DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

GOVERNMENT POLICY AND TECHNOLOGICAL INNOVATION: WHERE DO WE STAND AND WHERE SHOULD WE GO? *

Roger G. Noll

*In Government Policies and Technological Innovation, National Technical Information Service. National Science Foundation, Washington, D.C. 20550.

SOCIAL SCIENCE WORKING PAPER 86

May 1975

GOVERNMENT POLICY AND TECHNOLOGICAL INNOVATION: WHERE DO WE STAND AND WHERE SHOULD WE GO?

Roger G. Noll
California Institute of Technology

This paper is addressed to two questions: (1) Does the existing state of knowledge within the body of social science research provide an adequate information base for making decisions about public policies that affect technological change, and (2) What additional research might prove especially fruitful in its contribution to understanding the relationships between policy and innovation? In discussing these issues, every attempt will be made to push the existing state of knowledge as far as possible with respect to policy implications. Some of the propositions put forth are reasonably well established theoretically or emprically, but others should be accorded the status of unproved but plausible inferences.

The underlying disciplinary perspective of this survey is distinctly that of economics. The objective is to identify (1) microeconomic optimality conditions with respect to the rate and pattern of innovative activity and (2) institutions that would either cause the economic system to operate in a way that satisfied the optimality conditions or that, at least, unambiguously improved upon the current

* Financial support for the research reported here was provided by the National Science Foundation National R & D Assessment Program, grant No. DA 39495. I am particularly grateful to David Grether, Burton Klein, Charles Plott and James Quirk for comments on an earlier version of this paper

performance. The second objective requires that the costs of managing an institutional arrangement be incorporated into the analysis. Consequently, the literature on nonmarket decisions within organizations and on the behavior of institution-managing bureaucracies, notably government agencies, is also relevant.

The ideas presented in this paper are distilled from the papers written for the Caltech R & D project on specific aspects of the relation between innovation and public policy. This paper is the summary of that project. The more detailed papers should be consulted for references to the relevant literature and for complete defenses of the propositions or of the conclusion that, on some issues, the literature leads only to a conundrum.

THE INNOVATIVE PROCESS

Try as we may, there is really no sensible way to begin a discussion of the notion of "optimal technical progress" without reference to some fairly abstract concepts in economic theory. The reasons are, that, first, technical knowledge is an economic good -- it has value and is not infinitely available at zero cost -- and, second, the market for it is the embodiment of virtually every source of "market failure" that economists have ever investigated. The principle organizing concept behind most of the succeeding discussion is the principle lesson to be drawn from what is probably the most arcane topic in economics, general equilibrium theory, which can be expressed very crudely as follows:

<u>Proposition 0</u>: If perfectly competitive markets exist for all economic goods (including all futures and contingency markets), then the allocation of resources among productive activities will be optimal in the sense that any change in that allocation must make someone worse off; further-

more, any allocation of society's resources that is optimal in the same sense can be reached through a complete, competitive market system from some given initial distribution of resources among consumers.

The difficulty with Proposition 0 is that, for numerous reasons, its presumptions are not desciptions of any real economic system. Not all markets exist, and not all that do exist are perfect. Furthermore, markets systems are costly to maintain in that resources are required to accommodate market transactions. Nevertheless, public economic policy can be viewed in the context of the proposition as seeking to create insitutions that perform the functions of a market in situations where markets are unworkable or more costly than substitutes. Furthermore, since market failure can have many causes, the best choice of an institution to substitute for a market in any given case is likely to depend upon the nature of the market failure.

The preceeding conceptual base underlies most of the work done by economists in building a theory of the innovative process. A divergence from optimality in technical progress is viewed as emanating from a failure of the market for ideas. The sources of failure that have been identified are: (1) inappropriability: the inability of the innovator to capture the full economic gains made possible by his innovation; (2) uncertainty: the economic uses of the technical ideas that will emanate from R & D activities are not known fully in advance, so that the search for innovations is a gamble; (3) indivisibilities: the minimum efficient scale of R & D operations can be sufficiently large that the market for a particular class of ideas can not be perfect; and (4) indirect failures: if a good must be produced and distributed outside of a perfect market environment, the institutions created to bring this about may lead to inefficiencies in the advancement of knowledge with respect to production and distribution of the good. Each of these notions will be examined separately.

Before proceeding with an examination of these issues, a few general remarks and definitions are in order.

The conventional wisdom among public policy-makers and as expressed in much of the literature concerning technological innovation is that, somehow, there is not "enough" research and development, and that government can and should attack this problem through various policies that increase financial incentives for R & D and innovative behavior generally. The single most important conclusion of this paper, and of the Caltech R & D project generally, is that the existing literature does not establish either that too little innovative activity takes place, or that government can be particularly effective in devising cost-effective strategies to promote more R & D, or even which of the four sources of market failure listed above is the most important and, therefore, should receive the most attention from policy makers. Furthermore, existing policies do, on balance, promote R & D relative to other investments, at least in industries not subject to public utility regulation, so that even if the case is made that a private market economy generates too little R & D, it is by no means established that this is not offset by existing policy interventions. Much more solid theoretical and empirical research must be undertaken before any of these issues can be resolved sufficiently to warrant strong conclusions about the general stance of policy towards innovative activity.

The following discussion adopts two semantic conventions. First, the major ideas that have been distilled from the literature are stated as <u>propositions</u>. These are not necessarily restatements of theorems from the theoretical literature, although in a few cases they are. Some are empritical findings, and some are plausible but unproven inferences. Thus, a proposition is a summary statement of a research conclusion, and can have a status anywhere from a reasonable hypothesis or conjecture to a natural law.

