


available. It is even dangerous to argue that in some fields we are relying on discoveries of unfettered research made a century ago. That argument can quickly be turned around, to say that if we haven’t yet fully exploited the last century’s work in those fields, why should be go much further in them now?

A much more persuasive case for research, including relatively unfettered one, can be made by citing extended research programs that have only recently started to have an impact in the marketplace, but are crucial today. The Computing Research Association has compiled an excellent report on just such developments [CRA]. It shows how a decades-long collaboration of universities, government, and private industry has provided the basic tools for the information society. Broad programs, directed at areas that appear to be especially promising or important, but ones that leave substantial freedom for individual investigators, may be the most promising approach. This may not be exactly “the free play of free intellect” advocated by Vannevar Bush, but it should be easier to justify to the public, promote technological change, and advance human knowledge. As Langmuir is supposed to have said, “You can’t predict what you will find, but you can make sensible bets on where to look.”


References


Demand immediate payback from research in their companies recognize the value of fundamental research as a public good, and speak up in favor of government funding [Coy, NAM, PT]. The idea that we can proceed without looking far forward, dealing with technical problems only as they come up, seems implausible. While we can certainly do that for a while, living off the advances made in the past, such a strategy is unlikely to be successful for long. Even if it were possible to dispense with long range research, it would not be advisable. As a simple example, the discovery of public key cryptography, mentioned in the Introduction, could be called premature, in that it is only now, almost 20 years after the basic invention, that public key cryptosystems are coming into widespread use. (Computers and communication networks had to become widespread for the need for public key systems to become acute.) However, the knowledge that this technology existed was of tremendous value, as it showed that electronic commerce and related goals could be achieved at low cost. This, in turn, facilitated planning for the information age, and forestalled unnecessary research for alternatives.

While one can justify some support for unfettered research, the case is not easy. Further, there seems to be no way to get back to the days in which individual researchers had to make hardly any justification for their projects, and areas grew with the stimulation of bountiful funding for all. Choices will need to be made, especially choices between fields. Scientists can show that they have been making good choices, given all the uncertainties involved, in deciding on directions within subjects. (Even there, though, there have been many mistakes, such as a famous university eliminating matrix theory from the requirements for an undergraduate degree in physics in the early 1920s, just a few years before the invention of quantum mechanics.) However, scientists have been notoriously bad in deciding on priorities between subjects. Unfortunately, such choices seem inescapable. While scientists usually feel that knowledge is good by itself, the public is unlikely to support the large scale research enterprise we have without utilitarian justification. Claiming that the superconducting supercollider would have absorbed a smaller fraction of the US federal budget than Tycho Brahe’s Uraniborg observatory did of King Frederick II’s revenue did not persuade many people to vote for the $15B project. The lack of measurable payoff for the economy from the last 50 years of experimental high energy physics (combined with the lack of pork barrel appeal) seemed to play a bigger role.

We do not have a convincing way to justify any particular split between unfettered versus directed research. We do not even know what the optimal level of total R&D effort is. Many scientists and mathematicians feel that their personal projects are crucial for the development of human knowledge. However, while one can perhaps make a similar case in art (if Verdi had been diverted from music towards a more mundane occupation, such as banking, would anyone else have composed the Manzoni Requiem, for example?), it is hard to argue this in science or mathematics. Most researchers are Platonists, and believe that they discover more than invent. It is impossible to dismiss out of hand the argument that long-range unfocused research should be abandoned or at least sizably decreased in favor of shorter-term projects with greater payoff. Fundamental discoveries that are not made soon as a result of such a downsizing would simply be made later. One can even argue that in the long run, human knowledge might benefit from such redirection, since the improved technology developed during the next decade or two of the digital revolution, as well as the higher standard of living and better education that result from it, will enable much faster progress on a broader front in the future. There are numerous examples of ambitious projects, such as machine language translation in the 1960s or Ted Nelson’s hypertext Xanadu project in the 1970s and 1980s, that were attempted too early, before the tools (hardware, software, and basic knowledge) to carry them out were
large economy than in a single corporation. However, even there the issue of “free-loaders” or “the tragedy of the commons” arises, since basic scientific advances diffuse rapidly around the world, and the taxpayers in the country funding a long-term project in basic research may not benefit much from it. Questions are also being raised about the rationale for government support of all basic research. I will not deal with all the issues that arise in this context (see Armstrong, ByerlyP, NAS1, for example), but will assume that government funding will be provided, and will concentrate on the issue of what type of research should be supported.

Universities are still a stronghold of unfettered research. However, they are also facing increasing pressure to change. Most of the factors discussed in earlier sections apply to university research as well, and there are additional special ones that apply in academia. The increasing size and specialization of the research enterprise, even in academic settings, means that what professors do is getting further removed from what they teach. This raises into question the justification of research as a crucial qualification for a teacher. Universities are also still set up for continuing exponential growth, and there are few of the negative feedback loops in operation that would stop the overproduction of Ph.D.s (cf. [Goodstein]). Further, the Ph.D.s that are produced are trained primarily for unfettered research. That preparation is inadequate for the non-academic jobs that more and more of them are taking. Hence there is increasing recognition that graduate education will need to be rethought and reformed [NAS2]. The general growth of scientific knowledge means that interdisciplinary research offers increasingly attractive opportunities. However, universities, with their rigid division into departments, are poorly positioned to take advantage of this. (One can go even further, and say that the fierce competition for tenure and grants is forcing faculty into extremely narrow and esoteric areas of research, which undermines the rationale for unfettered research. For a discussion of the deficiencies of the present peer review system, see [Lederberg].) Further, the same factors that operate in industry (the high payoff from short-term work, the relative lack of dramatic new innovations, the increasing scale of research projects, etc.) also influence the decisions makers who fund university research. There are frequent calls for publicly funded work to be more relevant. With increasing reliance on industrial funding or collaboration, there will be additional pressure on academic researchers to prove that what they do is of value to society. Even military-funded research was often relatively unfettered compared to what is often required in work with industry [Ghoshroy]. The need for large scale efforts in some areas of research, and the commingling of potentially profit-making work with pure research, will continue to put strains on academic institutions.

