data that capture the variation in government spending on R&D. Direct government spending is well suited for this kind of analysis because it does not suffer from the concerns about additionality that are present for government subsidies for R&D. Surprisingly, using only these crude data, he finds strong effects on wages. For example, during the defense build-up between 1980 and 1984, federal spending on R&D increased, as a fraction of GDP, by 11%. His estimates suggest that this increased wages for physicists by 6.2% and aeronautical engineers by 5%.

In the face of this argument, defenders of demand-side R&D subsidies can respond in three ways. First, they can argue that people are not the only inputs used in R&D. If other inputs such as computers and specialized types of laboratory equipment are supplied elastically, then government subsidies for R&D could increase the utilization of these other inputs even if the number of scientists and engineers remains constant. If this were truly the intent of the various subsidy programs, it would be much more cost-effective for the government to provide the subsidies directly for these other inputs. Salaries account for the majority of total R&D spending. For example, in university-based research, annual research expenditures on equipment during the last decade have varied between 5% and 7% of total research expenditures (National Science Board 1998). If the goal of the subsidy program were to increase the equipment intensity of research and development and if the ratio of spending on equipment in the private sector is comparable to the figure for universities, a special tax subsidy for the purchase of equipment used in research would be substantially less costly than one that is based on total expenditures including salaries. Similarly, the government could achieve substantial savings, and still increase the use of equipment in R&D, if it restricted the grants provided by the SBIR and ATP programs so that these funds could be used only for additional purchases of equipment.

In the case of the targeted grant programs administered by the ATP or the SBIR, a defender could argue that even if the existing research subsidies do not increase employment of scientists and engineers in the economy as a whole, they can increase employment at the recipient firms, at the cost of a reduction in employment at other firms. If government agencies were able to identify an allocation across firms and projects that is better than the one the market would implement, the targeted grant programs could still be socially valuable. But even the strongest supporters of the subsidy programs are hesitant to make this kind of claim about the superiority of government allocation processes. Note also that because the research and experimentation tax credit is available to all firms, it cannot be justified on this kind of basis of any hypothesized ability of the government to improve the allocation of research inputs between firms and projects.

If the goal is not to encourage equipment investment in the R&D sector or to give the government a bigger role in deciding how to allocate scarce R&D personnel, some other motivation must lie behind these spending programs. The final response could be for a defender of these programs to dispute the basic assumption behind the supply-and-demand model outlined here and argue that, at least in the long run, the supply of scientists and engineers working in R&D in the private sector does respond to demand-induced changes in wage. But to make this case, one must confront some of the peculiar features of the educational system that actually produces these highly skilled workers and ask if there are more cost-effective ways to increase the supply of these types of workers.

IV. Overview of the Supply of Scientists and Engineers

Figures 7.4 and 7.5 give a broad overview of trends in the supply of scientists and engineers in the United States. Figure 7.4 updates data presented by Chad Jones (1995) on the number of scientists and engineers in the United States who are employed in research and development. These data are scaled by the size of the labor force. They show an increase in R&D employment as a fraction of the labor force from about 0.3% of the labor force in 1950 up to about 0.8% in the late 1960s, with no strong trend thereafter. The underlying data for this figure are collected by the NSF. (Data since 1988 are taken from Table 3-15 from National Science Board 1998.)

Official statistics on formal research and development capture only part of the private sector effort directed at innovation. Also, no consistent data series on employment in R&D is available in years prior to 1950. To give a more comprehensive overview of the proportions of skilled workers in the labor force over a longer time horizon, figure 7.5

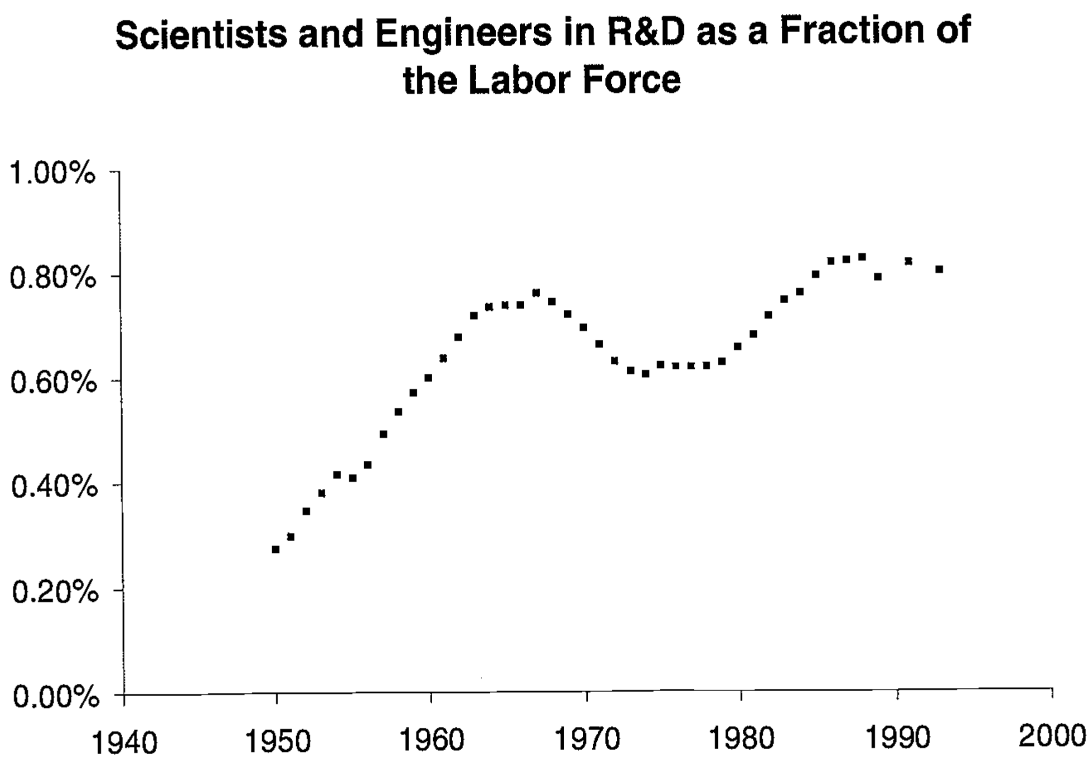


Figure 7.4
Scientists and engineers in R&D as a fraction of the labor force
Data from Chad Jones 1995.

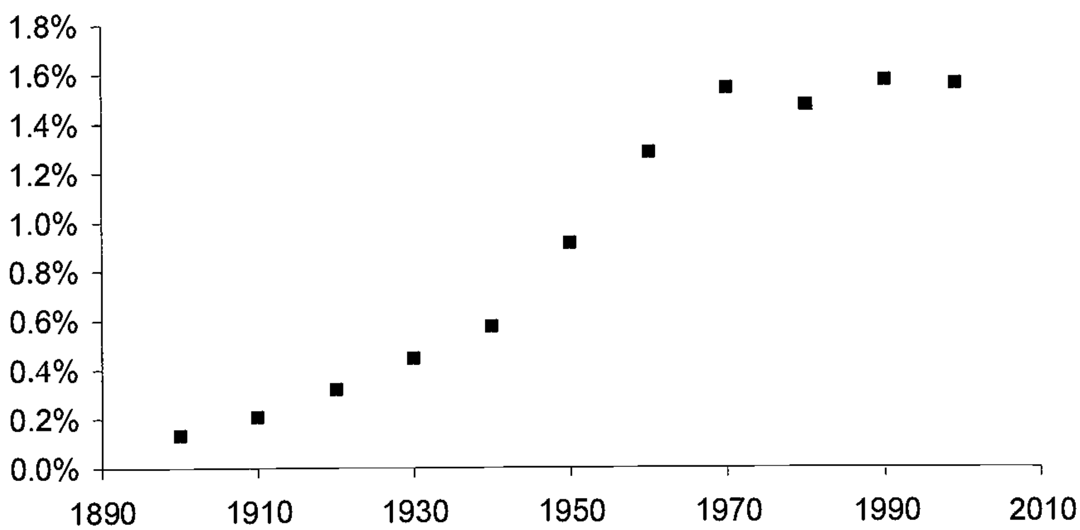


Figure 7.5
Engineers as a fraction of the labor force
Data are from Historical Statistics of the United States, 1975, and Statistical Abstract of the United States, various years.

presents data on the total number of engineers as a fraction of the labor force, using occupational data collected by the Bureau of the Census. This series shows a similar pattern. Engineers increase, as a fraction of all workers, from the turn of the century up until 1970, and remain roughly constant thereafter.

