THE EVALUATION OF TECHNOLOGY R&D: A CONTINUING DILEMMA

Peter deLeon

CSE Report No. 179
1981

CENTER FOR THE STUDY OF EVALUATION
Graduate School of Education
University of California, Los Angeles

The project presented or reported herein was performed pursuant to a grant from the National Institute of Education, Department of Education. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred.

ABSTRACT

The development of new and advanced technologies, especially those with significant potential social effects, needs to be assessed as part of an on-going process. This paper proposes a technology evaluation agenda for the next decade with special emphasis on four issues: appropriate technology R&D standards and criteria; the institutional context; comparisons to alternative options; and the social effects.

THE EVALUATION OF TECHNOLOGY R&D:

A CONTINUING DILEMMA*

by

Peter deLeon

Social Science Department

The Rand Corporation

INTRODUCTION

More than in any previous period, the American society is faced with a technological tidalwave of such breadth, depth, and (some would say) ferocity that it threatens to affect not only how we perform our everyday, workaday tasks but to fundamentally alter the very social structure and even how we conceptualize our future surroundings. Electronic transmissions will make the daily mail delivery a marvel of sophistication (if not reliability) as they have already changed the way we pay our debts;[1] solar energy technologies could undermine the great electricity grids that have been the basis of America's growth in productivity during the 20th century, a decentralization some claim

---

* Prepared for the Center for the Study of Evaluation, University of California, Los Angeles, as part of a series of papers proposing an evaluation agenda for the next decade.

[1] Kent W. Colton and Kenneth L. Kraemer, eds., Computers and Banking: Electronic Funds Transfer Systems and Public Policy (New York: Plenum Press, 1980).

could result in profound social changes;[2] and advances in computer technology and artificial intelligence address the very foundations of human thought and intelligence.[3]

Without implying that we all revert to Luddites, it is not unreasonable to suggest that these and other technological advances are, both taken in part and as a whole, more than a little frightening. Thoughtful citizens have the right, even the responsibility, to ask if nuclear power is worth the risk, if microwave ovens roast the chef as well as the casserole,[4] if supersonic transports will fatally pollute the heavens, if genetic engineering will cure cancer or breed automatons, or if the Green Revolution has merely postponed an inevitable war over food. The current and impending technological revolutions are not selective in their benefits or detriments; from the way we drive to work to the way we care for our sick, each of us is intimately affected. The key issue is no longer whether we intellectually and emotionally are prepared to manage these new technologies; Toffler argued that we were not, but this is surely debatable.[5] More to the fact, there is little choice. The challenge before us, then, is to suggest ways in which these new technologies and their effects upon society can be directed in paths most consonant with

---

[2] This assertion is most often voiced by solar energy advocates; see, inter alia, Amory B. Lovins, Soft Energy Paths: Towards a Durable Peace (Cambridge, Mass.: Ballinger for Friends of the Earth, 1977).

[3] Joseph Weizenbaum, Computer Power and Human Reason (San Francisco: W. H. Freeman, 1976), warns against the excesses of artificial intelligence.

[4] Paul Brodeur, The Zapping of America (New York: Norton, 1977).

[5] Alvin Toffler, Future Shock (New York: Random House, 1970).

traditional human values, most is keeping with and advancing accepted human aspirations.[6]

One might facilely suggest that the careful estimation or prediction of the effect of new technologies (in the jargon of systems analysis, their costs and benefits), would be the most straightforward means of alleviating these fears. There would be no fear of the unknown if it were, in truth, known. But this is hardly a feasible solution. The movement from general mechanical theory to technological R&D to later application is scarcely a predictable progression. Laws of physics do not promise benevolent applications to societal problems, even in the long run. Nor has medium-range technological forecasting offered much confidence in predicting technological breakthrough, costs, and usages that are of much value.[7] Even where careful pretesting has been mandated, later findings argue that man's ability to predict is severely circumscribed; witness the thalidomide tragedies and the continuing debate over the long-term effects of birth control pharmeceutics; the on-going and fractious arguments over the merits of the fast breeder reactor promise more rancor than resolution.

If our present abilities to predict the effects of various technologies are sorely limited, are other means available to assist us in choosing among alternative technologies and the futures they portend? The answer is less one couched in "yes" or "no" and more in "where" and "when." It must be acknowledged that our ability to peer far down the futures road becomes more clouded the further we attempt to see. This

---

[6] This is one of the basic charters of the policy sciences; see Harold D. Laswell, A Pre-View of Policy Sciences (New York: American Elsevier, 1971).