On the positive side, two factors (in addition to the inertia of the byzantine academic system) will serve to preserve at least some traditional unfettered research at universities. One is that the best way to prepare students for careers in science and technology in a world that is changing rapidly is not to train them in the narrow skills that happen to be in highest demand at the current moment, since those are likely to be obsolete in a few years. It is better to train them in more general skills, and in particular to concentrate on fundamental phenomena. That is also the way to attract the ablest students to science and technology, since they usually want to feel they are doing something basic that advances human knowledge, and not just tweak some process to increase profits.

Another factor that is likely to preserve a large fraction of the unfettered research being performed at universities, and perhaps even a small fraction of what used to be done in industry, is general unease with the trend towards short term research. Even industrial leaders who
for increasing their investment in research, including the unfettered type. These plans seems to have been shelved during the 1990s, and the question is whether this was a result of a reassessment of their needs, or of the recession and deflation of the Japanese financial bubble. Even though their profits have plummeted, Japanese companies have not decreased their R&D efforts (which already account for a larger fraction of GNP than in any other country), and the Japanese government is increasing its commitment to unfettered research [Normile]. The question is what will happen when the Japanese economy recovers. Japanese companies might be more inclined than Western ones to invest in long-term research. They tend to concentrate on the production of goods (where the lead times between initial ideas and marketplace applications appear to be longer than in systems and services). Also, their keiretsu structure makes its easier to justify long-term work, since the probability that some company in a grouping will be able to benefit from some unforeseen technological development is much easier than if just one narrowly focused company were involved. However, so far little information there is seems to point towards greater focus on short-term work even in Japan.

In American and European industry, the prospects for a return to unfettered research in the near future are slim. The trend is towards concentration on narrow market segments. Two decades ago, the hard disk market was dominated by vertically integrated manufacturers such as IBM and CDC. Today, leading firms, such as Seagate, basically do only system design, assembly, and marketing of their disks. They buy their disk controllers, motors, and other components from outside suppliers. The reason they can do this is that technology is widely available, and any basic research results are likely to be incorporated by the suppliers into their products and be available for anyone to buy.

Ford's River Rouge plant was the embodiment of vertical integration, with "iron ore and coal coming in at one end, and Model Ts rolling out the other." There was justification for this integration, since outside suppliers could not be relied on to provide the necessary quality and consistency of supply. Today the atmosphere is entirely different. The head of a business unit of a large corporation, whose division competes in a market where at least temporarily everyone appears to be losing money, mentioned publicly his efforts to persuade the other business unit heads to subsidize his division, to have assurance of supplies in the future. They were not interested, and even told him that they felt if his division went out of business, the other competitors would become stronger, and would therefore be able, through economies of scale, to lower their prices. This attitude is extreme, but apparently not uncommon.

At this point it is worth stating again that this essay is not meant to be a balanced account of research policies. It is intended to explain the reasons for the decline in unfettered research, and the turn towards more directed work. Therefore it emphasizes the negatives. There are valid arguments to be made for unfettered research, even in industry. I will not devote much space to explaining them, but they do include public relations, maintenance of ties with university researchers, and recruiting. Probably the most important reason is the need to have a window on future technological developments, to anticipate new threats and promising new directions for a company to pursue. All these reasons will presumably preserve some unfettered research in industry. However, these reasons have always been valid, and therefore in view of the negative developments listed in previous sections, they are not likely to lead to a reversal of current policies.

The arguments for taxpayer support of unfettered research are easier to make, since the benefits of wide-ranging undirected inquiry are more likely to be exploited someplace in a
is important to do as much as possible to ensure good selections. Industry estimates [Stewart] are that running a PC with the MS-DOS or Windows 3.1 operating system in a business environment for 5 years costs at least $5,000 more than a Macintosh one, which results in unnecessary total costs in the tens of billions of dollars per year. Does this not argue for much more research to enable better choices?

R&D spending as a whole has been increasing, and even spending on research itself has been increasing. Even when some companies or countries cut back, others more than pick up the slack. There seems to be little danger of a wholesale retreat from technology any time soon. It is true that China in ancient times repeatedly turned away from promising technological and economic developments. However, during those periods China was run by a centralized government that could exert tight control over society. Today’s world is broken up into many nation states that compete economically, a situation reminiscent of Europe during the formative stages of the current technologically oriented Western civilization. The success of governments is measured by the rate of growth of their economies. Further, economies of scale require participation in the world market. Therefore even though the changes that technology is bringing are socially disruptive, they are tolerated, as the primary goal is to maintain or increase international economic competitiveness. Since R&D is crucial for success in this environment, it is unlikely to decrease. However, the subject of this essay is the form of R&D that is undertaken, and especially how much of it will be D and how much R, and what form of R is to be favored.

To what extent is today’s decline in unfettered research a passing fad? Trendy policies do change, and forecasts are often wrong. As was mentioned in the Introduction, a century ago the U.S. patent commissioner thought that no significant discoveries remained to be made. Around 1940, it was widely thought that the nature of research had changed. A U.S. government report (quoted on p. 17 of [FloridaK]) stated that in “large industrial laboratories ... research has itself become a mass-production industry.” (Compare this to the Forrester 1995 quote from [TR] at the end of Section 2.) Even such perceptive observers of the economic scene as Joseph Schumpeter shared this view. Yet the following few decades saw the greatest flowering ever of industrial as well as university unfettered research. Are we likely to experience something similar in the future?