Taken together, these figures offer little reassurance that the aggregate supply of scientists and engineers responds efficiently to market demand. Of course, it is logically possible that the growth in the demand for scientists and engineers experienced a sharp fall starting in the late 1960s. However, other labor market evidence based on relative wages such as that presented by Katz and Murphy (1992) suggests that a process of skill-biased technological change that raised wages for skilled relative to unskilled workers continued at about the same pace in the 1970s and 1980s as in the 1960s. Other work (see for example, Autor, Katz, and Krueger 1998) suggests that, if anything, the rate of skill-biased technological change actually increased in the period from 1970 to 1995 relative to the period from 1940 to 1970. Taken together, these data on quantities plus the independent evidence on the demand for skill suggest that one look more carefully at other possible factors that could influence the supply of scientists and engineers.

Figure 7.6 gives a schematic outline of the process that actually determines the supply of scientists and engineers. The two key stages in the production process are undergraduate education and graduate education. (For simplicity, graduate programs that lead to a terminal master's degree are grouped in this figure with those that provide Ph.D. level training.) The first major branch in the process distinguishes undergraduates who receive degrees in the natural sciences or engineering (NSE degrees) from those who receive all other types of degrees. Section V below looks at the possible nonmarket forces that could constrain this decision. After a student receives an undergraduate NSE degree, she can either go to work in the private sector or continue on to receive graduate training. Section VI looks at recent developments in the market for people with an advanced degree in the natural sciences or engineering.

V. The Supply of Undergraduate Degrees in Science and Engineering

The market for education suffers from pervasive problems of incomplete information. Students contemplating a choice between different institutions typically have very little information about the

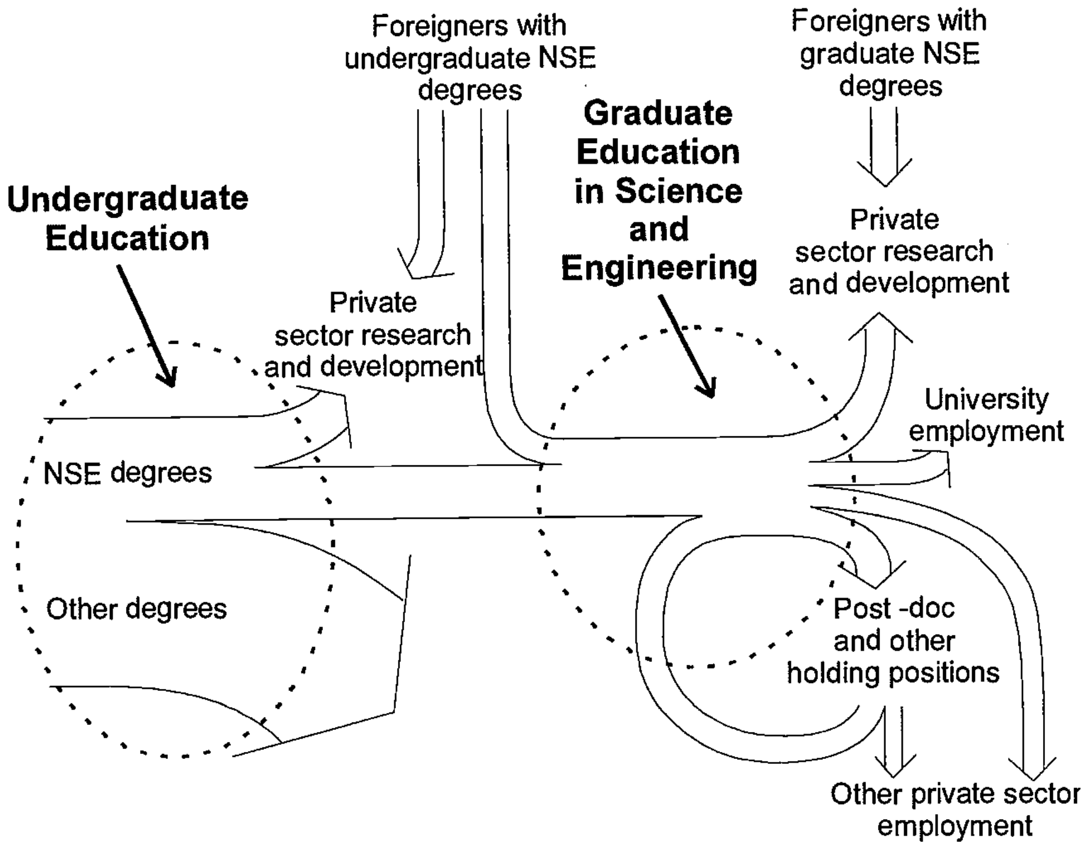


Figure 7.6
Schematic outline of the process determining the supply of scientists and engineers.

value-added they can expect to receive from one institution versus another. Employers selecting among graduates from different institutions also have very little objective basis for judging the absolute achievement levels of students from different schools, or even about students from the same school who have followed different courses of study. The competitive strategy that seems to have emerged in this market is one where undergraduate institutions have developed extensive systems for screening students by ability level. They enroll the most able students that they can attract. The schools compete for these students in large part by publicizing the degree of selectivity. Students, in turn, compete for admission to the most selective institutions because a degree from a more selective institution offers a stronger signal about the student's ability. Using data from the different campuses of the University of California, Robert Frank and Philip Cook (1995) suggest that competition along these lines has been getting more intense. For example, over time, SAT scores for students attending Berkeley, the UC campus that is perceived to be the most selective, have been increasing relative to SAT scores at other campuses.

In this kind of competitive environment, the traditional liberal arts university may face little pressure to respond to changing market demands for different types of skills. For example, imagine that government subsidies increased the market wage for scientists with several years of training beyond the undergraduate degree. Imagine that students are somehow informed about this change in the wage and respond by increasingly enrolling in undergraduate science courses that will prepare them for further study in engineering and science. A liberal arts university that has a fixed investment in faculty who teach in areas outside of the sciences and that faces internal political pressures to maintain the relative sizes of different departments may respond to this pressure by making it more difficult for students to complete a degree in science. Faculty in the departments that teach the basic science courses will be happy to "keep professional standards high" and thereby keep teaching loads down. Faculty in other departments will be happy to make study in their departments more attractive, for example by inflating the average grade given in their courses.

There is clear evidence that this kind of response currently operates on campuses in the United States. First, the number of students who begin their undergraduate careers with the intent of receiving a degree in science and engineering is substantially higher than the number who actually receive such a degree. For example, for white students, 12% of entering students intend to major in natural sciences and 9% plan to major in engineering. Only 8% of graduating students actually receive a degree in natural sciences and only 5% receive a degree in engineering (National Science Board 1998). For minority students, the attrition rate is even higher.

One additional indication of the pressure to shift students out of science and engineering degrees comes from the difference in the distributions of grades offered in courses required for degrees in these areas as opposed to grades in other courses of study. Measuring this difference is not straightforward because even within a department such as mathematics, and even within a specific subject area such as linear algebra, there are courses with easier grade distributions that are intended for people who will not continue toward a degree in science, and courses with a lower distribution of grades for people who will.

For example, students who place out of the basic calculus course on the basis of an advanced placement exam are more likely to take more difficult math courses than students who do not. This tends to lower the average grade they receive in the second-level math courses that they take. If one does not correct for this fact, one finds that math grades for the students who place out of calculus are not, on average, any higher than math grades for students who do not place out of calculus. However, if one holds constant the specific second-level math courses that students take and compares grades for students with different backgrounds who take the same course, it is clear that students who have placed out of calculus do receive higher math grades than other students taking the same class (Rick Morgan and Len Ramist 1998).