Second, an attempt has been made to be meticulous in differentiating between <u>absolute</u> and <u>relative</u> firm size. Adjectives such "large" and "small" refer to absolute size, while "monopolistic" and "competitive" refer to firms that do or do not possess market power. In an oligopoly, the firms with the most market power are "dominant" or "leading," while the others are "dominated." The two concepts of firm size -- absolute vs. relative -- play quite different roles in the literature, although authors do not always state clearly which concept they are discussing.

Appropriability

The appropriability problem arises from the presumption that to some extent the discoverer of new information can not prevent others from taking advantage of it without paying the discoverer the full value of the information. Since innovators do not capture all the benefits of their discoveries, it follows that:

Proposition A1: Because of the absence of full appropriability, too few resources will be allocated to developing new knowledge.

This proposition has been subject to challenge in two ways. First, some doubt the validity of its premise: that technical knowledge is, to an important extent, inappropriable. While unanimity can not be reached on the issue, consensus can probably be reached on two points:

Proposition A2: The extent of appropriability differs from industry to industry, and is likely to be greater for monopolies (no competitors will exploit the monopolist's ideas).

Proposition A3: Presuming that a continuum of R & D activities from basic research through developmental research through final product innovation makes sense, the more basic the idea the less appropriable it is likely to be.

One prediction arising from A2 and A3 is that the greater the market power of the firm, the more likely it is to be innovative since it can appropriate more of the results of its R & D. Empirical tests of

this prediction are inconclusive, in part because of the failure to date of researches to develop a satisfactory measure of market power.

The second line of attack on A1 is based on its implicit premise that innovation is economically motivated. The argument against profit motives is most convincingly made with respect to basic scientific research. Most basic research takes place in nonprofit institutions and its results are made freely available. This does not establish that the amount and type of innovative activity taking place is insensitive to financial incentives; instead, it merely suggests that even if financial incentives are lacking, some advancements in knowledge will still take place. Thus,

Proposition A4: While the rate and pattern of technical change is sufficiently sensitive to financial incentives that the workings of the market for ideas are a legitimate policy concern, nevertheless some important advances in knowledge are not in response to incentives and would occur even if there were total failure in the market for knowledge.

In principle, at least, A4 weakens A1. The absence of commercial R & D activities in an economic arena in which nonprofit institutions are active may be due to inappropriability, but it may also be because the nonprofit institutions are so active that the rate of return to further R & D in that arena is, even with full appropriability, uneconomically low. Although empirical studies show that the average return to investment in R & D exceeds the average return to other investments, the evidence is not conclusive that these differences persist at the margin (i.e., that they apply to the last dollar invested in each category) or that the differences are greater than is appropriate for the differences in the riskiness of R & D and other investments.

Several institutions are used to increase the appropriability of technical knowledge, notably the patent system, copyright laws, and protections of trade secrets. These systems, being voluntary (a firm need not take advantage of them to protect its rights to knowledge), must, since they are used, increase the appropriability of innovations; however, they can not eliminate the appropriability problem:

Proposition A5: The system protecting rights to knowledge, because it is costly to operate, still results in too little R & D since innovators must expect one cost of an innovation to be the deployment of the rights system to protect their discoveries.

At the same time, the existence of a system granting rights to knowledge generates an incentive to discover ways to accomplish the same end as another innovation, but in a manner that does not infringe upon the rights of others:

Proposition A6: The system for securing rights to knowledge generates socially unproductive R & D organized solely to invent around the discoveries of others.

That "inventing around" should take place is difficult to explain within the conventional economics paradigm. Both an original innovator and a copier should have an incentive to make a licensing arrangement instead of having the copier pursue secondary research activities. The copier expects to spend some given amount on R & D in order to copy successfully without infringing on the innovators rights; once he succeeds, the innovator receives no financial return from the sales of the copier. Thus, any licensing fee between zero and the R & D costs of the copier ought to benefit both sides.

Finally, the importance of the appropriability of innovation in leading to resource misallocations depends upon the elasticity of supply of innovative ideas. In the extreme, for research activities that are highly labor-intensive and in which there is essentially no possibility of improving the productivity of creative talent by giving talented researchers more resources -- an example might be abstract areas of mathematics, theoretical physics or even economic theory -- an increase in the appropriability of research results will, if all of the people with talent for that

activity are already engaged in it, lead only to increases in the income of the research personnel, with no increase in innovative output. More generally:

Proposition A7: The more inelastic the supply of innovative ideas at and above the current equilibrium, the less significant the divergence from optimality in innovative activity due to inappropriability of ideas.

An institutional arrangement that increases appropriability of resources can, in principle at least, actually worsen allocative efficiency, in part because of the effects of A5, A6 and A7 and in part because appropriability is a euphemism for monopoly:

Proposition A8: A fully appropriable innovation will be used less extensively than is economically optimal.

Only a perfectly discriminating monopolist can extract monopoly rents without destroying allocative efficiency. Otherwise, a monopolized economic good -- whether an idea or a more mundane product -- has too high a price. Since the social costs of allowing one more producer to have access to a fully-developed cost-saving idea are zero, any positive licensing fee for employing it excludes, in principle at least, some users who would find the idea economically valuable, but less valuable than the license fee.

Summarizing A5 through A8, it is not necessarily true that increasing the appropriability of innovations improves the rate of technical progress or increases the efficiency with which creative resources are deployed:

Proposition A9: Increasing the appropriability of innovations will reduce the rate of technical progress if the supply of innovative ideas is sufficiently inelastic that the new knowledge forthcoming because of greater appropriability does not offset the effect of greater appropriability on (a) incentives for "inventing around," (b) the costs of maintaining property rights to knowledge, and (c) monopolistic inefficiencies in exploitation and diffusion of new ideas.