The current trend towards directed, short-term research can be thought of as a triumph of the Japanese style of R&D management. It was Japan that perfected the incremental improvement approach to technology. By adopting the quality improvement approach pioneered by Shewhart at Bell Labs, and extended by Deming and Juran, the Japanese blurred the lines between R&D and production, with even assembly line workers contributing to quality and efficiency gains. Their research was extremely goal-oriented, and usually would have been called advanced development in the U.S. There were close ties between R&D personnel and production and marketing groups. This approach triumphed in the marketplace, and gave Japan a lead in producing high-technology hardware, a lead that is still growing, as is shown by trade figures (provided those are adjusted for Japanese exports of production tools and sophisticated components to other Asian countries). (For more evidence of this lead, see [Hamilton], which shows that most of the potential bottlenecks in the manufacture of PCs are in Japan.) This approach is also what was adopted by the rapidly industrializing countries, and is what American and European firms seem to be striving for. However, the Japanese were cognizant of their borrowing of basic science from the West, and during the 1980s were making plans
penalty for failure to adopt new technologies has been, and continues to be, severe, as companies such as Wang Labs or Smith-Corona have demonstrated recently. In that sense there is a high return on new investment, as it provides a chance for an organization to survive.) This would seem to argue for greater investment in long term research. Paradoxically, though, it seems to have just the opposite effect. Research is an investment in the future. It therefore requires belief that there will be a future. A farmer will not engage in crop rotation and other soil conservation measures if there is a high expectation that a flood will carry all the topsoil away next year. However, that farmer will also not invest in improving the land (or even in cultivating it at all) if a risk-free bank deposit will earn a guaranteed 50% return per year. High current returns mean that future payoffs have to be discounted at high rates to get their present values, and therefore even ventures that are expected to be very profitable won’t be attractive if their returns are far off in the future.

The high returns that are obtainable from building products and services based on present knowledge lead decision makers to de-emphasize research on basic knowledge, and to ask the R&D community to “just do it.” This phenomenon also seems to explain the paradox that while it was clearly primarily the work in the physical sciences that gave us the computing and communication capabilities that are creating the digital revolution, research in those sciences is de-emphasized. Even Intel, the prototypical hardware company, is increasing its efforts in systems and software, since it sees the main barriers in those areas. Although the gap from basic insight to a marketable product is short in some areas, in others, especially where an entire new infrastructure is required, it is often still 15-30 years. Optical amplifiers took around 20 years from concept to deployment. Combinatorial chemistry, which only now appears to be coming into widespread use, is about three decades old. In fields where such delays are common, it is hard for managers to justify basic research.

5. Future prospects

The factors listed in the preceding sections have led to a shift towards short term, directed research in industry. These factors are all operating, and seem likely to continue to be in force for several more decades. They are also affecting research in universities and government laboratories. The question is whether this is good or bad, and what should be done about it.

The decreasing number of “big hits” is not necessarily a good argument for cutting back on research. The story is told that somebody once asked Einstein, “Professor Einstein, where do you get all your wonderful ideas?” That sage replied, “I don’t really know, I have only had two or three in my life.” There simply aren’t that many brilliant new insights to be had, and the easy ones have probably already been found. Much greater effort may be an unavoidable price to pay for technological progress. Some of the negative arguments about research can be interpreted as arguing for greater support of relatively unfettered investigations. For example, Section 2 discussed the evidence that in many of the areas of greatest interest to society, research is not far ahead of the marketplace. Does this not argue, though, that in those areas much more research is needed? Also, Section 2 argued that research results often are difficult to implement because of the need for industry standards, interoperability with other systems, and related concerns. However, since this is so, and because technology choices have such a huge impact on society, and once made, continue to have an influence for extended periods, it
The final theme that appears to play a major role in the decline in support for unfettered research is that of the information revolution. (Just as the previous theme, it is reminiscent of points made by Alvin Toffler in his books [Toffler1, Toffler2, Toffler3].) The examples of technological developments that have been cited so far, as well as phenomena such as “Moore’s Law,” are drawn largely from the computing and communications areas. This reflects to some extent my own background, but I believe it also reflects the reality that those areas are the ones that drive technological, economic, and social change today. The 21-st century may well be dominated by biology, but today it is computing and communications that appear to be most important, especially since they pervade all other areas. (Interestingly enough, while most technological forecasts have been incorrect, primarily because of overoptimism, the one area where these forecasts have consistently been too conservative has been in computing [Schnaars].)

The developments in communication and computing are leading to a fundamental transformation of human life. The technological developments that make this transformation possible, the fruits of extensive research, have been going on at a steady pace for a couple of decades. However, they have now reached a critical point where they are having a revolutionary impact. For example, the Internet has been growing at a steady pace for the last 25 years. It has caught public attention only recently because it became large enough to be noticeable and because its usefulness increased dramatically as more people got involved. Similar threshold effects operate in other areas. Information storage, retrieval, and processing are now rapidly passing performance levels that make more and more tasks doable in new ways. Automatic teller machines eliminated many bank clerk positions a decade ago, relatively primitive word-spotting voice recognition eliminated most telephone operator jobs a few years ago, and middle-management jobs are increasingly being squeezed out through new information systems. This Schumpeterian “creative destruction” is bound to continue for a long time. Technology takes a while to diffuse through society. It took several decades before the introduction of electric power to factories led to big productivity improvements, since entire manufacturing processes had to be reengineered. Change is faster today, but it still takes time to develop the organizational framework that takes full advantage of the best information technologies. Even if all progress in integrated circuit technologies stopped, we would still experience rapid and fundamental change in our lives. However, progress in all the crucial technologies has been steady, and is widely expected to continue for at least another decade. Thus we are bound to experience at least two decades of turmoil, as new technologies are absorbed by society. Neither new fundamental physical phenomena nor dramatically new mathematical or software tools are needed to make this possible.