To do this kind of analysis, the College Board, which administers the advanced placement exam, collected data from a representative sample of 21 selective universities. Using these data, one can do a direct comparison of grade distributions across different fields of study. Take, for example, the sample of second-level math courses that students who place out of calculus attend. These tend to be biased toward the classes that students majoring in mathematics or the natural sciences will take. One can then compare the distribution of grades in these courses with the distribution of grades in second-level English courses taken by students who receive advanced placement credit in English composition; or with the distribution of grades in second-level history courses taken by students who receive advanced placement credit in American history. As table 7.1 shows, in the selected math courses, 54% of all students received a grade of A or B. For the English courses, the fraction with an A or a B is 85%. For the history courses, the fraction is 80%. For social science courses such as political science or economics, the fraction of students who receive a grade of either A or B is about 75%.

As figure 7.6 shows, immigration is an alternative source of supply for the labor market in the United States. If the domestic supply of scientists and engineers is constrained to a significant extent by our existing system of undergraduate education, one should see evidence that the response in terms of undergraduate NSE degrees differs from that of immigrants. Recently, much of the discussion of migration has focused on political pressure from technology-intensive firms for increases in the number of H1B visas that permit private firms to hire skilled workers from abroad to fill entry-level jobs in areas such as computer programming. This debate has obscured the extremely important role that immigration has long played in supplying scientists and engineers with the highest levels of skill. Moreover, immigration is clearly responsive to demand conditions. Fields such as computer science and engineering, where indicators suggest that market demand is high relative to the available supply, are the ones that have experienced

Table 7.1
Fractions of students receiving an A or a B in different subjects

Subject Area	Fraction of Students Receiving a Grade of A or B
English	85
History	80
Economics and Political Science	75
Mathematics	54

the largest inflows from abroad. For example, in 1993, 40% of the people in the United States who had a Ph.D. degree in engineering were foreign-born. In computer science, 39% of the Ph.D. holders were foreign-born. In the social sciences, where demand for new Ph.D. recipients is generally much lower (economics being a notable exception), only 13% of the Ph.D. holders were foreign-born (National Science Board 1998). These immigration flows stand in sharp contrast to the trends in undergraduate education. From the mid-1980s until 1995, the number of undergraduate degrees in engineering and in mathematics and computer sciences fell substantially. For example, in the 1980s and 1990s, as the personal computer and internet revolutions were unfolding, the number of undergraduate degrees in computer science showed no strong trend, increasing at first in the mid 1980s, then falling in the 1990s and ending at about the level at which it started in the early 1980s.

Engineering degrees follow a very similar pattern (National Science Board 1998: Table 2-20). Between 1981 and 1995 there is no change in the number of undergraduates who receive degrees in engineering. The number does increase in the late 1980s but then returns to the previous level. For future reference, note that the number of master's degrees in engineering behaves quite differently. From 1981 to 1995 the total number of master's degrees in engineering increased steadily so that the number in 1995 was about 1.7 times the number in 1981 (National Science Board 1998: Table 2-27).

Another sign that the domestic enrollment of students who are able to continue in science and engineering is a critical bottleneck comes from an examination of downstream developments in Ph.D. education. From the mid-1970s to the mid-1980s, the number of Ph.D. degrees awarded in the United States each year in natural sciences and engineering remained roughly constant at about 12,500. (Here, as elsewhere, natural sciences and engineering exclude behavioral and social sciences.) Then, starting in 1986, this number began a steady increase up to 19,000 per year in 1995.

We can use this expansion in the size of Ph.D. programs to gauge the elasticity of the foreign supply response and compare it to the domestic supply response. In 1986, U.S. citizens accounted for about 8,000 of Ph.D. degree recipients, and noncitizens accounted for the other 4,500. In 1995, the number of degrees for U.S. citizens had increased by about 20% to around 10,000 and the number of degrees awarded to noncitizens had more than doubled to 9,000 (National Science Board 1998: Table 2-35).

A similar, though less extreme picture emerges from an examination of master's degrees, particularly in the high-demand areas of computer science and engineering. As market opportunities for holders of the master's degree increased and universities added to the number of slots that they made available in master's degree programs, foreign students responded more strongly than U.S. citizens, just as they did when new positions in Ph.D. programs opened up. In 1975, foreign students received 22% of the master's degrees in engineering and 11% of the master's degrees in math and computer science. By 1995, foreign students accounted for 39% of the master's degrees in engineering and 35% of the master's degrees in math and computer science (National Science Board 1998). In both instances, increased downstream demand for undergraduates with NSE degrees does not seem to have induced a sufficient supply response. The system equilibrated by importing more foreigners.

VI. The Supply of Ph.D. Degrees in Science and Engineering

The sharp increase in the 1990s in the number of Ph.D.s granted has been accompanied by generally declining job prospects for degree recipients. In the most recent period, it is possible that part of the reason why undergraduate students did not pursue degrees in the natural sciences is that they were vaguely aware of the worsening job prospects that Ph.D. recipients faced. Note, however, that developments in the academic market for Ph.D.s cannot explain the absence of an increase in undergraduate degrees in engineering or in specialized areas such as computer science where job prospects for Ph.D. recipients have remained strong. Also, the weak market for new Ph.D.s would only have been a factor fairly recently, primarily since 1990 when the increased supply of Ph.D. recipients began to show up on the market. Nevertheless, going forward, the weak academic market for some types of Ph.D.s will certainly be a complicating factor in any attempt at increasing the number of undergraduate degrees that are awarded in natural science and engineering. To increase the number of undergraduates who receive an undergraduate degree in the natural sciences and engineering, they must be convinced that this kind of degree can lead to better career outcomes than the dead-end postdoctoral positions that have become increasingly common in some fields.

Independent of its role in influencing undergraduate degrees in the United States, understanding the behavior of the market for Ph.D.s is critical to the formulation of policy concerning the supply of scientists and engineers. The thrust of the possible programs outlined below is to substantially increase this supply. Yet many people in the academic community are convinced that the most pressing science policy issue in the United States is the Ph.D. glut. They have advocated measures that would reduce the supply of Ph.D.-level scientists and engineers. A careful look at the market for Ph.D.s is necessary to explain why increases in the supply of scientists and engineers with several years of graduate training are still called for even in the face of difficulties in the labor market for Ph.D.s. The key point here is to distinguish between people who are trained exclusively for employment in research universities and people who can work in research and development in the private sector.

Look again at figure 7.6. Events in the Ph.D. market can be summarized in terms of this figure. As noted above, the total flow of students through NSE Ph.D. programs increased starting in the late 1980s and continuing through the 1990s. Much of this flow has been directed at two of the alternatives upon leaving graduate school—university employment and postdoc and other holding positions. The challenge in this area is not to increase the total numbers of Ph.D. degree recipients, but to increase the fraction of them that can put their skills to work in private sector research and development.

This pattern of outcomes—increased numbers of Ph.D. recipients and steadily worsening academic job prospects—can be explained by increased subsidies for Ph.D. training. These subsidies derived from increased support for university-based research, which is complementary to Ph.D. training. As a result, the nature of the support for graduate students changed along with the level. Consider the sample of students who received their primary means of support for their Ph.D. education from the federal government. Between 1980 and 1995, the fraction whose primary mechanism of support was a traineeship fell from 25% to 15% and the fraction whose primary mechanism was a research assistantship increased from 55% to 63%. The fraction receiving their primary support from fellowships stayed roughly constant at about 10%. Among students whose primary support was from sources other than the federal government (primarily state governments), research assistantships also increased by about 10 percentage points (National Science Board 1998: Chapter 5).

Because this increase in supply consisted of people who planned to pursue academic research appointments, the increased supply of Ph.D. recipients was accompanied by generally worsening job prospects for Ph.D. recipients in the academic market. For example, consider in any year the sample of people with degrees in the natural sciences and engineering who were working in academic institutions and who had received their Ph.D. degree within the previous 3 years. In the early 1980s, there were about 17,000 of these recent degree recipients working in academic institutions. About half of them had faculty jobs. The rest held postdoctoral positions or some other form of appointment. By 1995, this same measure of recent Ph.D.s in academic institutions had increased to 23,000, but the number holding faculty positions remained roughly constant, at about 8,500. The entire increase of 6,000 recent degree recipients is accounted for by increases in nonfaculty appointments (National Science Board 1998: Table 5-29).