[7] For a review of approaches and difficulties, see Joseph P. Martino, Technological Forecasting for Decisionmaking (New York: American Elsevier, 1972).

is not to suggest that futures methodologies have no role in our professional mappings. Rather, it urges we realize that long-range predictions are accurate mostly by chance and that if we do mean to structure the future, we need to do so in much more modest, near-term time frames. This means that the long-term evaluation of predicted technologies is less likely to be useful than the evaluation of technologies as they enter, germinate, and later emerge from the research and development (R&D) phase of development. This reduced focus scarcely solves all our problems, as is argued below, but it does present a more discrete, manageable, and practical research agenda for those concerned with the interaction of technology with society, and particularly for those involved in shaping technology in real time through the mechanisms of public policy.[8]

Surely no one acquainted with the problems of evaluation could be misled into thinking that the evaluation of technological R&D is an easy task, particularly when those technologies have clear and forceful social implications. Yet two factors make such evaluations imperative. First, as outlined above, the march of technologies and their potential effect on the society which spawned them demands that we must gain some leverage on their development or risk being molded by them in ways not necessarily optimal for the society. This could be termed "Sorcerer's Apprentice" syndrome, except that the Sorcerer might not appear to rescue us, <u>deus ex machina</u>. The second reason is much more immediate and, from a public policy perspective, apparent. Virtually every field

---

[8] Peter deLeon, "Technology and Public Policy: Whither Side of Janus?" <u>Policy Sciences</u>, Vol. 11, No. 3 (February 1980), pp. 235-240, briefly sets forth some of these issues.

of research and development--social, physical, technological, whatever --is severely and increasingly resource constrained. The private sector is often and justifiably unwilling to support technological innovation where the possible applications are uncertain and the payoffs problematic. The government does not have the resources to develop and disseminate every "promising" idea. Public sector innovations in educational technology (e.g., teaching machines) are as closely monitored as energy technologies (e.g., solar-powered satellites), although the development processes and actors might be very different. This requires the evaluation of technology R&D throughout the R&D process; we literally cannot afford to await a market test for all the technologies under development in any given area; such a strategy would spread the R&D resource base too thin and thereby risk underattending the most promising technologies. Deprived of the traditional market tests for technologies, how can we best evaluate technologies still in the development stages?

This paper proposes means by which on-going technology developments can be evaluated in terms of specified objectives, both technical and social. As the title suggests, this is a continuing dilemma because during the R&D stages, by definition, we are unable to evaluate the end product. Indeed, we are often unable to identify what the objectives even are. For instance, have reading machines produced a more productive citizenry or advanced forms of DNA medical research saved more lives? We are therefore forced to render our best judgment on work in progress, always a chancy situation. The fundamental questions, then, are, can this be done and, if so, what are the most effective approaches? It is to these questions that this paper turns, but with one important caveat. Were the answers to these questions (and others that will be raised below) transparent, the evaluation community would be a much safer place to live. That they are not is what makes the evaluation exercise intellectually and professionally exciting (some would say risky). This paper does not pretend to present concrete answers. Rather, it will examine some methodological approaches, raise cautions when necessary, and, more specifically, propose research agenda and priorities for evaluating technological R&D over the coming decade. The thrust will be to juxtapose small steps with big ambitions. Much more at this time might be ill-founded and certainly presumptuous if one has any hopes of implementing such evaluation research and practices within the public policy arena.

THE EVALUATION OF TECHNOLOGY R&D

Four of the key underlying problems in evaluating the development of a new technology are:

-- Defining evaluation standards and criteria which are appropriate for a technology at a given stage of its development.

-- Identifying the institutional context into which the technology will be placed during the course of its development.

-- Comparing the technology to present and future alternative options; and

-- Amassing the means and evidence to measure with some confidence the effect of the emerging technology upon the society, especially its intended recipients (e.g., urban residents for computer dispatched emergency medical services).

Although these are listed separately, they are, of course, closely interrelated and need to be pursued simultaneously. For purposes of the present discussion, however, they will be treated individually.

