1995

AAAS Science and Technology Policy Yearbook

Albert H. Teich
Stephen D. Nelson
Celia McEnaney
editors

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

1 An Unpredictability Principle for Basic Research

Ralph E. Gomory

Outcomes from Science

Basic research is funded because of the belief that something broadly and directly useful will eventually come out of the scientific effort. Those who do research will not often think about usefulness, nor should they. They are motivated by the basic desire to understand and the hope for recognition by their peers for work that is good science, quite independent of its applicability. But scientific funding ultimately rests on society's hope and expectation of practical results.

This expectation has been amply fulfilled. Indeed, I am puzzled by the often expressed need to justify science now that the Cold War is over. It is absolutely clear that basic science has provided major practical returns. One example is the transistor, an invention that grew directly out of the most basic understanding of solid-state physics. Another is molecular biology, with its remarkable revelations about the basic functioning of all living things and its enormous and emerging consequences as a technology.

This *overall* success of basic research, however, has not prevented officials (this seems to vary with Administrations) from wanting to fund only those areas within science that can be seen as useful

---

Ralph E. Gomory is the president of the Alfred P. Sloan Foundation. *This article is based on the William D. Carey Award Lecture delivered at the 20th Annual AAAS Colloquium on Science and Technology Policy, held April 12-14, 1995, in Washington, DC.*

and questioning the support of areas that do not seem to be useful. Even for those who accept the idea that basic research should be supported without any reference to usefulness, that general acceptance does not answer the question of how much science to support. How much science is enough?

Lack of Rationale

The scientific community is at the mercy of external views, and these views are changeable. This is in part because there is no generally agreed on rationale for the support of science within the scientific community itself. There is always, of course, the assertion that basic science should be funded, preferably on a larger scale than it is now, and agreement that the budgets of National Science Foundation (NSF) and National Institutes of Health (NIH) should increase. But that level of consensus within the community does not distinguish scientists from any other special interest group. What is lacking is an accepted rationale showing that support for science is in the national interest.

I do not claim that if we had such a rationale it would entirely solve the scientific funding problem. We would still be affected by the attitudes of committee chairmen, and all the other elements of personalities and politics that are the staples of so much discussion in Washington. But a coherent rationale would help, and I want to describe one that rests on some observations that all scientists understand, but do not always articulate clearly. It involves the concepts of predictability, unpredictability, and dependence on detail.

Quantum Mechanics

But before discussing such a rationale, I would like to recall a bit of history about quantum mechanics. In the 1920s there was no scientific subject more pure and esoteric than the brand new field of quantum mechanics. It involved the uncertainty principle, and the baffling puzzle of electrons that behaved like waves one moment and particles the next. It was a subject of exciting scientific and even philosophical importance, but nothing could have seemed further from practical applications. By the 1930s, however, quantum mechanics began to influence solid-state physics, and after World War II, an improved understanding of the fundamentals of crystal and solids led to a better grasp of the role of trace impurities and their effect on the flow of electrons. The invention of the transistor soon followed, with all its tremendous significance for computers and electronic devices of every sort, and through these, for the lives of all of us. Not much more than 30 years separated the emergence of an esoteric and apparently useless field of scientific study from its enormous everyday impact.

Why We Find Things (on the average)

This example (and there are many others) shows that practical discoveries do turn up in the course of pursuing the most basic knowledge. What is more, we can expect this to keep happening. Sometimes people outside science think that this is pure serendipity and that what this means is that if you turn over enough rocks at random, you will occasionally find a diamond. But the actual research process is nothing like that; it is the systematic exploration of a significant piece of the natural universe. It is *not* surprising that when you start to understand important pieces of the universe in a fundamental way—for example, how solids hang together, or how living beings function at the molecular level—at *some* point this understanding will enable you to do things that you could not do before.

What We Don't Know

But even though basic research itself is reasonably systematic and its improvement in our understanding is reasonably predictable, the when and how of its practical impact is *not* predictable. Even the most expert scientist does not and cannot know how and when or even if the practical impact will occur. Scientists in the 1920s, no matter how talented, could not have foreseen the timing or the nature of the remarkable impact on society that came from quantum mechanics, via transistors and computers. That is simply not doable.