The problems in the academic market in the life sciences were documented in a report from the National Research Council (1998). In the last decade, this is the area that has benefited from the most rapid rate of growth of federal research support. Between 1970 and 1997, the median time to receipt of a degree increased by 2 years to a total of 8 years. The number of people who hold a postdoctoral appointment 3 or 4 years after receipt of the Ph.D. increased from 6% to 29% between 1973 and 1995. The fraction of Ph.D. recipients who do not hold a permanent full-time job in science and engineering 5 or 6 years after they have received their degree increased from 11% in 1973 to 39% in 1995. The 1995 data, which were the most recent available at the time that the National Research Council wrote its report, reflect long-term outcomes for the 1989–90 cohort of Ph.D. recipients. Because of the steady increase in the number of degree recipients throughout the 1990s, the competitive pressures in this field have probably worsened still further.

VII. An Interpretation of the Evidence Concerning Higher Education

The picture that emerges from this evidence is one dominated by undergraduate institutions that are a critical bottleneck in the training of scientists and engineers, and by graduate schools that produce people trained only for employment in academic institutions as a side effect of the production of basic research results. This description of the system as a whole hides a heterogeneous mix of different types of institutions. Not all of them will behave according to the description given above.

For example, the pressure to keep enrollments down in the natural sciences and engineering will not be present at institutions that specialize in this kind of training. They may therefore face different kinds of incentives and behave differently in the competition for students. The institution that my son attends, Harvey Mudd College, is one of these specialized institutions, and this may explain why it features information about the market outcomes for its graduates more prominently than traditional liberal arts universities. A quick check of data from other schools is consistent with this observation. MIT and Caltech, two selective schools that also concentrate in science and engineering, present information about median salaries for their undergraduates on the web pages that provide information for potential applicants. Harvard and Stanford, two comparably selective institutions that cover the whole range of academic disciplines, apparently offer no information on their web pages about salaries or enrollments in graduate school.

One natural question that the model outlined here does not address is why competition by entry of more schools like Harvey Mudd, MIT, and Caltech has not partially solved the bottleneck problem described here. Mudd, which is about 50 years old, is a relatively recent entrant in this market, but in general, entry seems to be a relatively small factor in the competition between undergraduate institutions. Presumably the incomplete information available to students and employers about the quality of the education actually provided at any institution is a big factor limiting the entry process, but the nature of competition between schools deserves more careful consideration.

There are also different types of institutions that provide graduate education. The description offered here focuses primarily on graduate education in the sciences, which takes place almost exclusively within institutions where the revenue and prestige associated with research are more important motivating forces than tuition revenue. Training in these departments differs sharply from the kind of training offered by professional schools where income from tuition is a much more important determinant of institutional incentives. It should not be surprising that, as my research assistant discovered, business schools and law schools follow very different strategies from the ones used by departments of science when they compete for students. In many ways, master's level training in engineering is like these professional schools. Much of the income associated with these programs comes from tuition. Departments that get to keep a portion of this master's level, but not of undergraduate tuition revenue, should therefore be willing to expand the size of their master's programs at the same time that they put limits on the size of their undergraduate programs. These kinds of incentive effects may help explain why master's degree programs in engineering have shown steady growth while undergraduate engineering degrees have not.

In its report on career prospects in the life sciences, the National Academy Board on Biology concluded that policymakers should restrain the rate of growth of graduate students in the life sciences. In my language (not theirs) they also recommended that graduate education in the life sciences be reshaped along lines that are closer to those followed by professional schools. They recommended that students be given more information about career prospects, that they be given training that prepares them for employment in jobs outside of university-based research, and that funds that support the training of graduate students be shifted away from research assistantships and toward training grants or other forms of support that give more control over a student's education to the student.

This last and most controversial recommendation is the one that has the greatest potential to shift the traditional science-based model of graduate education closer to the model that we see in master's level professional schools of tuition-paying customers who collectively can exert a significant degree of control over what happens during the process of education. Similar proposals for modifying Ph.D. training have been made by a variety of study panels. All have received mixed support at best from the scientific community as a whole. (See the discussion of this point in National Science Board 1998: 5-33.)

Opposition to any change in the form of support for graduate education is usually justified in public on the grounds that there is insufficient evidence about what the effects might be for any change in the system of funding for graduate students. A more fundamental problem—one that goes largely unreported in print but that prominent scientists are willing to justify in private—is that the current funding and training system, one that puts graduate students in the position of apprentices to established scientists and that does not prepare students for careers outside of science, is crucial to the maintenance of the institutions of academic science. Recent work by Scott Stern (1999) offers convincing evidence that recipients of Ph.D. degrees exhibit a strong preference for engaging in the activities in science and are willing to accept substantial wage reductions if doing so will allow them to continue to pursue these activities. This preference could be the result of a selection process that attracts people with this taste into Ph.D. training in science, a training process that cultivates this taste, or a combination of the two. Regardless of the mechanism, any attempt to make the training of Ph.D. students resemble more closely the training of students in business schools could have the effect of significantly undermining the commitment to the ideas and process of science that Stern is able to document. This commitment, which may be psychologically and functionally similar to the commitment induced by training for membership in a religious order or a military unit, may be critical to the preservation of the institutions of science. Unfortunately, it may also help explain why the existing system of graduate education seems so poorly suited to training people for employment outside of academic science. For this combination of reasons, the task of modifying the educational system that trains scientists and engineers may be both very important and very delicate.

VIII. Goals and Programs

To formulate growth policy, policymakers may want to start by distinguishing goals from programs. Goals should be conservative. They should represent objectives that are neither risky nor radical and for which there is a broad base of intellectual and political support. Goals should remain relatively constant over time. They should also imply metrics for measuring success. By these criteria, increasing the long-run trend rate of growth is not specific enough to be a goal. It is appropriately conservative and should be the subject of a broad consensus, but because it is so difficult to measure the trend rate of growth, it does not imply any workable metrics that we can use to measure progress toward the goal. In contrast, increasing the fraction of young people in the United States who receive undergraduate degrees in science and engineering could qualify as a goal. So could increasing the total quantity of resources that are devoted to research and development.

In contrast to a goal, a program is a specific policy proposal that seeks to move the economy toward a specific goal. For example, the Research and Experimentation tax credit is a specific program that is designed to achieve the goal of increasing the resources used in research and development. It should be possible to judge the success of a program against the metric implied by the goal that it serves. All programs should be designed so that they can be evaluated on a policy-relevant time horizon. If they are, they can also be less conservative and more experimental than the underlying goals. A variety of programs could be tried, including ones where there is some uncertainty about whether they will succeed. If the evidence shows that they do not work, they can be modified or stopped.

To illustrate how this framework could facilitate better analysis of the growth process, it helps to focus on a specific set of hypothetical goals. Imagine that policymakers and the public at large accepted the following goals because they want to increase the long-run rate of growth in the United States. (1) Increase the fraction of 24-year-old citizens of the United States who receive an undergraduate degree in the natural sciences and engineering from the current level of 5.4% up to 8% by the year 2010 and to 10% by 2020. (2) Encourage innovation in the graduate training programs in natural science and engineering. (3) Preserve the strengths of the existing institutions of science. (4) Redress the imbalance between federal government subsidies for the demand and supply of scientists and engineers available to work in the private sector.

Each of these goals suggests natural metrics for measuring progress. The NSF currently measures the fraction of 24-year-olds who receive undergraduate degrees in the natural sciences and engineering (NSE). These data are also available for other countries. Although the United States provides undergraduate degrees to a larger fraction of its young people than almost all other developed nations, many fewer undergraduates in the U.S. receive degrees in natural science and engineering. As a result, the fraction of all 24-year-olds with undergraduate NSE degrees is now higher in several nations than it is in the U.S. The United Kingdom (8.5%), South Korea (7.6%), Japan (6.4%), Taiwan (6.4%), and Germany (5.8%) all achieve levels higher than the 5.4%

level attained in the United States (National Science Board 1998: Chapter 2). The experience in the United Kingdom also shows that it is possible to expand this fraction relatively rapidly over time. In 1975, the figure there stood at only 2.9%.