It is rather important to recognize that these evaluation tasks are all <u>process</u> rather than end or product oriented. In assessing technology R&D, we are not able to judge the product of the R&D system until much too late in the process. Therefore, we must train our evaluative attentions on the component processes and modestly extrapolate from the evidence in hand to the next stage. For instance, if expected technological breakthroughs are not forthcoming, one can question the development's progress and predict that the forecast milestones will not be met. This evaluation, however accurate, may not be adequate for the decisionmaker. The R&D process is not conveniently linear; a development may lag, then take a quantum jump, unforeseen market conditions may intercede, or a delay may have been dictated by unreasonable scheduling earlier in the process. Government sponsorship of solar energy applications to facilitate what some call market penetration is illustrative. The somewhat paradoxical point to be stressed here is that the evaluation of technology R&D is very much a sequential, incremental activity, even if the R&D process itself is not. Evaluation attends to the relatively near-term objectives while--ideally--paying growing emphasis to long-term goals which can be matched against the technology development's objectives as the data become available. That this approach is more a model than a present reality should not prevent one from aspiring to it. This is hardly a casual aspiration, as will become apparent when we turn to the examination of the specific problems posed.

Finally, it should be noted that the evaluation of R&D technology should be a rather Darwinian exercise. In its normative form across multiple developments, R&D should be indifferent to the success or failure of its efforts. It is just as important to identify and terminate the failures (and perhaps learn from those lessons) as it is to recognize and expedite the successes. By the very nature of the uncertainties inherent in technology development, some technologies will be inadequate to their projected uses. The proper role of technology R&D evaluation is to identify these as early as possible, and thus permit the majority of R&D funds to support technologies which continue to display promise. This is, of course, an ideal situation. It assumes a consensual set of objectives and the ability to project technologies to those objectives, an unusual set of circumstances. Furthermore, the institutional involvements and organizational "sunk costs" often promote repetition and reduce technological Darwinism to relative ineffectiveness.[9] The unfettered results of technological competition are rarely observed, especially in those areas in which technology R&D is considered a "public good,"[10] or political investments are incurred.[11]

---

[9] See Michael H. Armacost, The Politics of Weapons Innovation: The Thor-Jupiter Controversy (New York: Columbia University Press, 1969), for one example.

[10] NASA's efforts to "sell" its space shuttle's capabilities and involve private industrial users is a good example; see Henry S. E. Cooper, "Shuttle-II," The New Yorker, Vol. 55, No. 52 (16 February 1981), pp. 65-113.

[11] An excellent example is provided by Melvin Webber, "The BART Experience--What Have We Learned?" The Public Interest, No. 45 (Fall 1976, pp. 79-106).

Appropriate Technology R&D Standards and Criteria

The working assumption of this paper is that R&D is a sequential, changing process, often with a dynamism of its own. Different actors with different objectives (usually corresponding but occasionally divergent) enter and exit the R&D process seemingly at will.[12] For this reason, no single criterion or standard is acceptable for evaluating the entire technology R&D process. Standards and criteria must be developed to fit the individual technologies at the various stages of their development. Measures of the progress of micro-circuitry would be inappropriate for evaluating advances in renal dialysis or fusion reactors. This, of course, makes generalization across the range of technologies almost impossible. What is more feasible is to generalize across certain limited classes of technology developments, such as alternative energy sources[13] or health care technologies. Furthermore, one can suggest that certain stages exist in the R&D process which might serve as the basis for additional generalization.[14] Still, it should be realized that the unpredictable nature of basic research, the broad range of technology, the changing

---

[12] This is documented in the development of the civilian atomic power plant by Peter deLeon, Development and Diffusion of the Nuclear Power Reactor: A Comparative Analysis (Cambridge, Mass.: Ballinger, 1979).

[13] There is of late an immense body of literature on this subject; see, among others, Roger Stobaugh and Daniel Yergin, eds., Energy Future (New York: Random House, 1978); Samuel H. Schurr et al., Energy in America's Future (Baltimore: Johns Hopkins Press for Resources for the Future, 1978); and Hans H. Landsberg (Ch.), Energy: The Next Twenty Years (Cambridge, Mass.: Ballinger for Resources for the Future and the Ford Foundation, 1979).

[14] For instance, see Walter S. Baer et al., "Government-Sponsored Demonstrations of New Technology," Science, Vol. 196, No. 4293 (27 May 1977), pp. 950-957.

set of actors in the R&D process, and the inability to agree upon a targeted set of objectives conspire to make a universal set of technology R&D standards and criteria a contemporary chimera whose pursuit could blind us to more reasonable and practical alternatives.