Rapid advances in science and engineering mean there is a high return on investment in new technology. (This does not have to mean a sustained increase in profitability of corporations, even though Wall Street seems to think it does. Earlier eras of large investments in new opportunities opened up by novel technologies, such as those of the canals, then the railroads, then cars, did not lead permanent increases in profits. Competition saw to that. However, the
3. Racing against a moving escalator

The idea of technological progress is only a few centuries old. Before then, all the way through the Renaissance and even a century or two after, there was an astonishing respect and nostalgia for the ancients. In the end, it seems that it was only the Industrial Revolution that led to a transformation in the mindset of most people, so that steady improvements in technology are now expected. However, until recently, such improvements were perceived as discrete steps, such as the invention of the Hall process for extracting aluminum or the discovery of penicillin. Today, in contrast, we are living in a world of constant, rapid, and to some extent predictable progress. Microprocessors double in power just about every 18 months, in accordance with “Moore’s Law.” This helps explain phenomena such as the continued dominance of the Intel x86 architecture. In a more stable world, with only occasional discrete steps forward, RISC chips, which used to double the performance of Intel x86 CISC ones for a comparable price, would have displaced the latter ones, just as jet planes displaced propeller-driven ones. However, in an environment that is governed by “Moore’s Law,” Intel chips, just like RISC ones, doubled in performance every 18 months, so that by sticking with the Intel architecture, a customer only gave up 18 months, less than the life cycle of computer equipment. (The distinction between x86 CISC chips and RISC ones is narrowing today, but there used to be a sharp distinction.) This allows other factors, such as compatibility, to play the dominant role in determining what products to use.

The IBM acquisition of Lotus can be understood only in this light. IBM paid $3.5B for Lotus Notes. Now Notes is an innovative and important product, and no competitor has yet been able to produce a package with all the features of Notes. In a more stable world, though, it would not take long to develop a competitive product, which would limit the profit potential of Notes. What IBM has bought for $3.5B is a place a bit ahead of everybody else on the moving escalator, in the hopes that this will enable it to dominate an area that will likely be a crucial one for the information society we are developing.

The idea of constant change and having to keep up with the competition helps explain why there are so many bugs in commercially successful software products. The speed of delivery of new features is more important than quality. (Cf. [StalkH].)

The idea of rapid technological progress is deeply embedded in the minds of managers in high-tech areas. Several leaders of advanced development projects have told me, when asked what the main technical barriers were in their projects, that they did not see any. Those views were not entirely true, in that those managers could not go to market with the technology that was available then. However, they were taking for granted that microprocessors would get faster, and various other advances would be made. It’s just that they were so sure these improvements would be available, they did not have to concern themselves with them.

What this means for a researcher is that much more effort is required to have an impact. A 25% improvement in some process is always valuable. In a static world, it might be a breakthrough that could lead to great payoff. However, in a world governed by “Moore’s Law,” a 25% improvement is just 6 months along the technology curve. Hence the researcher who has an innovative way to save 25% or speed something up by 25% has to find a way to incorporate this improvement into all the other systems that are involved without delaying the project by more than 6 months. In effect, radically new ideas have to compete against the
is a common phenomenon. To quote from p. 199 of [Schmookler], a careful study of this subject,

[despite the popularity of the idea that scientific discoveries and major inventions typically provide the stimulus for inventions, the historical record of important inventions in petroleum refining, paper making, railroading, and farming revealed not a single, unambiguous instance in which either discoveries or inventions played the role hypothesized. Instead, in hundreds of cases the stimulus was the recognition of a costly problem to be solved or a potentially profitable opportunity to be seized; in short, a technical problem or opportunity evaluated in economic terms. In a few cases, sheer accident was credited.

However, at least until recently, this was not understood properly. Great credit was given to the perceived breakthroughs. Today, though, there is much more evidence on the subject, and much more stress is placed on incremental work [Gomory, FloridaK].

2.6. “Moore’s Law” and the predictability of science

One form of “Moore’s Law” says that microprocessors double in computing power every 18 months. This “law” has been followed closely over the last 20 years, and experts assure us that it will hold for at least another 10 before we encounter any serious barriers to further improvements in processors. (The main barriers seem to be more economic than technological, with the costs of fabrication facilities increasing rapidly.) Tremendous progress has to be made on a range of technologies, not only in devices but also in circuit designs, architectures, and software to continue this rate of advance. However, what we should note is that all large companies have been able to keep up with the progress predicted by Moore’s Law. TI has done it, Motorola has done it, and so have Intel, IBM, Toshiba, and others. Further, not a single one of these companies has been able to get far ahead of others for an extended period. In addition, it is understood what resources are needed to keep up this rate of advance. As long as no big mistakes are made, anybody with enough money can hire enough knowledgeable scientists and engineers to produce state of the art microprocessors. This goes against the popular image (that fits in with the justification for unfettered research) of a lone inventor having that one critical insight that changes a whole area. Management is not telling a researcher, “You are the best we could find, here are the tools, please go off and find something that will let us leapfrog the competition.” Instead, the attitude is “Either you and your 999 colleagues double the performance of our microprocessors in the next 18 months, to keep up with the competition, or you are fired.” An extreme example of the view of what research has evolved into was expressed by a distinguished researcher, Jay Forrester [TR]:

“... science and technology is now a production line. If you want a new idea, you hire some people, give them a budget, and have fairly good odds of getting what you asked for. It’s like building refrigerators.”
2.3. There are many ways to skin a cat

With the huge growth in research, there are many competing technologies that can be used for solving most problems. A decade ago, Narendra Karmarkar invented interior point methods for solving linear programming problems. This was a great advance, since it made many more resource allocation problems accessible. However, his discovery spurred researchers working on the traditional simplex methods to improve them so that today they are competitive with the interior point methods on most commonly encountered problems.