Proposition A10: Increasing the appropriability of innovations can reduce the efficiency of the allocation of creative talent among alternative R & D activities if either of the following effects occur: (a) the increase in appropriability is greatest for innovative activities for which appropriability is already relatively high, or (b) for types of innovations for which appropriability is increased the most, the change in appropriability causes more resources to be committed to "inventing around" and maintaining the system of rights to knowledge than are added to R & D activities associated with the same types of innovations.

The principle lesson to be learned from the preceeding discussion is that the problem of creating an efficient market for ideas is exceedingly complex. The degree of appropriability that should be accorded to innovative activity of any particular type should depend upon the original appropriability of innovations generally and the type in question in particular, the extent to which exogenous developments can be expected, the supply elasticity of innovation, and the costs of greater appropriability in terms of maintaining rights and monopolizing ideas. It is likely that, in the real world, conditions will differ enough from industry to industry and along the continuum from basic to applied research that no single system of establishing rights to knowledge and markets for them will produce satisfacotry results in terms of its effect on the rate and pattern of innovation.

Uncertainty

The effect of uncertainty on innovative activity arises because of the absence of adequate insurance against the failure of an R & D project. There are basically two reasons why a contingencies market

may not develop for an activity subject to uncertainty. First, if the person buying insurance can control the likelihood that the event being insured against will transpire, the existence of insurance can create incentives to devote less than full effort to avoiding the contingency. Thus, if a firm could buy full R & D insurance -- that is, the insurance company would repay the firm the cost of the R & D program if the program did not produce a commercially viable output -- it would have less of an incentive to make its R & D project succeed. Second, the potential innovator is likely to have a different estimate of the likely success of a project than an insuror. Assuming that information and expertise add to one's ability to predict the potential of an R & D activity, the innovator is likely to perceive less uncertainty about the project than those from whom he might buy insurance; therefore the expected riskiness of insuring the project -- hence the insurance premium -- will be higher to the insuror than if he had more complete knowledge. As a result:

Proposition U1: Inadequately developed contingency markets for innovative activities cause innovators who are more risk-averse than average to engage in less innovative activity than is optimal.

Proposition U2: Inadequately developed contingency markets for innovative activity cause innovators with atypically great risk-taking propensities to over-invest in their own ideas (risk-taking innovators would do better by investing in the best ideas of risk-averse innovators than in the least productive investments based upon their own ideas).

The principal nonmarket substitution for contingency insurance is the "self-insurance" associated with investing in a diversified protfolio of projects. The rationale behind self-insurance, or portfolio theory, is, in its simplest case, that a series of very small bets is less risky than one large bet -- i.e. it is less risky to flip a coin for a dollar twice than to flip once for two dollars.

Self-insurance can take place only in organizations that are large enough to carry on a large number of diverse research projects. Assuming that R & D projects must be operated above some minimum rate per time period in order to conclude soon enough to be worth pursuing, then:

> Proposition U3: Abstracting from any effect that firm size may have on the effectiveness of a firm's organizational structure, its risk-taking propensities, or the extent of competition, large firms with diverse R & D programs will undertake more projects subject to risk and uncertainty than will small firms, and a given industry will be more innovative with a smaller number of large firms than with a larger number of small firms.

Uncertainty, per se, does not lead to the conclusion that if markets are monopolized, firms are more innovative, as does the inappropriability argument; however if one dimension of uncertainty is the extent to which rights to the new idea can be enforced, then a monopolistic firm in a market with blockaded entry -- for example, a regulated public utility -- will perceive less uncertainty in a given R & D project than will a competitive firm. Aside from this effect, the uncertainty of R & D projects leads to an argument for large firms, but not for uncompetitive ones.

In fact, one procompetitive argument is based on the presence of uncertainty. If competitors are engaging in R & D, a firm is subject to the risk that competitors will discover a process that will significantly reduce the profitability of the firm. If the expected return from R & D projects to the firm is zero -- that is, it anticipates equal future profits whether it invests in R & D or not -- then under certain plausible conditions failing to engage in R & D can be subject to more uncertainty than undertaking R & D projects:

> Proposition U4: Engaging in R & D activities at the norm of a competitive industry when R & D has zero expected return is less risky than engaging in no R & D as long as innovations are not fully inappropriable and most firms engage in more or less the normal amount of R & D for firms in the industry.

The point is simply that the uncertainty that projects will succeed must be weighed against uncertainties concerning the future market position of a firm with respect to its competitors. Thus, expectations about the behavior of other firms, which in turn are probably based upon the historical pattern of innovation in the industry, can be a crucial factor in determining each firm's decisions about future R & D and innovation.

In a monopoly in which entry is effectively blockaded, the firm's expectation about competition is that other firms are not a potential threat. Hence:

> Proposition U5: For innovations of predetermined appropriability (i.e. assuming market power does not contribute to appropriability) and a given degree of uncertainty with respect to technical and marketing success, a monopolist is less likely to pursue and adopt them than a competitor.

The reasoning behind the proposition is that, holding everything else constant, the absence of a competitive innovative threat reduces the uncertainty associated with doing nothing.