The principal problem with this seemingly unexceptional statement is the uncertainty inherent in technology R&D. Were it a more assured process, there would be little need for R&D evaluation, but uncertainties exist throughout the process. Technologies that seem inadequate at one point in time may experience a major breakthrough and leapfrog alternative approaches. The plutonium extraction process thought least promising in the early days of the Manhattan Project turned out to be the one ultimately used to produce the first atomic bombs;[15] conversely, the breeder reactor was thought to be the most immediately available form of atomic energy after the war.[16] Even as the technologies become more certain, problems with dissemination can add new and serious doubts. The study of technology innovation is populated with failure and characterized by dissention.[17] Therefore, one must be judicious about formulating hard and fast criteria for developments in which the bounds of uncertainty are significant.

---

[15] Leslie R. Groves, Now It Can Be Told: The Story of the Manhattan Project (New York: Harper and Row, 1962).

[16] Robert L. Perry et al., The Development and Commercialization of the Light Water Reactor, 1946-1976 (Santa Monica, Calif.: The Rand Corporation, R-2180-NSF, August 1977).

[17] See Robert Eyestone, "Confusion, Diffusion, and Innovation," American Political Science Review, Vol. 71, No. 2 (June 1977), pp. 441-448, and J. David Roessner, "Incentives to Innovate in Public and Private Organizations," Administration & Society, Vol. 9, No. 3 (Nov. 1977), pp. 341-365.

Clearly, the formulation of standards and criteria in technology R&D evaluation is not an easy task.

Nor is the task inconsequential in light of current (and rising) development costs. Most of the rules that are presently applied deal with the easiest, most quantifiable measures: schedule, costs, and projected performance.[18] The tradeoff between these three are most often found in military R&D; to reduce the almost endemic cost overruns, delivery schedules are extended and/or performance degradations are accepted.[19] Similar tradeoffs are found in civilian technology programs, and can be just as inhibiting, especially for those technology developments thought to have a large social impact.[20] The problem with these particular measures is that they do not reflect the changing stages of the R&D process; at best, they manifest measures most people can understand but with little agreement as to what they represent or portend in the overall development and ultimate worth of the technology. The American Petroleum Institute and the Solar Energy Industrial Association can possibly agree on the projected production costs of photovoltaic cells, but would be very far apart on assessing their

---

[18] Even these are not as concrete as they might appear; for instance, it is not always clear when a project begins or what the assumed cost inflation factors are; performance specifications or "deliverables" are sometimes written so loosely as to evade rigorous or even consensual definition.

[19] There is some evidence that uncertainties in production schedules and funding add substantially to unit costs. See Edmund Dews et al., _Acquisition Policy Effectiveness: Department of Defense Experience in the 1970s_ (Santa Monica, Calif., The Rand Corporation, R-2516-DR&E, October 1979), and G.K. Smith and E.T. Friedman, _An Analysis of Weapons Systems Acquisition Intervals Past and Present_ (Santa Monica, Calif.: The Rand Corporation, R-2605-DR&E/AF, Nov. 1980).

[20] For a representative set of case studies, see Walter S. Baer et al., _Analysis of Federally Funded Demonstration Projects: Supporting Case Studies_ (Santa Monica, Calif.: The Rand Corporation, R-1927-DOC, April 1976).

societal worth, or, in policy terms, estimating the amount of resources that should be devoted to reaching certain cost and performance goals. The message is unfortunate but hardly unexpected: the most readily available indices for evaluating technological R&D are not always the best given our expanded set of evaluative objectives.

These observations have certain practical implications for those attempting to evaluate on-going technology developments. The first recommendation is that the evaluation effort be initiated before the R&D. To begin to evaluate a project half-way through the R&D stage would force the evaluator to accept data and processes which might not be suitable to meet rigorous evaluation standards. If the evaluator can prepare an evaluation design during the earliest stages of the R&D, then the appropriate methodologies and data can be designed and designated for the various stages of the development. In other words, the evaluation (both in concept and practice) will fit the stage of the development. A second observation is that technology evaluation does not easily lend itself (if at all) to a single form of evaluation; what is good for education technology is not for communications satellites. Although approaches and methodologies may be roughly analogous, one should be cautious in assuming that similar measures and standards apply across a wide range of technology evaluations. A major metaevaluation effort should therefore be directed to developing congeries of technologies where similar evaluation frameworks are comparable, and thereby construct an experimental learning curve across the technology evaluation community. A third and final recommendation is that evaluators be alert to the changing priorities during the R&D process and design their approaches accordingly. This means that standards and metrics that are pertinent during the early R&D stages (such as the solution of basic engineering problems) become less important during the end stages (such as the reduction of production costs). This implies, once again, that there is no ubiquitous set of evaluation standards applicable throughout the technology R&D process.