The indicators for the next two goals will have to be more eclectic. Possible indicators of innovation in graduate education could include the creation of graduate training programs in new areas (bioinformatics, for example) where the private sector demand for graduates is high; or programs that involve new types of training (internships in private firms, perhaps); or programs that offer different types of degrees from the traditional master's or Ph.D. One would also like to see continued strength in the Ph.D. programs that form the core of our system of basic scientific research, measured perhaps by the quality of students that they attract both domestically and from abroad. The second and third goals explicitly allow for the possibility that developments in these two areas need not be closely linked. Universities might introduce new programs in an area such as bioinformatics that train people primarily for work in the private sector without affecting existing programs in biology. The new programs could have the same independence from Ph.D. training in biology that programs of chemical engineering have from Ph.D. training in chemistry. As a result, innovation in the sense of new programs need not imply any changes in the existing Ph.D. training programs and need not take any funding from those programs. If the country makes progress toward the first goal, and the number of U.S. citizens who pursue undergraduate studies in science increases, this could improve the quality of the domestic applicant pool for the traditional Ph.D. programs at the same time that it supplies people to the new alternative forms of graduate education.

It will take new funding from the federal government to encourage the introduction of new training programs and still preserve the strength of existing graduate programs. The last goal sets a rough benchmark that policymakers might use to set expectations for how much funding might be allocated on a permanent basis toward these goals. In the last 2 decades, the primary programs that have subsidized the private sector demand for R&D have been the research and experimentation tax credit, the SBIR program, and the ATP program. Rough estimates of the costs for these programs are $1 billion each per year for the tax credit and the SBIR program and between $300 and $400 million per year for the ATP program. The fourth goal suggests a starting target of around $2–2.5 billion per year in subsidies for the supply of scientists and engineers.

If policymakers adopted these kinds of goals, then it would be a straightforward process to design programs that might help achieve them and to evaluate these programs after they are implemented. The following list of programs only begins to suggest the range of possibilities that could be considered. (1) Provide training grants to undergraduate institutions that are designed to increase the fraction of students receiving NSE degrees. (2) Finance the creation of a system of objective, achievement-based (rather than normed) tests that measure undergraduate level mastery of various areas of natural science and engineering. (3) Create and fund a new class of portable fellowships, offered to promising young students, that pay $20,000 per year for 3 years of graduate training in natural science and engineering.

The details for all of these programs would have to be adjusted based on more detailed prior analysis and as experience with any of them is acquired in practice. Many alternative programs could also be proposed. These three are offered here primarily to indicate the wide range of possibilities and to move the debate about government programs out of the rut in which it has been stuck for some 20 years.

Training grants could be very flexible. They could follow the pattern that has already been established for training grants at the graduate level. Formally, grants could still be given to a lead principal investigator, but in effect, they would offer financial support to a department at a university or college. The details of the proposed training program would be left open to the applicants. Like all grants, they would be peer reviewed, with fixed terms but renewable. One of the central criteria in evaluating any proposed grant would be some estimates of its cost-effectiveness as measured by the expenditure per additional undergraduate NSE degree granted. At this point, undergraduate institutions in the United States award about 200,000 NSE degrees each year. The vast majority (roughly 95%) of these degrees are awarded to U.S. citizens. It will take an increase of about 100,000 NSE degrees to U.S. citizens per year to meet the goal of having 8% of 24-year-olds receive an NSE degree. If the federal government devoted $1 billion per year, or about $10,000 per additional degree recipient as a reward to schools that could increase the numbers of NSE degrees that they award, universities would surely find it in their interest to reverse the existing pattern of discouraging students from pursuing NSE degrees. Existing liberal arts universities could reallocate resources internally. Specialized science and engineering schools could use these funds to expand. New institutions could enter the educational marketplace.

One of the obvious risks associated with a goal of increasing the number of NSE degrees is the risk that universities would simply relabel existing degrees as NSE degrees or would substantially reduce the content of the NSE degrees that they award. One additional criterion for evaluating training grants would be the presence of metrics that verify whether the quality of the degree from the recipient institution is being eroded. But eventually, it would be more efficient to have objective, national measures of student mastery of science rather than the kind of implicit, idiosyncratic, institution-specific assurances of the quality that universities now provide. The model for this system of measures would be the advanced placement tests offered to high school students by the College Board. This organization has shown that it is possible to construct reliable tests with the property that when teachers teach to the test, the students actually learn the material that they should learn. Just as the AP system is guided by high school and college educators, one would expect that any such system for measuring undergraduate achievement in science would be guided primarily by the professors who teach science at the undergraduate and graduate level. Presumably, scores on these kinds of achievement-based exams would not replace other indicators like course grades, letters of evaluation, and general measures of intellectual ability such as are provided by the existing graduate record exams. Nevertheless, they would provide a new and useful piece of information about performance by individual students, by different educational institutions, and by the nation's educational system as a whole. Given the pervasive problems of incomplete information in higher education, it would surely be of value to students, employers, and faculty members to have access to objective measures of what students actually learn.

The new fellowship program is intended primarily to encourage the process of innovation in graduate education by providing a ready pool of funds that could be spent on any attractive new programs that are created. It would also create additional incentives for students to pursue undergraduate NSE degrees. Possible details for such a program could be as follows: The government could select a sample of graduating high school students who show promise in science, say the more than 100,000 high school students per year who pass the advanced placement exam in calculus. It could offer to a randomly selected treatment subgroup a fellowship that will pay $20,000 per year for 3 years of *graduate* education in natural science or engineering if the student receives an undergraduate NSE degree. (There would be little reason to pay them a subsidy for undergraduate education. Virtually all of these students already go on to get an undergraduate degree.) Granting the award before they begin their undergraduate study would allow them to take the science courses that prepare them for graduate study. Because the treatment group would be randomly selected, it will be easy to verify whether these grants increase the likelihood that a student receives an undergraduate NSE degree. One could also look among the students who continue their studies in graduate school and see whether the recipients of the portable fellowships select career paths that differ from the students who are supported under the existing RA and TA positions. To the extent that fairness is a concern, one could give some other award to the students in the control group, a new personal computer perhaps.

These fellowships would be portable both in the sense that they could be used to pay for training in any field of natural science and engineering and in the sense that they could be used at any institution that the student selects. Some of the students who receive these fellowships would no doubt pursue a traditional course of Ph.D. study, but some may be willing to experiment with other kinds of degrees. Because these funds would represent new funds, not subtractions from the funds that are already used to support graduate students, and because they would only cover 3 years of training, they should not pose much risk to the traditional training system in basic science.

If the government paid for a total of 50,000 of these fellowships each year, or about 16,700 for each annual cohort of students, this would represent an annual expenditure of about $1 billion. (To pay for 16,700 new fellowships each year, the government would presumably have to offer many more because the take-up rate would be less than 100%.) It is possible that the availability of these funds would not lead to the introduction of new courses of study that cater to the recipients. If this were the outcome, the fellowships would be judged a failure and would presumably be discontinued. But, a priori, it seems quite likely that a flow of funds of this magnitude would induce at least some innovative response from our educational system. It should not take many years of observation to verify whether this conjecture is correct.

IX. Conclusions

The analysis here is driven by two basic observations. The first is that better growth policy could have implications for the quality of life in all dimensions that are so large that they are hard to comprehend. The second is that in the last several decades, the efforts that our nation has undertaken to encourage faster growth have been timid and poorly conceived.

We owe it to our children and their children to address questions about growth policy the way we would approach a major threat to public health. We must use the best available evidence and careful logical analysis to frame new initiatives. We must then be willing to run experiments and to see what actually works and what does not.

References

Autor, David, Lawrence F. Katz, and Alan B. Krueger. 1998, November. "Computing Inequality: Have Computers Changed the Labor Market?" *Quarterly Journal of Economics* 113:1169–213.

Frank, Robert, and Phillip Cook. 1995. *The Winner-Take-All Society*. New York: Free Press.

Goolsbee, Austan. 1998, May. "Does Government R&D Policy Mainly Benefit Scientists and Engineers?" *American Economic Review* 88:298–302.

Griliches, Zvi. 1992. "The Search for R&D Spillovers." *Scandinavian Journal of Economics* 94:347–520.

Hall, Bronwyn, and John van Reenen. 1999. "How Effective Are Fiscal Incentives for R&D? A Review of the Evidence." Working Paper no. W7098, National Bureau of Economic Research, Cambridge, MA.

Historical Statistics of the United States: Colonial Times to 1970, Part 1, U.S. Bureau of the Census, Washington, D.C., 1975.