The application aspect of basic research is a process that in many ways resembles the spin of a roulette wheel. It is inherently not predictable in detail. You cannot tell when the useful result will occur, but, like results from the spins of a roulette wheel, the average is predictable, and history shows us that the average results from basic research are tremendously rewarding.

Recognizing the Practical

However, when the roulette wheel spins, and something of immense practical importance does occur, we can recognize it. This is what happened when solid-state physics gave birth to transistors, and today there is the immense and obvious practical impact of molecular biology. When this kind of result happens, the field has moved to a different category. It has established its practical importance.

The Unpredictability Principle

We can describe both what we can recognize and what we cannot predict in this way.

We *can* see when some area of science is useful, but we *cannot* see that some area of science won't be useful.

We can see when some area of science is useful when the field has shown it can produce something significant outside of itself. On the other hand, we can never know that some area of science won't be useful. This has to do with the unpredictability of useful results, of which quantum mechanics is an example. I think most scientists know this at some level, but do not articulate it with precision. If you accept this proposition, certain deductions follow.

The first conclusion is that if the United States wants to have the advantage of a scientific field when it starts to have a practical impact, then, since the practical impact is unpredictable:

The United States should be *among the world leaders*, at least, in *all* major areas of science.

If we do this, without regard to whether these fields have a track record of practicality or not, we will be there when the roulette wheel spins and we can participate in a possible fortunate outcome. We will be in a position to profit as we did from condensed matter physics and molecular biology.

Some argue that scientific information is open to every nation, so why bother with a national position in science at all such as the one we have described. But science is not a free good. You cannot benefit from science unless you have paid the price of entry any more than you can receive radio waves without a radio. Without training, your scientists will not be able to read the papers in their fields.

However, if the roulette wheel of science does spin and significant outcomes occur in the practical sense, we might decide that we want more than to be just one of the leaders.

When a field has become productive in that way, we may determine that its usefulness is such that we wish to lead the world in that area.

Thus, the second conclusion is:

The United States should maintain *clear world leadership* in *some* selected areas of science.

These two conclusions are, in fact, two of the main conclusions of the 1993 National Research Council (NRC) Committee on Science, Engineering and Public Policy (COSEPUP) report entitled "Science Technology and the Federal Government, National Goals For a New Era," of which I was one of the authors. I do not rule out the possibility of deciding that we should be clear world leaders in all important fields. That is an issue of affordability. But it is important to understand two points.

First, our first conclusion sets a minimum. It asserts that in all significant scientific fields we should be among the world's leaders. And, as we will see below, this is more than rhetoric—this policy can be translated into real budget terms.

Second, I would suggest that the selection of fields for clear world leadership should generally be a social, not a scientific, judgment. It involves a decision that money spent in these areas will yield a large societal return. These benefits may be in industry leadership, or a contribution to national health or a better environment—whatever it is that our society values.

The SSC and Molecular Biology

Two examples will illustrate this way of thinking. They will show that if we take these leadership goals seriously, they have many consequences.

The first example relates to the Superconducting Super Collider (SSC). Is particle physics a field where we must be out ahead of the rest of the world as clear leaders? If so, we should build our own SSC. But if we are content just to be among the leaders, we should try to work with the Europeans in a cooperative arrangement to advance the field. Do we need clear leadership for societal reasons? This is not a judgment to be made by scientists, although it does need scientific input. The simplest test is to ask if there could be a large demonstrated payoff to the economy, to medicine, or to any other area. I think that the record of particle physics simply does not support such a prospect now. We are just spinning the roulette wheel here, as in any other field which has not yet yielded practical results. So I conclude that this is not a field that requires clear leadership, that the SSC should not have been built, and that we ought to work out an arrangement with other countries that will allow us to move forward together in research on particle physics.