Twenty years ago, the most common modems were 300 bps ones, with acoustic couplers. Ten years ago, 9.6 kbps modems showed up. Now, as a result of improvements in microelectronics and in mathematical coding algorithms, 28.8 kbps modems are common, and 33.6 kbps ones are beginning to show up. There is work on modems that might get even closer to the absolute limit of 64 kbps that is imposed by central office equipment. If the 64 kbps limitation were truly absolute, each advance would be hailed as fantastic progress. However, for the customer, each advance has to be judged in light of other possibilities, and there are alternates to the use of modems. ISDN already offers 128 kbs, coax from cable TV companies promises several megabits per second, and optical fiber will eventually provide several hundred megabits per second. Thus advances in modems are valuable, but cannot be judged alone, and have to be compared to what other technologies offer.

2.4. It’s a complicated world

In most situations, a new product or service has to interoperate with others. This severely limits what can be done, and in particular limits potential profits for the inventor. A new coding scheme that leads to higher speed modems has to be accepted as an industry standard before consumers will buy it. Similarly, most control schemes for ATM networks have to be adopted by the whole industry before they can be used. Therefore the company that comes up with even a great invention can usually only obtain profits from licensing the patents and from a slight lead in marketing a new product.

2.5. Incrementalism wins

Incremental improvements have been much more important than striking new inventions. For example, commodity microprocessors have killed (at least for a while) the high performance computers based on exotic parallel architectures. Also, over the last two decades, there has been a series of predictions that progress in silicon integrated circuits was about to stop, and that new materials and devices had to be developed. Yet silicon still reigns supreme, while work on challengers, such as Josephson junctions or gallium arsenide (of which it has been said that “Gallium arsenide is the material of the future, and always will be”) is languishing. Formidable technical obstacles had to be overcome in silicon technologies to bring them to the present state, but at least to the public, this work appears incremental.

Incremental improvements have probably always been more important than new inventions in economic growth. Watt’s contribution to steam engine development is famous, but it was dwarfed in effectiveness by the cumulative impact of many other inventions in that area. This
The change in the role of research can be summarized in several interrelated points.

2.1. There are few big “hits”

Neither unfettered research nor any other kind has been producing the kinds of striking results that truly impress the public. Jonas Salk’s recent death led to recollections of the dramatic impact his vaccine had in defeating polio 40 years ago. Today, in spite of tens of billions of dollars spent on the “War on Cancer” over the last two decades, we have yet to see any treatment for cancer that can compare in its definitiveness to that of the Salk vaccine. Our knowledge of cancer has advanced tremendously, and current techniques are far more sophisticated than anything that Salk had at his disposal, but no “Magic Bullet” has been produced. The nice easy solutions have largely been found already. The problems we are facing are much harder. Therefore the payoff from investment in research is lower.

There are great successes across the whole spectrum of scientific knowledge, from Wiles’ proof of Fermat’s Last Theorem to the elucidation of the functioning of aspirin. However, they are hard to explain to the public.

2.2. Better technology does not always win

The world does not always beat a path to the door of the inventor of the better mouse trap, as Dvorak and other inventors of keyboards more efficient than the traditional QWERTY have found out. In video recorders, VHS beat the Beta format in spite of Beta’s technical superiority. Nobody seriously claims that the Intel x86 chip architecture was superior to that of RISC microprocessors, which typically had twice the performance with much smaller development efforts than Intel processors of the same period. Similarly, Microsoft Windows (3.1, 95, etc.) operating systems are only now beginning to catch up to where the Macintosh systems were several years ago. However, Intel has the lion’s share of the world microprocessor market, and Microsoft of the operating system market. One way to interpret this observation is to say that the basic technology is not the crucial issue, and that interoperability and related problems are. What that means, though, is that even a brilliant invention in chip design or operating systems is unlikely to make a big impact unless it is incorporated into either the Intel or the Microsoft products.

Even in the scientific and technical marketplace, tools require polished interfaces, so that it is often said that a tool for professionals will sell if it has a nice GUI, no matter what the basic technology is at the back end, whereas a wonderful new invention that requires the user to work at mastering it will be neglected. This again shows the decreasing importance of basic technology, a result of the greater competition in research that leads to competing products being close to each other. In contrast, there were few people who opted to consult witch doctors in preference to taking the Salk vaccine.
the performance of the best athletes has been improving in all fields, and the reason Babe Ruth stood out so much among his contemporaries is that he was the best from a smaller, less selective, and less well trained crowd. Today, the variation in performance among the leaders tends to be much smaller. Similar phenomena appear to operate in other fields. In science, Einstein attained a degree of public reverence that has not been accorded to any other researcher. However, if we could resurrect Einstein and clone 100 copies of him, the public would not treat each of these individuals with the same respect they accorded the original.

It is not only the public perception of research that has changed. The very nature of research has changed. A few decades ago, independent individual investigators or at most small labs were the norm. Today we are dealing with elementary particle accelerators that cost billions of dollars, and require teams of hundreds of scientists to operate. Even in mathematics, there is much more collaborative work, with the extreme example being the classification of finite simple groups, a great achievement of modern algebra that required dozens of researchers to work for several decades, and took 15,000 journal pages to document.

Large research projects are hard to fit into the traditional model of unfettered research. Just how much freedom to pursue “the free play of free intellect” (in the word of Vannevar Bush [Bush]) does a scientist working on dampening vibrations for the $300 million laser interferometry gravitational observatory have? The entire project may be aimed at unlocking Nature’s deepest secrets, and may be without any foreseeable practical application, but isn’t she just as fettered as an engineer developing a new air bag?