Jones, Charles I. 1995, May. "Time Series Tests of Endogenous Growth Models." *Quarterly Journal of Economics* 110:495–525.

Jones, Charles I., and John C. Williams. 1998, November. "Measuring the Social Return to R&D." *Quarterly Journal of Economics* 113:1119–35.

Katz, Lawrence F., and Kevin M. Murphy. 1992, February. "Changes in Relative Wages, 1963–1987: Supply and Demand Factors." *Quarterly Journal of Economics* 107:35–78.

Langlois, Richard N. and Mowery, David C. 1996, "The Federal Government Role in the Development of the U.S. Software Industry." In David C. Mowery, ed., *The International Computer Software Industry: A Comparative Study of Industry Evolution and Structure*. New York: Oxford University Press.

Lerner, Josh. 1999, July. "The Government as Venture Capitalist: The Long-Run Impact of the SBIR Program." *Journal of Business* 72:285–318.

Maddison, Angus. 1995. *Monitoring the World Economy, 1820–1992*. Paris: Development Centre of the Organisation for Economic Cooperation and Development.

Morgan, Rick, and Len Ramist. 1998, February. "Advanced Placement Students in College: An Investigation of Course Grades at 21 Colleges." Report no. SR-98-13, Educational Testing Service, Princeton, NJ.

National Research Council. 1998. *Trends in the Early Careers of Life Scientists*. Washington, DC: National Academy Press.

National Science Board. 1998. *Science and Engineering Indicators—1998*. Arlington, VA: National Science Foundation.

Rosenberg, Nathan. 1969. *The American System of Manufactures: The Report of the Committee on the Machinery of the United States 1855, and the Special Reports of George Wallis and Joseph Whitworth 1854*. Edinburgh: Edinburgh University Press.

Statistical Abstract of the United States. U.S. Bureau of the Census, Washington, D.C., various years.

Stern, Scott. 1999, October. "Do Scientists Pay to Be Scientists?" Working Paper no. 7410, National Bureau of Economic Research, Cambridge, MA.

Wallsten, Scott. 2000, forthcoming, Spring. "The Effects of Government-Industry R&D Programs on Private R&D: The Case of the Small Business Innovation Research Program." *The Rand Journal of Economics*.

7 Appendix C

Process, Responsibility, and Myrons Law

From, *In the Wake of the Crisis: Leading Economists Reassess Economic Policy* Edited by Olivier Blanchard, David Romer, Michael Spence, Joseph E. Stiglitz, The MIT Press, 2012.

12
Process, Responsibility, and Myron's Law

Paul Romer

In the wake of the financial crisis, any rethinking of macroeconomics has to include an examination of the rules that govern the financial system. This examination needs to take a broad view that considers the ongoing dynamics of those rules. It will not be enough to come up with a new set of specific rules that seem to work for the moment. We need a system in which the specific rules in force at any point in time evolve to keep up with a rapidly changing world.

A diverse set of examples suggests that there are workable alternatives to the legalistic, process-oriented approach that characterizes the current financial regulatory system in the United States. These alternatives give individuals responsibility for making decisions and hold them accountable. In this sense, the choice is not really between legalistic and principle-based regulation. Instead, it is between process and responsibility.

The Dynamics of Rules

The driving force of economic life is the nonrivalry of ideas. *Nonrivalry* means that each idea has a value proportional to the number of people who use it. Nonrivalry creates a force that pushes for increases in the scale of interaction. We see this force in globalization, which relies on flows of goods to carry embedded ideas to ever more people. We see it in digital communication, which allows the direct sharing of ideas among ever more people. We see it in urbanization, which allows us to share ideas in face-to-face exchanges with ever more people.

A new slant on an old saying expresses the updated essence of nonrivalry of a technological idea: give someone a fish, and you feed them for a day; teach someone to fish, and you destroy another aquatic ecosystem. This update reflects what has happened throughout most of human history and warns that we need more than new ideas about technology to achieve true progress.

We need to broaden our list of ideas to include the rules that govern how humans interact in social groups (rules like those that limit the total catch in a fishery). *Rules* in this sense mean any regularities of human interaction, regardless of how they are established and enforced. Finding good rules is not a one-time event. As academics, policymakers, and students of the world, we need to think about the dynamics of both technologies and rules.

To achieve efficient outcomes, our rules need to evolve as new technologies arrive. They must also evolve in response to the increases in scale that nonrivalry induces. Finally, and perhaps most important, they also need to evolve in response to the opportunistic actions of individuals who try to undermine them. Myron Scholes once captured this last effect in a statement he made in a seminar, a statement that deserves to be immortalized as Myron's law:

Asymptotically, any finite tax code collects zero revenue.

His point was that if there is a fixed set of rules in something like a tax code, clever opportunists will steadily undermine their effectiveness. They will do this, for example, by changing the names of familiar objects to shift them between different legal categories or by winning judicial rulings that narrow the applicability of the existing rules.

In sum, rules have to evolve in response to three distinct factors—new technologies, increases in the scale of social interaction, and opportunistic attempts at evasion. Any social group has higher-level rules—*metarules*—that determine how specific rules evolve. The metarules that govern the tax code, for example, allow for changes through legislation passed by Congress, regulations written by the Internal Revenue Service, and rulings handed down by courts. In some domains, the three forces that call for more rapid change in the rules may operate with greater force. In those domains, we presumably want to rely on different metarules.

Why Rules Lag Behind

As the number of people who use the Internet has increased, the rules that govern behavior have lagged far behind actual practice. This case offers helpful illustrations about the general problem that we face ensuring that rules keep up.

New technologies are part of the problem. Digital communication has created many new possibilities for criminal activity that crosses national borders. Our systems of criminal investigation and prosecution, which are based on geographical notions of jurisdiction, are ill-suited to this new world.

Scale also has an independent effect. Email is based on a set of rules that worked well when dozens of academics were communicating with each other. These informal rules were based on norms and reputation, so the Internet protocol and associated protocols for managing email failed to include even the most basic protections. Now that the Internet has scaled from dozens of people to billions, different rules are needed. For example, there is no built-in way for the recipient of an email to be sure about the identity of the sender. In a "spear-phishing" attack, an email is carefully tailored to resemble the authentic emails that the recipient normally receives. Because none of the usual warning signs are present (there are no offers of millions of dollars stranded in a stranger's bank account), the recipient is more likely to open an attachment with malicious code. Even RSA, a company whose business revolves around computer security, was compromised through this kind of attack.

Engineers at the Internet Engineering Task Force, a loosely defined voluntary organization with little formal authority, are the rule setters for the Internet. In 1992, they began to work on improving security protocols. They devised a patch called IPsec that reverse-engineered some basic security measures into the existing protocol. They also developed an update to the basic Internet protocol, known as IPv6, that has built-in support for IPsec. The basic specifications for these protocols were completed in 1998. Unfortunately, larger scale not only creates the need for better security but also makes it much harder to implement a change in the rules. The adoption of both sets of protocols has been held back by coordination problems among large numbers of users and vendors.

Even if these protocols are widely adopted, new attacks will still emerge. Bigger scale means that traditional mechanisms like reputation no longer operate and that more people are working to undermine and subvert all the existing security measures. Because a new vulnerability is a nonrival good that can be shared among predators, an increase in scale can increase the rate at which predators circumvent any given security system.

Financial Markets

Rules in financial markets need to evolve for all of the reasons identified above. Technology is creating entirely new opportunities—for example, in high-frequency electronic trading systems. The scale of financial markets continues to grow, and private actors in these markets will surely seek clever ways to evade the intent of existing rules. The gains from opportunism in these markets are so large that the total amount of human effort directed at evading the rules will presumably be at least as large as that devoted to a low-return activity like cybercrime.

Electronic transactions were supposed to offer liquid markets and unified prices that can be accessed by everyone, but they have not lived up to this promise because they have also created new opportunities for manipulation. For example, some firms now submit and withdraw very large numbers of electronic quotes within milliseconds in a practice known as *quote stuffing*. It is not clear what the intent of these traders is, but it is clear that any electronic trading system will have capacity constraints in computation and communication. Any system will therefore be subject to congestion. In the May 2010 stock market flash crash, congestion added to the anomalous behavior that firms were observing, and this apparently encouraged many high-frequency traders to stop trading, at least temporarily. This seems to have contributed to the temporary sharp fall in prices.