If we now turn to the second example of molecular biology, with its application to health and its clear relation to an emerging industry, we would, I think, come to the opposite conclusion. This country might well decide that in the interests of national health and of the emerging biotechnology industry, this is an area in which we wish to be well ahead of the world.

Comparisons and Other Consequences

The goal of being among the leaders in a scientific field is a measurable goal. It involves a comparison of the level of science in the United States, in a particular field, with the level of that field in other countries. While numbers of researchers and other such counts are relevant to such a comparison, they should be ancillary to informed scientific judgment. We are among the leaders if we are roughly on a par with the work done abroad. There are lots of questions to be answered as to how we compare with other countries, or with Europe as an entity, and so on. But these questions need to be worked out in accord with the basic rationale of whether we are in a position to react and participate if the field suddenly changes in productivity.

Note that when we compare ourselves with other countries, we should do it within a field. Testing whether we are among the leaders in a given field of physics—for example, in condensed matter physics—does not call for a comparison of funding for condensed matter physics with funding for particle physics or for a field within chemistry. *It does not involve arguments about whether one field is more exciting than another.* It means that we should measure ourselves against the world standards in each field.

These criteria do not entail a comparison of big science with little science. Our goal is to be among the leaders (or the clear leader) in a field, and this should determine the proper mix of big science and individual investigator science *in that field.* The mix that is right for particle physics is not appropriate for condensed matter physics. What matters is to ascertain the proper mix for each field, and not to add up the big science and individual investigator science across the board for all fields and then make a meaningless comparison of those totals.

This approach does provide an answer to the question of how much science is enough. It does not do this in terms of increases or decreases from whatever today's budget happens to be, but in terms of supporting science at a level that provides the desired level of leadership and thus contributes to the strength of the nation. If we were to adopt such an approach, I strongly believe that it would be an affordable and stable basis for the funding of basic research.

Benefitting From Basic Research Leadership

Now let us examine what we would have to do as a country to benefit from the leading position in basic research.

Flow of People

Certainly, one of the most important mechanisms for benefitting from research leadership is through the flow of people. University people played a major role in the early days of Silicon Valley and Route 128. Biotechnology companies today are full of university people who play decisive roles.

It is important, therefore, especially in those scientific areas which have shown their practicality, that we keep a steady flow of graduates from our universities into U.S. industry. Whether these individuals are U.S. citizens or foreign nationals is not the issue; what matters is the infusion of their ideas into U.S. industry. This country clearly benefits from this flow from our universities. The history of some of our industries would have been very different if Silicon Valley had been located somewhere in the British or German countryside, or just outside Tokyo.

Government Support

In addition to ensuring the continued flow of graduates from our universities, there is the possibility of government support of advanced work. Let me briefly review some of the arguments that are commonly brought up in connection with government support of R&D. The point I will try to make is that their validity or lack of validity is quite dependent on detail.

One theme that runs through this article is detail. Let's look at things in more detail. Don't focus on a total for science funding; instead, take science field by field, to see what we need to be leaders in each field. That is measurable, and then we can add it up and determine an overall science budget. Similarly, in the area of benefitting from scientific leadership, detailed knowledge and a more detailed analysis make a difference.

Externalities

I am going to review some of the sweeping notions. One of these is the externalities argument, which is that the total benefit to society of an investment in R&D cannot be captured within a firm. Others outside the firm will learn about it, and they can copy it and benefit from it. Since the firm cannot capture all the benefit its research gives to society, it underinvests in R&D, relative to the total societal return. Thus, the government should provide an incentive for R&D and somehow make up the difference. This argument is often brought up in connection with an R&D tax credit, but it also appears in other contexts.

This very high level argument needs to be examined in more detail. In its simple form, the externalities argument applies to almost anything that can be copied. New manufacturing techniques become known and are reproduced to spread through a wider world. Even marketing techniques—such as selling computers through an 800 number—once started by one firm are copied quite rapidly by others, and the benefit of that convenience becomes widespread. Clearly, in such cases, the innovating firm captures only a small part of the overall returns. So, why not have a manufacturing tax credit, or, even more unappealing, a marketing tax credit?