In summary, the formulation of standards and criteria for technology developments is a function of two elements--the stage of the development and the type of the technology. The first is a variable, the second a constant. Together, they require that each technology development must define evaluation measurements and methodologies appropriate for that technology at a given stage of development. The criteria should be generalizable; how else might one compare and evaluate technologies directed towards the same end use (e.g., coal liquefaction vs. oil shale refraction)? The effort must be tailored to the particular technology development, for the development stages should be peculiar to each effort. Finally, the defined standards cannot be too strict or absolute, for rigidity (especially in the early stages of the development) would only undermine the uneven progress that is part and parcel of technology R&D.

The Institutional Context

Most of the research on evaluating technical R&D has been devoted to the technical nature of this process. For instance, earlier R&D evaluations have asked if a supersonic transport could fly over certain distances at a predicted speed, if a mechanized refuge collector could pick up so many tons of trash per day for a given price, or if solar collectors could reliably produce a certain temperature over a specified period of time. These perspectives have generally ignored the institutional constraints present in the development and dissemination of a technology. This is a serious oversight, one that recent literature on technological innovation has begun to correct.[21] However, this realization does not necessarily guarantee solution; overt and covert institutional objectives can influence a technology R&D program as profoundly the technical problems; Baer's case studies of federally-funded demonstration projects are replete with examples, and Morison's illustrations of the "not invented here" syndrome provide ample evidence of the institutional obstacles to technology development.[22]

These institutional constraints have obvious import for the technology developer, but little has been said in this context about the technology evaluator. The institutional problems in technology R&D must be considered as an integral part of the technology evaluation effort if we are to gain an accurate assessment of a technology's value. The evaluator cannot rely on strictly quantitative measures, such as costs or positions on the logistics curve (even assuming they could be accurately estimated), for these will not completely reflect or explain

---

[21] Two very different accounts of institutional problems obstructing the development and later dissemination of technologies are presented in Elting E. Morison, Men, Machines, and Modern Times (Cambridge, Mass.: The MIT Press, 1966), and W. Henry Lambright et al., Technology Transfers to Cities: Processes of Choice at the Local Level (Boulder, Colo.: Westview Press, 1979).

[22] Baer et al., Analysis of Federally Funded Demonstration Projects, Morison, Men, Machines, and Modern Times.

what is happening during the R&D process and, therefore, not completely reflect the merit or promise of the technology.

The institution is rarely a benign or even neutral observer. It can, however, be either a positive or a negative influence. Depending on its institutional bias, it could either accelerate or impede technologies. The U.S. Atomic Energy Commission (AEC) was a forceful advocate of nuclear energy during the 1950s and 1960s while sorely neglecting alternative energy forms; solar energy technologies had to rely on the largess of the National Science Foundation until well into the 1970s. Certainly different school districts are more receptive than others to innovations in education.[23]

The question is not whether the evaluation of technology R&D should encompass the institutional considerations which will surely shape the program. The challenge is how these factors can be included with any degree of methodological rigor. The evaluator must judge if the development is being conducted in a receptive institutional environment; again, the AEC was not noted for being evenhanded in its sponsorship of alternative energy sources.[24] The evaluator might wish to examine what has been described as a "technology delivery system," or the progression of institutional actors who are nominally responsible for a

---

[23] See the multiple volumes, principally authored by Paul Berman and Milbrey W. McLaughlin, reporting on the education Change Agents program; a summary of their findings is contained in Paul Berman et al., Federal Programs Supporting Educational Change; Vol. V: Executive Summary (Santa Monica, Calif.: The Rand Corporation, R-1589/5-HEW, April 1975).