Quote stuffing could be one of many different strategies that traders use to influence local congestion and delays in the flow of information through the trading system. These, in turn, could affect liquidity, as they did during the flash crash. As a result, transactions could take place at prices that depart substantially from those that prevailed just before or just after they occurred.

After an extensive analysis, the Securities and Exchange Commission (SEC) reported that quote stuffing was not the source of the cascade of transactions that overwhelmed the systems during the flash crash. The SEC is still equivocating about whether this particular practice is harmful and, more generally, about systemic problems that high-frequency traders may be causing. Even if it had tried to address the specific practice of quote stuffing, the type of rule that had first been mooted—forcing traders to wait 50 milliseconds before withdrawing a quote that they had just submitted—would probably have been too narrow to limit the many other strategies that could be used to generate congestion or influence liquidity.

It seems implausible that the kind of behavior that occurred in the flash crash is an inevitable consequence of electronic trading. (But if it is, it seems implausible that the switch to electronic trading has brought net welfare benefits for the economy as a whole.) One year later, it also seems implausible that any of the changes implemented so far has fully addressed the underlying issue. Individual stocks continue to suffer from instances where trades take place at prices that are dramatically different from those that are prevailing seconds before or seconds later.

After the flash crash, trades were canceled if they took place at prices that differed from a reference price by more than a discretionary threshold, set in that particular case as a 60 percent deviation. Under new rules that try to be more explicit, transactions for some individual stocks will be allowed to stand if they take place at prices within 10 percent of the a reference price. In a multistock event, where many prices move together, the band of acceptability widens to 30 percent. Some have criticized these new rules because they still allow some discretion in setting the reference price. Others have expressed concern about the potential for manipulation that could intentionally trigger the looser rules that apply in a multistock event.

As the discussion below about rule making at the Occupational Safety and Health Administration (OSHA) shows, even in a simple setting it is difficult to develop rules in a timely fashion that meet legal standards for clarity and do so following procedures that meet legal standards for due process. The Security and Exchange Commission's attempts to clarify the rules for breaking trades suggest that it is much harder to live up to these standards in a complicated and dynamic context. The SEC seems to have settled for a rule-setting process that leaves ample room for opportunism for extended periods of time. Perhaps some other, less legalistic approach deserves consideration.

Process versus Responsibility in Other Domains

One way to think about how the metarules that govern financial regulation might be adjusted so that the system can respond more quickly is to examine a broad range of social domains and observe the outcomes under alternative metarules. Here are four influential organizations in the United States that set rules and a specific goal that each organization's rules try to promote:

- Federal Aviation Administration (FAA): flight safety
- Federal Reserve: stable economic activity
- U.S. Army: combat readiness
- Occupational Safety and Health Administration (OSHA): worker safety

The Federal Aviation Administration works in a domain with the potential for rapid technological evolution. It has responsibility for passenger airplanes, which are among the most complex products ever developed. It approaches its task of ensuring flight safety with rules that specify required outcomes but that are not overly precise about the methods by which these outcomes are to be achieved. This is one way to interpret what principle-based regulation should look like. In practice, this means that some person must have responsibility for interpreting how any specific act, in a specific situation, either promotes or detracts from the goal that is implicit in the principle. That is, someone has to take responsibility for making a decision.

The general requirement that the FAA places on a new plane is that the manufacturer demonstrate to the satisfaction of its examiners that the new airplane is airworthy. The examiners use their judgment to decide what this means for a new type of plane. Within the FAA, the examiners are held responsible for their decisions. This changes the burden of proof from the regulators of a new technology to the advocates of the technology and gives FAA examiners a large measure of flexibility.

This approach stands in sharp contrast to one based on process. There is no codified process that a manufacturer can follow and be guaranteed that a new plane will be declared airworthy. Nor is there a codified process that the FAA examiners can follow in making a determination about airworthiness. There is no way for them to hide behind a defense that they "checked all the boxes" in the required process.

One obvious requirement for a plane to be airworthy is that the airframe be sufficiently strong. There are no detailed regulations that specify the precise steps that a manufacturer must use to make a plane strong or show that it is strong. For example, there are no regulations about the size or composition of the rivets that hold the skin on the airframe, nor should there be. On an airplane like the Boeing 787, which is made of composite materials, there are no rivets. Instead, as part of the general process of establishing airworthiness, the employees of the FAA have technical expertise in areas like materials science and testing procedures and are responsible for making a judgment about how to test a particular design and determine whether it is sufficiently strong.

Moreover, because new information about an airframe can emerge for decades after it enters into service, the granting of a certificate of airworthiness is always provisional. Operators of aircraft are required to report evidence that emerges over time that might be relevant to airworthiness. At any time, the FAA can withdraw a plane's airworthiness certificate or mandate changes that must be made to an aircraft for it to continue to be airworthy. No judicial proceeding is required. There is no appeal process for an owner that unexpectedly receives an airworthiness directive that mandates an expensive modification. There is no way to get a judge to issue an injunction that would let the plane keep flying because the FAA has not satisfied some procedural requirement.

It is also clear that the rate of innovation in technologies is a choice variable, along with the rate of innovation in the rules. If social returns are maximized when technologies and rules stay roughly in sync, good metarules might require that those who develop new technologies also have to develop the complementary rules before the new technologies can be implemented. A larger plane such as the Airbus 380 will generate more air turbulence in its wake. This means that the FAA has to implement new rules about the spacing between planes that follow each other on a flight path. The FAA will not let a plane like the Airbus 380 fly until the manufacturer has demonstrated the size of its wake and the FAA has had time to put in place new systemwide rules about separation. This is the polar opposite of the approach that the SEC takes with regard to the introduction of major changes in the architecture of the electronic trading system.

The FAA implements a system based on individual responsibility by organizing itself as a hierarchy. People at a higher level can promote and sanction people at lower levels based on how well they do their jobs. At the top of the hierarchy, the secretary of transportation and the administrator of the FAA are appointed by the president and confirmed by Congress, both of which are held accountable by the electorate.

The Federal Reserve, like every other central bank, is also organized as a hierarchy. Its leaders are held accountable by democratically elected officials who specify a mandate. In their day-to-day decisions, the employees at lower levels in the hierarchy have a lot of freedom to take actions that will achieve the organization's mandate. They are rewarded or punished based on the judgment of those one level higher in the hierarchy. There is little scope for the legislature to micromanage decisions, and there is no judicial review of the process by which decisions are made. As was seen in the financial crisis of 2009, this kind of system allowed for a much quicker response than the parallel mechanism involving legislation passed by Congress. The Fed's response to the failure of Long Term Capital Management also showed that it could manage what amounted to a bankruptcy reorganization far more quickly than a court could.

Like the Fed and the FAA, the U.S. Army is run as a hierarchy, with accountability at the top to elected officials. After a period during the 1970s when racial tensions in the army were seriously undermining its effectiveness, the leaders of the army decided that better race relations were essential for it to meet its basic goal of combat readiness. In less than two decades, they remade the organization. Writing in 1996, the sociologists Charles Moskos and John Butler observed that among large organizations in the United States, the army was "unmatched in its level of racial integration" and "unmatched in its broad record of black achievement" (2). To illustrate how different the army was from more familiar institutions such as the universities where they worked, Moskos and Butler (1996, 3) tell this story:

> Consciousness of race in a nonracist organization is one of the defining qualities of Army life. The success of race relations and black achievement in the Army revolves around this paradox. A story several black soldiers told us at Fort Hood,

Texas, may help illustrate this point. It seems that one table in the dining facility had become, in an exception to the rule, monopolized by black soldiers. In time, a white sergeant came over and told the blacks to sit at other tables with whites. The black soldiers resented the sergeant's rebuke. When queried, the black soldiers were quite firm that a white soldier could have joined the table had one wished to. Why, the black soldiers wondered, should they have to take the initiative in integrating the dining tables?

The story has another remarkable point—that a white sergeant should take it on himself to approach a table of blacks with that kind of instruction. The white sergeant's intention, however naive or misdirected, was to end a situation of racial self-segregation. Suppose that a white professor asked black students at an all-black table in a college dining hall to sit at other tables with whites. This question shows the contrast between race relations on college campuses and in the army.