If we get down to a finer level, we can make some distinctions between these cases. In the marketing techniques case, the investment to set up an 800 number and to advertise it is not large, and if it works, there is an immediate return. Companies are likely to do this if they think of it, and they will not be deterred by the fact that society will benefit disproportionately when others copy them. In the case of new manufacturing techniques, action depends on how quick the return on the investment is, and sometimes for large investments this case would be hard to distinguish from the case for R&D.

The R&D case itself can be split into investments on continuous improvements and attempts to do something really new and different. Continuous improvements are easier. so a company is more likely to do them on its own. But often this work will not meet the externalities test. Much R&D done by a company is only applicable to that company's product, and no one else can use it. So, since lots of R&D will not meet the externalities test, why have an overall R&D tax credit?

Another important element is the structure of the company itself. In a small company, the R&D tax credit may make it possible to do something that otherwise would not be done. But in many large companies, it has no real effect: Decisions are made as usual throughout the company, then at the end of the year the tax department calculates the R&D tax credit. This suggests that fostering R&D is not helped by such a broad, sweeping approach of tax credits.

Attempts to do something really different in R&D tend to require long periods of investment. The more radical the departure from the established product, the longer the period of gestation and the more uncertain the result. Here we are closer to what emerges from university research, and here I think the externalities argument applies better. Companies find this kind of R&D harder to do, because the first effect is a loss, and the gain is more distant and uncertain. Whether or not the sweeping notion of externalities should apply in particular cases is actually rather detail dependent, and we have not reached anywhere near the bottom level of the details that matter.

How Much R&D

R&D as Percentage of Total Output (GDP)

Another widely used grand argument has to do with R&D as a percent of total national output, or Gross Domestic Product (GDP). Many Administrations have talked about the U.S. *underinvestment in R&D* and its effect on U.S. productivity. It is pointed out that Germany spends 2.5 percent of its GDP on R&D, Japan spends 3 percent, and the United States only spends 1.9 percent. Statements of this sort are often very persuasive. We hear much about the short-range mentality of U.S. industry, so connecting this to cutbacks in R&D seems both plausible and alarming.

However, we are dealing here with a difficult subject, and keeping in mind the importance of details, we should approach such a sweeping statement with caution. In this case, being cautious is rewarded, as I will explain.

Almost all R&D, about 93 percent, is done in the manufacturing sector. As a proportion of GDP, the manufacturing sectors of Japan and Germany (30.8 percent and 30.6 percent, respectively) are bigger than the manufacturing sector of the United States (19.0 percent). If all manufacturing firms in these three countries were *equally* R&D intensive, R&D as a percent of the GDP would come out in the same proportions as the size of the manufacturing sectors. So the difference in R&D spending would not mean that foreign firms are more R&D intensive, only that the manufacturing sector is larger in Japan and Germany.

If this factor is taken into account, it immediately explains the numbers, which are completely consistent with an interpretation that the firms in all three countries are on the average equally R&D intensive. Thus, the apparent national underspending on R&D by the United States relative to Japan and Germany is a reflection of the size of the manufacturing sectors of those countries rather than actual underspending on the part of firms. For the United States to have the same R&D percentage as Japan and Germany, U.S. manufacturing firms would have to spend on the average 1.5 times what their foreign competitors spend on R&D.

In both the examples that I have given, the externalities argument and R&D as a percentage of GDP, the sweeping conclusions do not really work. What is needed is more detailed knowledge of what is really happening. But this is not easily acquired.

Too Big to See

There is, in fact, an inherent difficulty in understanding what the R&D system really does. What is this difficulty? We are used to the idea that there are things that are too small to see with the naked eye; we look at them with the aid of microscopes. However, we are much less used to the idea that there are things that are *too big* to see. But there are, and a national economy or the R&D system are both examples. We cannot see the functioning of the R&D system. That functioning depends on the efficient or inefficient actions of hundreds of thousands of people scattered through thousands of plants, labs, and offices.