[24] For a more recent example, see U.S. General Accounting Office, Management Problems Impede Success of DOE's Solar Energy Projects (Washington, D.C.: Government Printing Office, EMD-81-10, 22 December 1980).

technology's development and dissemination.[25] Are labor unions likely to be predisposed (either pro or con) to the emerging technology and are large interest groups involved with the development? A final question might concern the role of the government in a technology development; as soon as the notion of "public goods" enters the developmental sequence, a new set of actors and criteria is introduced which greatly complicates the institutional calculus. This last point raises a basic question for metaevaluation research: namely, what are the proper roles of the public and private sectors regarding technologies basically destined for the civilian marketplace? Although the answer to this question might not be essential for the evaluation of specific individual programs, it fundamentally underlies the evaluation effort for it affects the continued governmental sponsorship of the programs and how the private sector enters the development process.

Ignoring for a minute the methodological problems such evaluation questions present, the institutional considerations of a technology R&D evaluation force the practitioner away from the comfort of hard, numerical data and square onto the agony of political or business realities. Developments of modern technologies of the types envisioned here are rarely without political overtones or undercurrents. To evaluate them, the evaluator must directly confront this fact and be aware of the political games and guises that could subvert his or her

---

[25] The concept of the "technology delivery system" was advanced by Edward Wenk, Jr., "Technology Assessment in Public Policy: A New Instrument for Social Management of Technology," _Proceedings of the Institute of Electrical and Electronic Engineers_, Vol. 63, No. 3 (March 1975), pp. 371-379. For one application, see Arthur A. Ezra, "Technology Utilization: Incentives and Solar Energy," _Science_, Vol. 187, No. 4187 (28 February 1975), pp. 31-38.

efforts. For instance, an assessment might be commissioned with the underlying political motivation of terminating the program or results might be requested before the study is complete because of attending political exigencies.[26] In any case, the institutional agenda for R&D technology evaluation should be recognized for what it is--the insertion of politics into what the academic evaluator might have hoped was a strictly methodological exercise.

In brief, it is certain that the evaluation of large R&D programs must begin to include institutional factors, or, in other words, the political element. In a sense, this is nothing new. The politic evaluator has long recognized that his or her work always had political implications and would couch the evaluation design or report's recommendations to take such sensitivities into account. There is, however, a difference here, in that I am proposing that part of the metaevaluation research agenda for the coming years include an explicit acknowledgment of the institutional considerations inherent in large technology development programs and devise ways in which these can be formally incorporated into research and report methodologies. Failing this test might be tantamount to failing the relevancy test or, perhaps even worse, seeing the results of one's evaluations used in ways never intended. Certainly rigorous program evaluations should not be subject to the whims of political fortune, but they cannot afford to trade methodological perfection for policy irrelevancy. Again, the key

[26] A typology of political reasons for policy evaluation which generally fall outside the objectives of the evaluation community is enumerated by Edward A. Suchman, Evaluating Action Programs (Boston: Allyn & Bacon, 1972).

research issue is how to include the institutional considerations into a set of formal evaluation approaches.

Comparing Alternative Technologies

Evaluating a single R&D technology is, as we have seen, rarely an easy task. However, it is increasingly inadequate in the technology development process and a technology-intensive society. Again, the issue of scarce resources dictates that not all technical alternatives under development be permitted to take the final market test. Therefore, analysts are frequently being asked to evaluate and compare alternative technologies so that some can be accelerated while others terminated.

In some cases, this comparison of technologies might be relatively straightforward; reading machines or wide-body aircraft are rather comparable technological approaches to well-defined objectives, even if the inevitable tradeoffs vitiate a dominant solution. In other cases, however, where the objectives are more ambiguous or where radically different means to an objective are proposed, the evaluation and comparison of alternative technologies become much more difficult. Imagine, for instance, an objective centered around technological means for reducing American dependence on imported oil or, even more concretely, reducing the amount of home heating oil consumption. Several technology-intensive means can be proposed to this end--solar and water and space heating units, new ways to generate electricity (e.g., fluidized combustion beds), or enhanced energy conservation techniques--which need to be first evaluated individually and then against one another. This analysis is scarcely straightforward. Even the decision as to what unit of common measurement is used could bias the entire evaluation. Similar examples are easy to cite: technologies to ease mass urban transit problems (better cars vs. rapid transit systems); health care systems; communications networks (e.g., land lines vs. microwave towers vs. communication satellites); and advanced energy technologies (e.g., fusion power vs. solar power transmission satellites vs. magnetohydrodynamics).[27]