Finally, uncertainty in the outcome of R & D activities has an important implication with respect to selecting the best R & D strategy. In an uncertain environment, R & D programs are more likely to be profitable if their organization and operations are characterized by flexibility, mixed strategies, and sequential decision-making, all of which are designed to use the added information that is acquired as R & D progresses. As more information is collected, the better able are decision-makers to estimate the likely success of alternative actions, and hence an R & D program is more likely to succeed if more options remain open as R & D proceeds. Hence:

Proposition U6: Assuming firm size does not affect decision-making procedures in a counter-acting fashion, large firms are more likely to be able to pursue suitable R & D strategies than small firms since the more diversified the research program of a firm, the more numerous are the other activities that might be profitably redirected on the basis of information acquired during an R & D project.

The defense of this proposition depends on the same kind of arguments with respect to the absence of adequate contingency markets that are the underpinning of the self-insurance notion expressed in U3.

The upshot of the investigations into the effects of uncertainty is somewhat more conclusive than that of the appropriability discussion, although not without its own dilemmas. Uncertainty is more likely to be dealt with optimally if firms are large enough so that their R & D programs are large compared to particular R & D projects, thereby allowing greater self-insurance and opportunity for responding to new information. At the same time, the threat of competitive innovation is an important factor in determining the willingness of the firm to accept R & D uncertainty; hence the effects of the minimum size requirements in U3 and U6 can be offset if they require that firms be so large that markets are not competitive.

Indivisibilities

The indivisibilities argument is an assertion about the empirical realities of the production of new ideas. It holds:

Proposition I1: In some significant number of cases, R & D activities exhibit sufficiently large scale economies that only large organizations can support efficient R & D operations.

These indivisibilities involve a number of effects other than the self-insurance and flexibility arguments made with respect to uncertainty.

For example, some advantage is said to be gained from having a group of scientific researchers in the same organization so that they can benefit from mutual interaction. Or, in the case of highly complex products, such as airplanes, developmental costs may be large compared to the total demand for the product.

At the same time, plausible explanantions have been offered for precisely the opposite phenomenon: that R & D has diseconomies of scale.

Proposition I2: The larger the organization, the greater the cost of maintaining adequate information flows and control of subordinate behavior, with the effect being larger the less amenable is the activity of the organization to routinization and rule-making.

Since in the presence of uncertainty activities must be flexible and decisions must be based upon a continuing stream of new information, R & D (especially basic research) activities would seem particularly unsuited to large organizations. Furthermore, the less well-defined the problem facing researchers, the greater the difficulty of carrying out the activity in large organizations. Thus:

Proposition I3: Smaller organizations are more likely to be an important source of advances in more basic research and of more revolutionary innovations, while large firms are more likely to be an important source of innovations at the more developmental end of the R & D continuum and of more incremental kinds of technical advances.

Empirical investigations provide some support for I3: firms of average and slightly above average size seem to do a disproportionately large share of innovating in many industries; rarely is the largest firm in an industry the source of major technical progress; and very small firms, often from outside the industry, frequently account for major research breakthroughs, although on average small firms commit proportionately fewer resources to R & D than do other firms.

These observations weaken the case for large firms made in the preceeding section on uncertainty. Propositions U3 and U6, in particular, depend upon assertions about the unimportance of I2 and I3.

Secondary Effects of Government Interventions

A relatively small part of government intervention into private markets is motivated by the desire to improve innovative performance. Yet all interventions alter the incentives faced by firms, and hence the rate and pattern of innovations they produce. Unfortunately, the study of this process is one of the most neglected areas of social science research. Nearly all of the systemmatic theoretical and empirical work is on one of three topics: the incidence of taxes, the behavior of the firm that is subjected to rate-of-return regulation, and the consequences of antitrust policies. Very little work has been undertaken on other aspects of regulation, such as standard-setting and licensing, or on governmental production and procurement.

Antitrust Policies. Commentary on the effects of antitrust activities on innovation follows essentially the lines discussed in the preceeding section with respect to the relationship between the size and market power of the firm and the propensity to innovate. If large firms that possess substantial market power are more innovative because they enjoy greater appropriability, more opportunity to capture scale economies, and a larger pool for self-insuring risk and uncertainty, than active antitrust activity, however beneficial in terms of static efficiency, reduces the rate of technical progress. But if technical competition, internal flexibility and a diversity of attacks on technical problems are more important factors in determining technical progress, then both static and dynamic efficiency are served by vigorous antitrust policy.

One other aspect of antitrust activity is its effect on an especially well-managed firm. A talented managerial group may adopt

policies that make its firm dominant in an industry by virtue of the superiority of the price/quality combination of its product. Once such a firm increases its market share to the point where antitrust action becomes a possibility, management will lose the incentive to make further improvements in the firm's market position. If the market dominance of the firm is due to especially creative R & D activities, the result is reduced technical progress by the leading firm:

Proposition S1: Antitrust activities reduce the incentive of firms with strong market positions to generate more rapid technical progress than is normal for firms in the industry.

Even dominated firms can find that the ceiling on firm size reduces R & D incentives:

Proposition S2: The ceiling on firm dominance limits the maximum potential profitability of revolutionary technical advances for all firms; and

Proposition S3: Limitations on the dominance of firms reduce technical competition, so that for dominated firms the potential profitability of engaging in no innovative activities is increased.

At the same time, antitrust policies can have positive effects on the incentive to innovate. As a dominant firm approaches maximum allowable size, the potential profitability of employing creative resources converges to the expected loss in profits that would come about if the resources moved to another firm and made that firm a more effective competitor. The incentive faced by dominated firms is greater, since the potential profitability of the resource would include gains captured from all other firms in the industry. Furthermore, the reduced expected rate of progress of the leading firm in some ways increases the potential profitability of R & D in other firms by reducing the chance that the R & D programs of the latter will produce less successful innovations than the program of the former. Hence:

Proposition S4: Antitrust policies increase the incentive of dominated firms relative to dominant firms to engage in innovative activity, and can increase the absolute incentive of the former if reduced technical threats from the latter are sufficiently valuable to offset the effects described in S2 and S3.