The gap between leading researchers and the general public has widened in most areas, with scientists pursuing topics that are increasingly esoteric. On the other hand, the gap between what researchers can do and what is available to the public in areas that the public sees first hand has narrowed considerably. As recently as 20 years ago, the best computational device that was widely available was a non-programmable pocket calculator, whereas researchers at leading institutions had access to supercomputers. This was a gap in computational capability of 8 to 10 orders of magnitude. Today, the Pentium processor in a home PC is not all that much less powerful than the fastest computers available at supercomputing sites, with the gap only two or three orders of magnitude. The Pentium PC is often more powerful than the workstation that a typical engineer or scientist uses. Similarly, in speech or optical character recognition, all that existed 20 years ago were research systems, available only in a few labs. Today, in contrast, one can buy much more capable “shrink-wrap” software for a PC. What is most remarkable, though, is that this software does not differ all that much in performance from the most advanced research systems. The time between invention and a commercial product has shrunk dramatically in many areas, so that research is not far ahead of the rest of the world, and therefore has fewer advanced toys to impress the public with.

The main justification for unfettered research was that scientific discoveries could not be predicted, and that allowing researchers to follow their intuition in selecting problems to work on was the best policy, one that would result in enough significant new results that some of them would pay off handsomely for the sponsoring organization. Roentgen’s discovery of X-rays and Fleming’s of penicillin are two examples of such unpredictable discoveries that both had great impact. For that argument to be valid, though, many significant scientific advances have to occur, a large fraction have to be of interest to the sponsor, and there have to be opportunities to exploit them. These assumptions are no longer believed by industrial R&D managers, and are being questioned by national policy makers.
could not be sustained indefinitely, but it continued for a long time. Now, however, there are clear signs of a leveling off, at least in the developed countries. The end of the era of exponential growth appears to have arrived, and this all by itself may be responsible for many of the complaints about poor job prospects and low morale.

While exponential growth had to stop at some point, why did it have to happen now, and not after another doubling in the population of scientists and engineers, say? There does not seem to be any persuasive answer. Nobody even has a good theory of what the optimal rate of investment in R&D should be. As an empirical matter, no advanced industrial country invests more than 3% of its GNP in R&D. The prospects of going above that figure seem dim. Cutbacks in R&D spending are talked about more frequently than boosts.

To maintain some balance, we should mention that the traditional image of unfettered research is not completely accurate. Research has seldom been totally unfettered. Ehrenreich’s experience at GE, cited in the Introduction, can be ascribed to GE’s interest in the then brand new field in semiconductors, where almost any research was likely to be of commercial value. Had Ehrenreich suddenly decided to switch to genetics, he would surely have found his freedom much more limited than it seemed. Even university researchers are much more fettered than is popularly imagined. Except for tenured professors in theoretical areas, almost all researchers depend on grants or contracts in their work, and to get them, they have to survive peer review scrutiny that sharply limits what they can do.

Cocke’s invention of RISC (reduced instruction set computers) is an example of unfettered research, in that IBM allowed him to work on whatever he chose. However, Cocke had already made valuable contributions to the design of IBM mainframes, and was working on computer architectures, which were of obvious relevance to IBM. Thus it was not surprising that he was given the freedom to pursue “the free play of free intellect.” Industrial research laboratories have been giving such freedom to a few select individuals for a long time (Steinmetz at GE over a century ago may have been one of the earliest ones) and surely will continue to do so in the future. What is remarkable about the era of unfettered research that is ending now is that even brand new Ph.D.s were being offered such freedom in large numbers. The purpose of this essay is to explain why less freedom is being offered today.

The general decrease in support for R&D and for unfettered research in particular are surely strongly connected to the rapid growth in research. Whatever the optimal level of R&D spending is, even if it is twice the present level, we are surely much closer to it now than we were three or four decades ago. With the large current community of scientists and engineers, the public and the decision makers are no longer dealing with a few souls laboring in obscurity in their ivory towers, but with a large community that can easily be seen as just another “special interest group” looking for its “entitlements.”

The growth and increasing competitiveness of any field can easily affect the public perception of that field. In sports, for example, it is common for commentators to talk of how some famous figure of old, such as Babe Ruth, was the greatest player of all time. Such assertions are made only about sports such as baseball or boxing, where teams or individuals compete against each other, and there is no objective measurement that can be used to compare performance over time. In sports such as swimming or running, where the clock determines the winner, such assertions are never made, since the evidence there is clear, that the performance of the top athletes has been steadily improving. The explanation for this phenomenon is that
experimental areas. (Recall the years that Einstein spent working in the Swiss patent office, before he could obtain a university position.) During the period between the two world wars, scientists and engineers were devoting an inordinate effort to chasing after the few funding sources that were available [Burke].

After World War II, support for research expanded tremendously, first in the United States, and then, as their economies recovered or developed, in other countries. Science and technology had played a vital role in the war. The Bomb was the most famous development of that period, but there were many others, such as radar, plastics, and jet engines. There were also striking advances in non-military areas, such as the almost total (for a time, at least) conquest of infectious bacterial diseases by penicillin and other antibiotics. The technologies and products that captivated the public’s attention were the results of concentrated development efforts. (The Manhattan Project was more of an engineering than a scientific enterprise, although many of the best physicists, chemists, and mathematicians spent their full war-time careers in it.) However, they were correctly perceived to be the culminations of the unfettered research carried out in the preceding decades. The nuclear era, for example, is often traced back to the serendipitous discovery of radioactivity by Becquerel at the end of the 19-th century. The policy makers and the general public responded with an unprecedented increase in support for scientific and engineering research. Governments proceeded to fund extensive research and development in all areas, not only in universities, but also at their own laboratories and through industry. Universities vastly expanded their own commitment to research, decreasing teaching loads, and placing more emphasis on scholarly achievement than on teaching. Industry also proceeded to build up R&D staffs. Under the leadership of people like Vannevar Bush [Burke], who had bitter memories of the lean years between the wars, policy makers allocated substantial fractions of available funding to unfettered research, in the belief that scientists would themselves be best able to select the most promising direction for their work.