Statistics and the Macroscope

It would be wonderful if we had a *macroscope* to observe, in real time and in glorious detail, how this part of our huge economy functions. Then we could see what functions right and what functions wrong. But we don't have a macroscope. *Statistics* is, in fact, our attempt at a macroscope, and it only functions erratically, because if we have the right overall picture, then the statistics can size it right for us and tell us more about it, but if we don't have the picture right, the statistics won't tell us that we don't. The earlier example of R&D as a percentage of GDP illustrates this problem. If you have a picture of R&D being done across the economy, statistics tells you one thing, but if you have another picture which gets down to sectors and their R&D tendencies, it tells you quite another thing. Unfortunately, we don't always have the right picture of these large complex things.

Why People Believe

With an incorrect picture, people will tend to believe either what they want to believe or what makes rough sense to them. If you believe that companies are shortsighted and selfish, you will see any report of a shortage of R&D as just another manifestation of these inherent tendencies, confirming that there is no alternative to getting government deeply involved in making up for and controlling those shortcomings. Or if you believe that free markets are universally applicable, then you don't want any government involvement, because you are convinced that the R&D issues will take care of themselves. Although the complex realities do not match *either* of these homey beliefs, the homey beliefs die hard.

What Do We Do?

If we do acknowledge the complexities of dealing with R&D and of benefitting from basic research, we must answer the question, What is the proper role of government in this area? Ultimately, the *goal* of government should be to ensure that U.S. industry is not disadvantaged, and perhaps is sometimes advantaged, in the area of advanced technology. Other goals are to make sure that relevant research benefits U.S. health care, aids in protecting the environment, and helps us to educate our people.

When do we leave things to the free market, and how do we find out whether the free market works? When, if ever, should we use R&D tax credits? In what areas, if any, do we create a civilian DARPA-ARPA? Should we support initiatives like the clean car initiative? Do we encourage research and development consortia? The realistic answer to these types of questions is that today we really know very little about what works and what doesn't—where there is a realistic as opposed to an ideologically determined role for the government. We are also unclear about what to do when the market does not work.

Basic research is an area where there is considerable agreement that the free market is deficient. Here the government has learned over time to play a constructive role. But outside of this area, there is very little history in modern times of the U.S. government and industry working together to further nonmilitary technical goals. We do not have in government a large cadre of experienced people who know how to work effectively with industry in the national interest, or even how to move new technology into education. Nor do we have realistic criteria for when such actions are needed. Therefore, we need to learn first, and we also need to experiment.

Learning

Learning means, for example, that we should determine field by field whether there is a flow of people and ideas from our basic research establishments into U.S. industry. Is this happening now? When is it happening? Do foreign graduate students still stay in the United States or do they head home to areas where there are now new opportunities? Are there good ties between academic advanced research areas and U.S. companies? Is there a flow of new ideas into our hospitals or our school systems? We will find some places where this is happening and some places where it is not happening.

Experimenting

When we see an area that is not working, we should experiment. There could be industries that are too scattered to do effective R&D either on their own or in connection with universities. There could be important health items for which the market is too small, or areas like high-temperature superconductors in which the materials work needed to develop products is too long-range for potential user-companies to do it. We should experiment in such areas to see what works and what doesn't.

The Results of Knowledge

If we follow this path, we will develop a base of knowledge and experience on the outcomes of certain government actions. We can then have a debate, not about process or ideologies, but about whether the results of a given set of actions are worth the money and the effort. People and parties with different social views will weigh these outcomes differently. But nothing would be more appropriate than to have a political debate about the relative importance of different outcomes, rather than a debate about methods.

Summary

There is a strong case in practical terms for the support of basic research. We can determine on a leadership role for the United States in this area—and by being serious about what leadership means, we can determine a stable level of funding that is in the national interest. But we will not fully benefit from this leadership if we do not also look at the flow of people and ideas from basic research into industry, health care, and other areas that matter to our society. In looking at how we can benefit from basic research, we should be pragmatic, not ideological. Let us do what is needed and what works. In this area we have much to learn. If we in the scientific community can better learn to articulate our own views, we will be less at the mercy of swings of public opinion. We will be better off, and so will our country.