The preceeding discussion abstracts from the reasons for the dominance of the leading firm except insofar as it might arise from more rapid technical progress. Dominance can arise for reasons having nothing to do with technical performance, such as monopolization of an input market or purely pecuniary scale economies (such as advertising barriers to competition). The profitability of R & D for dominated firms in these kinds of markets is reduced to the extent that technical virtuosity is limited in its impact due to other advantages of the dominant firm. To the extent that antitrust policy limits or even eliminates these other advantages, the potential profitability of R & D to all firms, including the leading one, can be increased.

The most important point to remember in attempting to assess the impact of antitrust policies on innovative behavior is that the effects upon the leading firm do not necessarily imply anything conclusive about the effects upon other firms or innovative individuals. Even if antitrust policy reduces the incentive for dominant firms to innovate, as long as other firms believe that technical competition will still be pervasive antitrust policy can increase their incentive to innovate. And, from the point of view of innovative managers with less risk-aversion than typifies the industry, procompetitive policy actions will serve primarily to induce them to move from dominant to dominated firms, particularly if R & D is not subject to scale economies and if the dominance of leading firms arises for other than technological reasons.

Taxes. While few investigations have been undertaken of the effect of taxation on R & D, the literature on tax incidence deals with the problem of risky investments in sufficient detail that some of the effects on R & D can be inferred as special cases. The most obvious aspect of taxation is that it reduces the profitability and the risk of any form of investment, including investment in R & D. Even a "neutral" tax -- i.e. one that leaves unchanged the ranking of investments according to expected rate of return -- can affect the selection among investment alternatives by causing a reduction in the differences among them in riskiness. Consequently:

Proposition S5: An increase in taxation will make all investments less attractive, but it will improve the relative attractiveness of more risky investments.

Whether an increase in taxes leads to more or less R & D depends upon: (1) whether R & D is more or less risky than other types of investment and (2) whether the effect of the tax increase in narrowing risk differentials among investment alternatives offsets the effect of generally reduced after-tax profitability for all investments. A tax increase can increase R & D if R & D is, in general, both more profitable and more risky than other forms of investment and if the tax increase reduces the relative riskiness of R & D more than enough to offset the reduction in its absolute expected profitability.

The relevant issue for tax policy is more complicated than the general principle enunciated in S5, since it involves the selection from among numerous taxes the combination that will generate some target amount of revenue with the minimum attendant socially undesirable effects on economic incentives and income distribution. What policy makers need, in this context is a set of comparative statements about the impacts of various types of taxes on investment behavior, R & D and productivity advance. Unfortunately, the literature provides few insights into this issue.

Under current practice, R & D expenditures are, for tax purposes, normally treated as current expenses, as are training expenses (which conceptually are investments in human capital), whereas fixed capital investments are amortized over the useful life of the asset. Furthermore, the depreciation allowances for capital investments are, to some extent, earmarked for future replacement of the capital good, whereas new knowledge does not "wear out" in the sense that machinery does.

<u>Proposition S6</u>: If the tax treatment of R & D and investment were perfectly symmetrical in terms of their effects on incentives, R & D would be classified as a capital investment, and amortized as an intangible asset with a useful life that ends when an invention becomes obsolete.

To treat R & D expenses as current expenditures, not as investments, is to subsidize them relative to investments in fixed capital. Thus, an argument that, because of the theoretical problems mentioned above, economic efficiency requires a tax subsidy of R & D does not necessarily imply that subsidization should be increased, since the tax system already makes an investment in R & D of given gross profitability worth more in terms of net, after tax profits than a comparably profitable investment in fixed capital.

An increase in capital gains taxation serves to reduce the attractiveness of R & D relative to fixed capital investments. Since R & D expenditures are regarded as current expenses, any increase in the value of a firm that they cause is, if the knowledge or the firm is sold, taxed entirely as a capital gain. An investment in fixed capital that is depreciated more rapidly than the actual decline in its economic value will, if sold, be subject to income tax recapture of excess depreciation. Hence:

<u>Proposition S7</u>: The higher the capital gains tax rate, the less the value of the preferential treatment of R & D for tax purposes.

Finally, a few things can be said about specific tax rules that might be imposed to induce more R & D. Some advocates of tax subsidies for R & D have proposed a counterpart to the investment tax credit for expenditures on R & D. Some fixed percentage of R & D expenditures could then be deducted from tax liabilities, rather than from gross income before taxes, in calculating the total tax bill. These kinds of proposals have the benefit of being automatic, thereby avoiding bureaucratic decision-making processes except insofar as tax returns are audited; however they also have one large disadvantage:

<u>Proposition S8</u>: Automatic, universal tax subsidies are inefficient in that some of the tax subsidy inevitably goes for investments in R & D that would have taken place anyway without a subsidy.

One mechanism for avoiding this inefficiency is to apply the tax subsidy only to net increments in R & D spending. For example, the historical average annual rate of growth of R & D spending in real terms would be calculated, and a tax credit given to firms equal to some proportion of R & D expenditures representing an increase over that historical average growth rate. This would provide an especially strong incentive for R & D on the part of a firm that historically had done very little innovating, since nearly all of any substantial R & D program would, when initiated, receive the favorable tax treatment. But even this proposal does not escape a second fundamental problem of tax subsidy proposals:

<u>Proposition S9</u>: A firm must have sufficient after tax profits to utilize the tax subsidy system in order for such a program to affect its R & D decisions.