To get a graphic appreciation for the growth in the research establishment, it is instructive to look at pictures of participants at any of the Solvay congresses held between the world wars. There are only a few dozen people in any one of these pictures, but they usually contain most of the creators of modern physics, scientists like Bohr, Einstein, and Heisenberg. Today, a typical physics conference has hundreds or thousands participants, and there are many more conferences than before. When we talk of the decline in unfettered research, we should remember that there would be no difficulty in providing an unfettered research environment for Einstein today, were he still alive. The difficulty is that there are now thousands of theoretical physicists who would like to be treated like Einstein.

The growth of the research establishment is not a new trend, but it accelerated after World War II. The number of scientists has been increasing at an exponential (in the strict mathematical sense of the word) rate for centuries. The number of scientific papers published annually has been doubling every 10-15 years for the last two centuries [Price]. With the spurt in funding after World War II, the rate of increase rose. For example, the number of abstracts in Chemical Abstracts just about doubled every decade from 1945 to 1985, when it reached about half a million per year (and has increased at a lower rate since then). The volume of technical publications today is more than 10 times what it was at the end of World War II. Although there have been frequent complaints about the inadequate level of federal government support for science, the U.S. National Science Foundation has doubled the number of researchers it supports in the last 20 years. It was always clear that this rate of increase
there has been growth [Coy], with cutbacks at some corporations more than balanced by growth in others. Worldwide, even unfettered research is probably increasing, with universities in the rapidly industrializing countries expanding their faculties. The only area where there is a clear decrease is in unfettered research in industry. This type of research has never been a large part of the total R&D budget. However, it has been the most visible type of research, in that it tended to garner public attention and Nobel prizes. That is one reason for discussing its decline. Another is that the decrease in unfettered research is symptomatic of a general shift in R&D towards much more directed and shorter-term type of work. Not all directed investigations are short-term (the invention of the transistor is a good example of successful long-term directed research), and the correlation between unfettered research and long-term one is not perfect, but there is a correlation, so that the decline in unfettered work serves as a measure of the how far into the future one looks. The turn towards short-term work is a phenomenon that has occurred at most industrial laboratories, even those that had never engaged in unfettered research [Coy]. Is this a trend that will continue much further? Will it spread to research at universities? An examination of the reasons for the decline in industrial unfettered research suggests that the answer to both questions is yes.

2. Research as a commodity

The basic reason the role, style, and image of research are changing is that there is much more research than a few decades ago. There used to be much less competition, and the intervals between invention and marketing of a product were long. As an example, xerography was invented by Carlson in 1937, but it was only commercialized by Xerox in 1950. Furthermore, there was so little interest in this technology that during the few years surrounding commercialization, Xerox was able to invent and patent a whole range of related techniques, while there was hardly any activity by other institutions. This enabled Xerox to monopolize the benefits of the new technology for over two decades. Xerography was not an isolated case. When the transistor was invented by Bardeen, Brattain, and Shockley at Bell Labs in 1948, several years elapsed before other laboratories acquired enough expertise in the semiconductor area to make significant contribution. Had AT&T not been prevented both by its culture and by regulation from exploiting this dramatic discovery, it could have erected a patent barrier around its discovery that would have allowed it to keep most of the profits from commercial developments. Today such opportunities are extremely rare. For example, when Bednorz and Mueller announced their discovery of high-temperature superconductivity at the IBM Zurich lab in 1987, it took only a few weeks for groups at University of Houston, University of Alabama, Bell Labs, and other places to make important further discoveries. Thus even if high-temperature superconductivity had developed into a commercially significant field, IBM would have had to share the financial benefits with others who held patents that would have been crucial to developments of products.

The era of extensive unfettered research started only after World War II. Until then, “curiosity-driven” research was practiced at universities and a few academies and research laboratories, but it was done on a small scale. During the last century universities developed the view that research is one of their main missions, along with teaching. However, their resources, while adequate for small scale and low-overhead work in theoretical areas, could not cope with the increasing number of researchers and their ever more costly equipment in
systematize and justify such procedures.

Competition and pressure for quick financial payoffs are often cited as causes of the decline in unfettered research in industry. However, Wall Street can be remarkably sensitive to technological advances and extremely patient in waiting for profits. This is demonstrated by the recent initial public offering of Netscape Communications, Inc. A company that had been in existence for less than two years, had racked up total sales of less than $20 million, and had only losses and no profits, was suddenly judged to be worth $2 billion. The reason was that the Netscape Navigator browser that Andreessen, Bina, and their colleagues built won the popularity contest among Internet users. Several companies were pursuing the same strategy as Netscape, of giving out their browsers for free. It was Netscape’s technical superiority that won the contest (even though this superiority was slight, and much of the programming was of poor quality, as was shown by the security flaws that were discovered in the Netscape program). It will be a long time (if it ever happens) before Netscape earns enough profit to justify its initial stock market valuation. However, the prospects of leveraging control of the most popular World Wide Web browser into control of Internet software were sufficiently enticing for Wall Street that it was willing to assign an outlandish value to this company. Thus in cases where there is a clear connection between technical excellence and market applications, financiers can temper their demand for profits. Hence financial pressure alone does not explain the cutbacks in unfettered research.

This note proposes a somewhat different explanation for the decline of unfettered research. I feel that it was caused primarily by internal developments in the world of science and technology, not by arbitrary outside decisions. There seem to be three main (and closely interrelated) themes that help explain the turn towards short term, directed research:

(a) dramatic increase in volume of research,

(b) steady and rapid progress in all areas of technology,

(c) unprecedented opportunities in applying existing knowledge.