Implicitly, these and other technology alternatives are constantly being evaluated and compared against one another. The issue professional evaluators must face is whether the implicit comparison is methodologically adequate or sufficient by most public policy analysis standards. I propose that it is not but this assertion inserts additional, certainly unwanted complications into the already intractable evaluation of technology R&D. We have just seen how evaluation under conditions of technology uncertainty poses several serious methodological problems and tenuous results; at the very least, wide uncertainty bands exist around most evaluations of technologies still in their R&D stages. Now the evaluation community is being tasked to compare competing technologies at a given (or perhaps differing) stage of the R&D process when the variations around most evaluation evidence are great. And this assumes that there is sufficient agreement regarding the end use or objective of the technology that it can serve

[27] Linda Berry and Lois Martin Broufman, "Research Strategies for Evaluating the Adoption Potential of Energy Technologies," Policy Studies Journal, Vol. 9, No. 5 (Spring 1981), pp. 721-734, examine non-economic research approaches for comparing energy technologies.

as an accepted benchmark. These are not easy tasks even under the best of circumstances. When the reality of tradeoffs enters the evaluation picture, comparative evaluation with any semblance of methodological rigor or verifiable standards seemingly becomes almost impossible.

In the face of this rather bleak assessment, however, there seems to be little alternative but to accept the comparative gauntlet. Technologies are constantly being compared and winnowed during their R&D stages, sometimes for apparently very shortsighted, irrational reasons. The Soviet Union, the United States, and Great Britain each advanced a number of different nuclear power reactors well into the R&D process; the first two nations finally settled on two basic types of reactors for reasons that would make little sense to the professional technology evaluator while the British are still seeking to settle on a standard reactor.[28] If choices are to be made, and clearly choices need to be made in light of diminishing financial and industrial resources, then they should be made on the basis of sound technological evaluation standards and criteria rather than unarticulated comparisons, naive selection, or simple happenstance.

This mandate, if accepted, presents profound intellectual and professional challenges for the evaluation community. As argued above, the evaluation of R&D technologies requires that the evaluator specifically tailor or custom an evaluation to a given technology at a specified stage of development; this evaluation needs to be judged against a rather well-defined objective. In short, it must encompass

---

[28] These decisions are discussed in deLeon, _The Development and Diffusion of the Nuclear Power Reactor_.

many unique features of the technology under study and the R&D process. We are now asking that a number of these unique (by definition) evaluations be explicitly compared and contrasted. By any standard, this represents a formidable research agenda for both those involved in the daily evaluation practice and those interested in metaevaluation.

Certain necessary steps towards comparative evaluation of developing technologies are clear, such as the requirement to define a technology's end use or objective so that there is at least a common benchmark for comparison. Methodologically, the need for constructing technology typologies which are suitable for comparison is an essential first step, if for no other reason than that they can serve as a test for more complicated comparisons. And, of course, the key to valid assessments among technologies is accurate evaluations of the individual technologies, which returns us to many of the points raised in the previous sections. But, in this case, these evaluations of the individual technologies acquire a new requirement, for they must somehow be made compatible with the assessments of the competing technologies. Therefore, the technology evaluator must explicitly take the comparison element into account and design and conduct the methodology evaluation accordingly. Such assignments are increasingly present and pressing; the issue is how they can best be met within the constraints of professional and methodological standards.

Assessing Social Effects

Perhaps the most difficult task in the evaluation of technology R&D agenda is the identification of the possible social ramifications and effects of emerging technologies. As argued above, it is extremely difficult to assess with any confidence the final shape and proportions a developing technology will assume; to predict its position in the appropriate technological marketplace is even more problematic.[29] To go one step further, to forecast the social implications and effects of a new technology--even in the short run--certainly strains one's analytic credibility. The English Industrial Revolution arguably had a more profound effect on society than the French Revolution. Some claim that the current U.S. energy crisis and its technological remedies portend major social changes on the American society;[30] more basically, Toffler contends that the technological revolution taken as a whole fundamentally augers an entirely new American culture.[31] Even if one acknowledges a certain literary hyperbole, it is safe to assume that specified technology advances could have a significant effect on various elements of the population. Therefore, the conscientious technology evaluator must develop means to assess the social and political implications of these products of the technology R&D system before they occur.