Any subsidization system based upon reducing income taxation is biased against small, new firms. In light of Proposition I3 (major innovations tend to come from small firms), this is an important shortcoming of tax subsidy proposals. It can only be overcome by imposing a "negative income tax" for corporations -- i.e., instituting a procedure whereby firms without sufficient profits to take full advantage of R & D tax subsidies would receive payments from the Treasury to make up for this shortcoming. Unfortunately, such a system would have some tendency to reward the inefficient as well as making life easier for the struggling new innovative firm.

Regulation. As used herein, regulation refers to two types of governmental control over business: the setting of prices or profit rates (public utility regulation) or the establishment of minimum performance criteria (standards regulation). Always the former and often the latter involve entry controls as well. To illustrate, the Federal Communications Commission has mainly public utility responsibilities with respect to telecommunications and mainly standards responsibilities with respect to broadcasting. It controls entry in both cases. These examples also illustrate the theoretical, rather than strictly realistic, dichotomy of regulatory responsibilities, since performance objectives are at least implicit in telecommunications decisions, and entry control in broadcasting is based in part on considerations of economic viability. Nevertheless, a useful conceptual distinction can be made between policies intended to compensate for external effects and information problems that cause failure of competitive markets (interference among broadcasters, social consequences of program content) and policies designed to limit the exercise of "natural" monopoly power.

With respect to public utility regulation, the principal proposition offered in the literature is the familiar Averch-Johnson hypothesis that monopolistic firms whose profits are limited to a fixed proportion of capital investment will use excessive capital in producing their outputs. Applied to technical change, this becomes:

Proposition S10: Rate-of-return regulation causes firms to engage in excessively capital-using innovation.

A related proposition has to do with the optimal strategy of a regulated firm with respect to depreciation. Because profits are a proportion of capital investment, firms have an incentive to depreciate the capital stock more slowly than would otherwise be the case. As a result:

Proposition S11: A regulated firm will replace old equipment more slowly than is optimal, leading to a suboptimally small embodiment of current knowledge in the existing capital stock.

In a sense, these propositions predict that innovation among regulated firms will be both too fast and too slow. There will be too much exploration of new ways to substitute capital for labor, but once a useful innovation is identified, it will be diffused too slowly through the capital stock of the firm. These notions are not necessarily contradictory: the first refers primarily to a bias in technical change, while the latter refers more to its rate -- and then with respect to the replacement of existing capital with new equipment embodying current knowledge.

A-J models are a particular example of cost-plus regulation. While telecommunications, power and airline firms tend to be subject to rate-of-return regulation, other major regulated industries -- most notably surface transportation -- formally have profits limited to some percentage of revenues or costs. While this type of regulation avoids the capital-intensivity bias of rate-of-return regulation, it can create its own kind of inefficiencies.

Proposition S12: If, using the most efficient methods, a monopolistic firm, facing cost-plus regulation and setting price equal to long-run average costs, finds that the equilibrium price and output is on the inelastic portion of its demand curve, it has an incentive to use less efficient production methods (e.g. to pad costs).

When the conditions of S12 hold, a firm can increase total revenues (and thereby total profits, which are a fixed proportion of total revenues) by selling less output at a higher price and a higher cost per unit of output. Of course, such a firm would not be particularly thrilled by the prospect of a cost-reducing innovation; however, it would be excessively motivated to adopt innovations that increase the quality of service, but at a greater cost. Conversely, if the demand curve is elastic at the most efficient regulatory equilibrium, the firm will have an extra incentive to adopt cost-reducing (and service-degrading) innovations.

Models of regulated firms that postulate profit maximization subject to a cost-based constraint on profits are a simplified characterization of the regulatory process, and probably should not be taken very seriously. They assume that a firm faces a profit constraint that is stable and known during its planning horizon, an assumption that is obviously unrealistic for four reasons. First, regulatory policies may change in unpredictable ways during the twenty to forty year planning horizon of the regulated utility because of unpredictable changes in the external environment (e.g. the energy crisis). Second, regulatory policies may be to some degree altered by strategic actions of the firm, in which case a decision by a firm must take account not just of its effect on profits given the regulatory rules, but also on the rules themselves. Third, the monopoly position of the firm is presumed to be given and unalterable, whereas much of public utility regulation has to do with deciding upon the optimal extent of competition and the identity of the firms permitted into a market. Fourth, the model presumes that a regulatory authority can, in fact, measure costs and profits with reasonable ease and accuracy, removing from consideration what is probably the single most time-consuming regulatory activity, the estimation of allowable costs and a "fair" rate of return.

This is not to say that the inefficiencies predicted by A-J and cost-plus models can not or do not exist. It is simply to say that investment decisions, whether in equipment or in knowledge, are

likely to reflect a far more complicated set of strategic considerations than the ceteris paribus effect on allowed profits.

Studies of the airline industry, in particular, are the most suggestive of the complexities of decisions by regulated firms. Airlines are to some extent competitive, and policy, cost and demand factors are changable over time. Research on the airline industry typically concludes that attempts to maintain prices above the competitive equilibrium result in the provision of excessive service, including too rapid adoption of new planes:

> Proposition S13: Regulated competition which maintains prices above competitive levels encourages excessively rapid innovation as firms substitute competition with respect to technical advance (among other things) for price competition.