These themes arose from the success and growth in research.

I am not implying that research is finished as an important activity. I do not share the opinion of a former head of the U.S. Patent Office, who resigned a century ago and recommended that his position be abolished, since “everything that can be invented has been invented.” Nature has a limitless supply of secrets to be explored and exploited. I am convinced that research is important to society’s economic welfare, will continue to advance human knowledge, and can be an intellectually rewarding career. This note is not even intended as an attack on unfettered research. My intention is to explain why this type of research is on the decline, and so I concentrate on the negative side. I am not attempting a balanced treatment of the role of R&D and the optimal level or form of support for research. Most of this note is descriptive, not prescriptive.

It is necessary to emphasize that R&D spending as a whole, and even research by itself, have generally been going up at a steady pace. There is a sense that cutbacks might come, and that the relationship of science to society might have to be reexamined (cf. [ByerlyP]), but no substantial cuts in overall R&D budgets have materialized so far. Even within U.S. industry,
A good recent example is the invention of public key cryptography by Diffie, Hellman, and Merkle in the 1970s. Their work was funded by Hellman’s NSF grant in information theory. Since the grant terms were flexible, without negotiated work plans or deliverables, they were free to work on an audacious idea that was not mentioned in the grant proposal, and that formed a foundation stone for the information society.

The unfettered research that Ehrenreich encountered at GE in 1956 is almost universally taught at universities as the only real research. However, this type of research is now almost totally gone from industrial and government laboratories, and is under pressure even in academia. Industrial leaders stress the need to “focus on customers’ needs.” The pressure is to do work with quick payback, and to justify everything that is undertaken on the basis of its relevance to the corporation [Coy]. The trend is not confined to the U.S. For example, the Canadian Finance Minister introduced a proposed new budget with the stipulation that “[i]n the future, [Canadian] science and technology efforts will be concentrated more strategically on activities that foster innovation, rapid commercialization and value-added production ... to stretch government’s science dollars further and more effectively.”

Today researchers, even when they are not working on well defined projects, are increasingly required to submit plans for their research, and to explain not just how their work might help their employer, but what steps they are taking to ensure their results are utilized to the fullest. Scientists are in effect asked to become engineers, where the term engineer is not used in a derogatory sense, but, in the words of Vannevar Bush (as quoted in [Zachary]), describes a person who is

“not primarily a physicist, or a business man, or an inventor, but [someone] who would acquire some of the skills and knowledge of each of these and be capable of successfully developing and applying new devices on the grand scale.”

Much of the distress experienced by scientists is caused by this pressure to understand and justify how their work fits into a much larger setting. At a time when growth in knowledge requires more training and much greater specialization, they are being asked to become generalists, and to de-emphasize leading-edge work in their fields.

Many of the popular explanations for the decline in unfettered research are unsatisfactory. Short-sighted management is often blamed. However, the decline has been going on for a long time, and different companies have followed varying policies, so that if traditional unfettered research were the best method for a corporation, its superiority should have become evident by now. That has not happened, and most of the rapidly expanding high-tech companies appear to do little long-term research, unfettered or not. Another explanation that is often proffered is that the end of the Cold War has cut back funding for R&D in general. Again, though, one might expect that reductions in military-related R&D would have led to lower demand for scientists and engineers in general, and therefore to lower salaries, which would have made civilian R&D cheaper and thus more attractive. Thus this explanation also seems to be inadequate. (In addition, this explanation leaves open the question why the Cold War should have spurred unfettered research in the first place. Why weren’t all available resources put into building more tanks and submarines?) Novel management theories (cf. [RSE]) are often blamed by themselves, but as is true of most management theories, they tend to be derived from observations of what companies perceived to be successful are doing, and only
The decline of unfettered research

Andrew M. Odlyzko

AT&T Bell Laboratories
Murray Hill, New Jersey 07974
amo@research.att.com

revised version
October 4, 1995

1. Introduction

We are going through a period of technological change that is unprecedented in extent and speed. The success of corporations and even nations depends more than ever on rapid adoption of new technologies and operating methods. It is widely acknowledged that science made this transformation possible. At the same time, scientific research is under stress, with pressures to change, to turn away from investigation of fundamental scientific problems, and to focus on short-term projects. The aim of this essay is to discuss the reasons for this paradox, and especially for the decline of unfettered research.

What do I mean by unfettered research? In the discussions of federal science policy it has also occasionally been called “curiosity-driven.” It is exemplified by the following reminiscences of Henry Ehrenreich [Ehr].

When I arrived at the General Electric Research Laboratory at the beginning of 1956, fresh from a PhD at Cornell, I was greeted by my supervisor, Leroy Apker, who looked after the semiconductor section of the general physics department. I asked him to suggest some research topics that might be germane to the interests of the section. He said that what I did was entirely up to me. After recovering from my surprise, I asked, “Well, how are you going to judge my performance at the end of the year?” He replied, “Oh, I’ll just call up the people at Bell and ask them how they think you are doing.”

In this style of work, the researcher is allowed, and even required, to select problems for investigation, without having to justify their relevance for the institution, and without negotiating a set of objectives with management. The value of the research is determined by other scientists, again without looking for its immediate effect on the bottom line of the employer. The assumption that justifies such a policy is that “scientific progress on a broad front results from the free play of free intellects, working on subjects of their own choice, in the manner dictated by their curiosity.” (This quote is from the famous report of Vannevar Bush [Bush] that formed the cornerstone of U.S. federal funding for research after World War II.)

Unfettered research is not a boondoggle, simply indulging scientists’ curiosity. There were good historical reasons for Vannevar Bush to advocate it. Many of the stunning scientific and technological advances that have shaped our world have come from unfettered research.