The system in the army makes such individuals as the sergeant in this story responsible for the state of race relations in any unit they supervise. This system holds them responsible for both their decisions and accomplishments, through occasional ad hoc review of their decisions by superior officers and through more formal decisions about promotion to a higher rank. Any particular decision like that of the sergeant in the story could easily be second-guessed, but the system as a whole has clearly been effective at achieving both integration and good race relations. Both direct judicial intervention in the operation of public school systems and the combination of legislation and regulations that guide behavior on university campuses have been far less successful.

The approaches to safety at the FAA, to macroeconomic stabilization at the Fed, and to race relations in the army all stand in sharp contrast to the legalistic, process-centered approach to safety followed by OSHA. To improve safety on construction sites, which have a bad safety record, OSHA follows a detailed process that leads to the publication of specific regulations such as these:

1926. 1052(c)(3)
The height of stair rails shall be as follows:
1926. 1052(c)(3)(i)
Stair rails installed after March 15, 1991, shall be not less than 36 inches (91.5 cm) from the upper surface of the stair rail system to the surface of the tread, in line with the face of the riser at the forward edge of the tread.
1926. 1052(c)(3)(ii)
Stairrails installed before March 15, 1991, shall be not less than 30 inches (76 cm) nor more than 34 inches (86 cm) from the upper surface of the stair rail

system to the surface of the tread, in line with the face of the riser at the forward edge of the tread.

These regulations are enforced by OSHA inspectors, who can issue citations that lead to fines and that can then be challenged in court. The regulations are supplemented by guidance about enforcement. For example, in the early 1990s, someone also added a note in the *Construction Standard Alleged Violations Elements (SAVE) Manual* that guided OSHA inspectors on how to apply these regulations on stair rails:

NOTE: Although 29 CFR 1926.1052(c)(3)(ii) sets height limits of 30″–34″ for stairways installed before March 15, 1991, no citation should be issued for such rails if they are 36″ maximum with reference to 29 CFR 1926.1052(c)(3)(i).

This change in enforcement patterns avoids the awkward situation in which a 35-inch-high rail could be cited either for being too low or for being too high, depending on when it was installed, although it still leaves a puzzle about why a 38-inch-high rail might still be cited if it had been installed too early.

It is tempting to ridicule regulations like these, but it is more informative to adopt the default assumption that the people who wrote them are as smart and dedicated as the people who work at the FAA. From this, it follows that differences in what the two types of government employees actually do must be traced back to structural differences in the metarules that specify how their rules are established and enforced. The employees at the FAA have responsibility for flight safety. They do not have to adhere to our usual notions of legalistic process and are not subject to judicial review. In contrast, employees of OSHA have to follow a precise process specified by law to establish or enforce a regulation. The judicial checks built into the process mean that employees at OSHA do not have any real responsibility for worker safety. All they can do is follow the process.

One possible interpretation of the regulations about stair rails is that the regulations once specified a maximum height of 34 inches and that new evidence emerged showing that a higher rail would be safer. As they considered new rules they could propose, the regulation writers faced the question of what rules to apply to stairways that had been installed in the past. Rather than make an ex post change to the regulations for existing stairs, they may have chosen instead to stick to the principle that the regulations that were in force when a stairway was installed would continue to apply to that stairway but to suspend enforcement for some violations.

The caution about ex post changes in the regulations may derive in part from a concern about judicial review of the new rules. Or it could have come from a concern about judicial review of penalties that had already been assessed or violations under the old rules that would no longer be violations under the new rules. The change in enforcement at least made sure that no judge saw cases where 35-inch-high rails were sometimes cited for being too high and sometimes for being too low.

You can get some sense of how difficult it is to be precise in writing rules by digging into an area like this. From published inquiries that OSHA received, it seems that the decisions here were complicated by ambiguity about the rules for handrails, which a person uses as a grip and should therefore not be too high, and stair rails, which mark the top of a barrier designed to prevent falls and which therefore should not be too low. The top of a stair rail might be but need not be a handrail. It looks as though the rules morphed over time to distinguish more explicitly between the two types of rails.

It is striking that safety officers for construction firms who wrote to OSHA for clarifications about apparent discrepancies between different sections of the regulations waited four to six months to receive answers. (One wonders what happened at the construction sites during those many months.)

Even more striking is the fact that the rules cited here were first proposed in 1990 or 1991, but judging from a 2005 notice in the *Federal Register* calling once again for comments, they did not come into force until sometime after 2005. (The notice in 2005 makes a brief reference to other agency priorities that took precedence over the rules for stair rails.) This required the application of a further enforcement instruction that a stair rail that conforms to the proposed regulation for stairs built after 1991 but that violates the existing regulations (which were not changed for another fifteen years) would be treated as a de minimus violation and would not result in an enforcement action.

The principle-based approach to the regulation of air safety lacks all of the procedural and legal protections afforded by the process of OSHA, but in terms of the desired outcomes, the FAA approach seems to work better. Air travel is much safer than working on a construction site. The

Fed and the army also seem to have been much more effective in addressing complicated challenges. Despite the more extensive judicial protections afforded the construction firms under the OSHA process, firms find the process infuriating. Construction sites are still very dangerous places to work.

Conclusion

People from the United States take pride in a shared belief that theirs is "a nation of laws, not of individual men and women." Taken literally, this claim is nonsense. Any process that decides what kind of planes can carry passengers, what to do during a financial panic, how people of different races interact, or how a construction site is organized will have to rely on decisions by men and women.

Because of combinatorial explosion, the world presents us with a nearly infinite set of possible circumstances. No language with a finite vocabulary can categorize all these different circumstances. No process that writes rules in such a language can cover all these circumstances. Laws and regulations always require interpretation. Giving judges a role to play in making these interpretations or reviewing them does not take people out of the process.

We could have a system in which individual financial regulators have the same kind of responsibility and authority as the sergeant in the cafeteria. If they saw behavior that looked harmful to the system, they could unilaterally stop it. We could have a system like the one we use to certify passenger aircraft, in which the burden of demonstrating that an innovation does not threaten the safety of the entire trading system rests on those who propose the innovation. In such a system, the people that the innovators would have to persuade could be specialists who would have the same kind of responsibility and authority as FAA examiners. The opportunists in the financial sector would presumably prefer to stay with an approach that emphasizes process, but this leaves the other participants in the sector at a relative disadvantage. More seriously, it leaves those outside the sector unprotected, with no one who takes responsibility for limiting the harms that the sector can cause.

The right question to ask is not whether people are involved in enforcing a system of rules but rather which people are involved and which incentives they operate under. There may be some contexts where a legalistic approach like that followed by OHSA and the SEC has advantages, but we need to recognize that this approach is not the only alternative and that it has obvious disadvantages.

A careful weighing of the costs and benefits will involve many factors, but the factor that seems particularly important for the financial sector concerns time constants. As the OSHA example suggests, the legalistic process is inherently much slower than a process that gives individuals more responsibility. Moreover, clever opportunists can dramatically increase the delays and turn the legalistic approach into what Phillip Howard (2010) calls a "perpetual process machine."

Under this approach, rules for the financial sector will never keep up. The technology is evolving too quickly. The scale of the markets is enormous and continues to grow. There may be no other setting in which opportunism can be so lucrative. It is hard to understand why technologically sophisticated people devote any effort to committing cybercrime when the payoffs from opportunism in financial markets seem to be so much larger. If we persist with the assumption that a legalistic rule-setting process is the only conceivable one we could use to regulate financial markets, then the opportunists will thrive. We will settle into a fatalistic acceptance of systemic financial crises, flash crashes, and ever more exotic forms of opportunism.

"No one can predict how complicated software systems will behave" (except in airplanes). "You can't change behavior" (except in the army). "Financial systems are just too complicated to regulate" (except in countries like Canada, where instead of running a process, regulators take responsibility).

References

Howard, Philip K. 2010. "To Cut the Deficit, Get Rid of Surplus Laws." *Washington Post*, December 10.

Moskos, Charles C., and John Sibley Butler. 1996. *All That We Can Be: Black Leadership and Racial Integration the Army Way*. New York: Basic Books.