In addition, the presence of some sort of cost-plus regulation can reduce the downside risk of failures, and the presence of competitive license awards can, when prices are too high, cause firms to be excessively innovative if to do so enhances their chances of being viewed favorably by regulators. With respect to airlines, all of these arguments have been used to support the contention that service innovations and advances in aircraft have come more rapidly than is economically efficient.

Organizational studies of regulatory bureaucracies emphasize still another feature of regulation: the slowness with which regulators make decisions. "Regulatory lag" is the term given to the observation that policy responses to changes in the regulatory environment are often significantly delayed. If an innovation can not be adopted without approval of regulators, regulatory lag delays it, thereby reducing its expected profitability, and thereby reducing the incentive to be innovative. (This argument is used to explain the lethargy of railroads with respect to innovation.) Innovations that can be adopted without regulatory approval can give the firm the opportunity to earn profits in excess of those allowed by regulators until the latter have responded to the cost and demand structure created by the innovation. (This argument has been advanced to explain the value of Bell Labs to AT & T.) If the duration of the regulatory lag is roughly commensurate with the rate of diffusion of innovation in a competitive industry, it could cause regulated firms to face the same incentives for innovation as do perfectly competitive firms.

These and other ideas about the efficiency of technical progress among regulated firms seem more fragmentary, contradictory and unrealistic than is typical of even the generally unsatisfying literature on unregulated firms. Virtually no theoretical work has been done on the extent to which the conclusions of the A-J type models depend upon their exclusion of the numerous other strategic and informational issues attendant to the regulatory process. Certainly such theoretical work should have a high priority. In addition, little attempt has been made to connect regulation to other policies to which regulated firms are subject, most notably taxation. For example, if the use of capital is taxed more heavily than the use of other productive resources, A-J effects might be offset within the context of the model.

With respect to standard setting regulation, the Averch-Johnson conception of regulators imposing constraints on firms has a clear counterpart in the literature about standards. Here regulators are assumed to impose a performance standard on the firm, which reacts accordingly to maximize profits.

One issue discussed in the literature, particularly with respect to environmental policy, deals with the problems of imposing new standards on the production of established goods. In comparison with taxation, standards are criticized because they provide no incentive to seek innovations that more than just satisfy the standard, whereas emissions taxes give firms a possibility of reducing tax costs by further innovation beyond the policy target implicit in the standards.

The other major issue with respect to standard setting pertains to the requirement that new products or facilities be given a seal of approval, as is the case with drug licensing and nuclear safety regulation. In these cases standards increase the cost of innovation. Their effect on the amount of innovation is not clear on a purely theoretical basis, since the higher costs tend to retard innovations but the standard setting process increases their appropriability by making competitive innovations and copying more expensive. Efficacy standards for new drugs have, in particular, been singled out as having a detrimental effect on consumers. They are said to constrain the introduction of much cheaper but slightly less efficacious drugs. In addition, because patents are said to take care of appropriability, they are alleged to reduce the profitability of even those new drugs that successfully treat a previously untreatable condition.

In some cases standards regulation prescribes which inputs and processes are to be used, as opposed to rules that establish minimum output standards. Such rules are quite common in the medical care sector, where certain services must be performed by specified occupational groups and where medical malpractice liability is often determined on the basis of the extent to which traditional and generally accepted practices were followed in treating a patient. This type of standard setting is, of course, anathema to innovation. For a new process to become legal in such a circumstance, the standards must be changed, and to change them not only involves waiting out the regulatory decision processes but als winning an adversary contest against those with a substantial personal stake in maintaining the primacy of the established method.

When the motivation for setting standards is an informational imperfection in the market, the alternatives to standards based on process, comparative safety and efficacy are twofold: (1) the requirement, with threat of serious penalty, to supply truthful and complete information about a product, coupled with free entry, and (2) comprehensive producer liability and antifraud policies. Examples of the former are the "truth in lending" law, statements about the nutritional content of breakfast cereals and warnings on cigarettes, while the medical malpractice suit is an example of the latter. If the person making the decision to use a potentially hazardous or ineffective product is assumed to be capable of responding rationally to complete information, the former strategy can be both inexpensive and successful. For example, the textile labeling requirements of the Federal Trade Commission are apparently successful in providing adequate warning at minimal cost to people with severe allergies. Resort to liability and damage lawsuits seems less promising, in part because it is an extremely expensive process and in part because of the uncertainty introduced by the tenuous connection between damage awards by juries and the actual damages suffered.

The preceeding discussion leads to the following conclusions about standard-setting regulation:

> Proposition S14: Economic incentives and, when externalities are not involved, information requirements are likely to generate more innovative behavior by regulated firms than are performance standards, and performance standards are likely to rank ahead of input or process standards.

POLICY IMPLICATIONS

The preceeding discussion obviously produces little in the way of direct implications with respect to the government actions that might improve the innovative performance of the American economy. On the other hand, it suggests numerous opportunities for further research into the innovative process and the consequences of existing government policies with respect to technological change.

Government Policy Towards R & D

The main lesson to be learned from the existing literature is that numerous factors are likely to affect the innovative behavior of firms and the extent to which innovative performance, when guided only by private market decisions, is likely to diverge from economic optimality. Since the conditions in which firms operate vary with respect to these factors, the only firm conclusion about generalized, economy-wide policy is a negative one: no universal policy covering firms in differing market and technological environments is likely to lead to an efficient rate and direction of technological innovation.

Each industry operates in a particular environment with respect to market competition, the opportunities for technical progress, the extent to which nonprofit institutions are advancing the science upon which the industry is based, the appropriability of new ideas, the expectations of each firm with respect to the liklihood of innovation in the industry, and numerous other factors that will affect its own program of research, development and market innovation. A rational government policy would be predicated on an investigation of the extent to which the conditions of particular industries were congenial to an acceptably efficient rate of technical progress. For each industry policy interventions would then be tailored to the specific aspects of its economic and scientific environment.

Although it is certainly beyond scientific proof given the existing development of knowledge, the most promising form of governmental intervention in most instances is probably through the award of grants and prizes for specific projects and accomplishments. In general, directly increasing the financial incentive for promising R & D is less likely to create inefficiencies than are programs designed to increase the monopoly power of the innovator. And financial incentives tied to specific projects, intended to tilt the scale slightly in favor of marginal projects, will avoid the expense of subsidizing activities that would take place anyway or that are not of high social priority. As a theoretical matter, the maximum potential impact of a given number of dollars spent by a grant-giving agency to subsidize specific research projects is greater than the maximum possible impact of a tax credit for R & D having an equal impact on federal revenues. The latter policy, even if provisions are made so that it supports only increments to R & D, will still not be efficient since it will not be focused on those industries and projects where innovation is most retarded by the various market imperfections discussed above. For example, because of its greater potential span of applicability, lesser appropriability and greater uncertainty, basic research probably should receive a larger proportion of subsidy than developmental research, but as a practical matter tax policy is unlikely to reflect this distinction.

The principal obstacles to capturing the theoretical advantage of a system of grants and prizes are the slowness and risk-aversion of bureaucratic decision-making processes. In a world in which politicians with 20-20 hindsight can easily point with alarm to risky research ventures that after the fact did not prove productive, executive bureaucrats are faced with an incentive to stick to the safe, sure proposal. Prizes for significant innovative accomplishments present less of a problem to bureaucrats by eliminating some of the need to identify ex ante which projects are likely to be successful; however prizes given for work at any stage of the innovative process other than the final market test will still run some risk of proving later to have gone for work that proved economically barren. Nevertheless, further exploration of the use of prizes for spurring innovative activity appears warranted.

Another mechanism that might improve both the speed and risk-taking of bureaucrats is to establish several governmental entities for promoting the same types of research. In essence, the argument in support of this proposal is similar to that in support of competition in private markets: that it will improve the incentives for efficient

behavior on the part of the agencies supporting R & D activities. For example, an alternative to the present structure for supporting research on new energy sources would be to establish several agencies that could award grants for energy R & D. Presumably there would be some incentive for an agency to try to be identified as the source of financing for a technological development that significantly increased the availability of economical energy, rather than strictly to avoid supporting research down blind alleys. If this conjecture is valid, then it is a mistake to collect nearly all developmental reseach programs with respect to energy technology into a single organization such as the Energy Research and Development Agency.

Finally, governmental programs for supporting research and development probably should be budgeted in part over several years, rather than on the conventional annual cycle. The annual budgeting process induces agencies to favor projects that can generate at least some short-term results which can be used to justify maintaining or increasing next year's budget. This generates a bias towards projects promising incremental gains in knowledge, and away from more fundamental and potentially more beneficial projects requiring several years of effort. Since one of the main implications of the literature on the innovative process is that the more basic the research the more likely that private market decisions will lead to too little of it, this consequence of the annual budgeting process is particularly unfortunate as it only reinforces an inefficiency in the allocation of research effort that is already likely to be present.

The R & D Assessment Program

There is obviously much room for further research along the lines supported by the R & D Assessment Program of the National Science Foundation. If, as is argued above, policy must be tailored to specific market and technical conditions in order to be efficient, a prerequisite to the development of rational policy is to delineate the conditions prevailing in major industries and how their performance diverges from that which is efficient. Thus, case studies of innovation in particular industries, focusing on the extent to which the problems delineated in this paper may be present, should have high priority.

Case studies on industries are likely to be more productive than cross-industry studies of the effects of policies. For one thing, studies that focus on policies require as an input data on the innovative performance of industries that, in most instances, simply are not yet available. Such data will only be forthcoming if a rather complete set of industry studies, with similar underlying objectives and methodologies, is forthcoming. Second, the state of theoretical understanding of the innovative process in general is so rudimentary that well-grounded empirical studies of the effects of policies are probably impossible at this juncture.

For example, the relationship of antitrust and patent policies to technological change depends in part on the relationship between market structure and the propensity to engage in innovative behavior. To shed light on the latter issue requires far better measures than we now have of the extent of competition in various industries. It also requires a workable model of how expectations are formed about the innovative behavior of other firms and the extent to which profitable R & D can be built upon current scientific knowledge. Only when this is accomplished can reliable estimates be made of the ceteris paribus effects of increased competition on technological change, and only then can some attempt be made to assess the effects of making markets more competitive through antitrust actions or reductions in patent protection.

Finally, the development of rational policy towards R & D depends critically on the elasticity of supply of innovative activity for the economy generally. The difficulties of estimating the extent to which productive R & D could be significantly increased in the aggregate

are surely immense, but the effort is worthwhile. The potential benefits of government programs to induce more innovation in lethargic industries will be greater if the resources attracted into R & D by such policies represent in some measure a net increment to the nation's innovative activities.

The scope of an inquiry into the supply of innovative activity must be broad, indeed. It must identify areas of economic activity that currently drain off significant numbers of talented people who might otherwise engage in innovative activity, and the extent to which the social productivity of their current activity may diverge from their earnings. For example, a major potential benefit of deregulation and no-fault automobile insurance could conceivably be their effects on the demand for lawyers and the possibilities thereby created for increasing the response of creative talent to increased opportunities in business and scientific activities more closely tied to innovative behavior.