In some cases, the possible societal effects of new technology could be rather limited in a direct causal sense, although more widespread in an indirect manner. For example, the move towards increased industrial automation could result in growing sectorial

---

[29] Morison, _Men, Machines, and Modern Times_, speaks to this condition.

[30] Denis Hayes, _Rays of Hope_ (New York: Norton, 1977), and Lovins, _Soft Energy Paths_.

[31] Toffler, _Future Shock_. For his more recent vision, see Alvin Toffler, _The Third Wave_ (New York: Norrow, 1980).

unemployment in (say) the automobile industry. Geographic areas like Detroit would be particularly and directly affected in a social sense. On a broader scale, automobiles produced on an automated production line might be safer and result in a lower number of traffic fatalities, a rather diffuse but tangible societal benefit; on the other hand, increased unemployment has certain economic and social costs. In other cases, the social costs and benefits could be causally linked and of a much larger magnitude; even the minor (in retrospect) problems encountered at the Three Mile Island nuclear power station generated a full-fledged public debate on the relative merits of nuclear energy vice the potential hazards.

The central questions to this evaluation issue are what is the technological context into which the new technology is most likely to fit, who are the groups most directly affected, and what are the potential benefits or costs of the technology to those groups? All of these, of course, are matters of degree and require careful forecasting. None represents an easy avenue of inquiry; forecasting is a notably unreliable craft, even where models and data seem to exist.[32] Although methodologies like risk analysis and technological forecasting are available, their applicability to these social research endeavors is questionable because of their distinct reluctance to include societal issues or variables. Simply, the state of the art in evaluating the social implications of new technologies is the most primitive among the four evaluation topics discussed. More than in any of the other areas,

---

[32] See William Ascher, Forecasting: An Appraisal for Policy-Makers and Planners (Baltimore: Johns Hopkins Press, 1978).

the critical questions have yet to be posed with any semblance of analytic rigor; conceptual frameworks and evidence are still largely matters of speculation; and appropriate methodologies and data still await development and application.[33]

If this is a pessimistic assessment, it is not meant to be one of undue despair. The potential for social mischief presented by some new technologies (many would cite DNA research but much more concrete examples, for instance computers and the threatened invasion of personal privacy, can be offered) is too apparent and foreboding for the evaluation community to permit this facet of technology R&D evaluation to continue unattended. Relatively discrete evaluation tasks can presently be identified, such as the need to specify the recipient or most affected groups or to develop psychological and/or sociological tools to predict the most immediate implications of the technologies for such groups (e.g., will only the wealthy benefit from solar residential applications?). This information would permit policymakers to devise programs to ward off possible social problems or anticipate previously unforeseen externalities arising from the new technologies. For instance, advances in medical science might make contraception and population control in the Third World nations a practical reality, but they could exacerbate the inability of these governments to care for their aged population which would lack the traditional family geriatric

[33] That Otis Dudley Duncan, "Social Forecasting - The State of the Art," The Public Interest, No. 17 (Fall 1969), pp. 88-118, is still one of the best essays on this subject, a full decade after it was published, supports this assessment. Also see Olaf Helmer, "On the Epistemology of the Inexact Science," Management Science, Vol. 6, No. 1 (1959), pp. 25-52.

support. Improvements in forecasting techniques are surely important, especially as one is asked to peer further down the time line.

The key to this approach is the realization that evaluation is presently not methodologically prepared for a quantum jump to evaluate accurately the social implications of new technologies. We should not delude ourselves or public policymakers into thinking that the mere recognition of the problem is tantamount to resolving it. Incremental methodological steps must be taken in this area for any substantive progress to be made.

In summary, more than any other area of evaluative research, the issues presented by the evaluation of social implications for advanced technologies present real problems and unlimited opportunities for metaevaluation research over the next decade. In an area in which speculation is rife and potentially lethal, great if cautious progress can be made. The metaevaluation research agenda for this topic demands a multidisciplinary approach, for the issues fall well outside the context of the narrow legal, engineering, sociological, economic and political scopes of investigation. Methodological rigor will be available only on rare occasion, for the necessary models (to say nothing of data) are still inchoate.[34]

---

[34] Although this is not always admitted; see D. H. Meadows et al., The Limits of Growth (New York: Universe Books for Potomac Associates, 1972), for an example of inadequate models and questionable data still managing to generate extensive public debate. A brief but cogent rejoinder is Carl Kaysen, "The Computer that Printed Out W*O*L*F*," Foreign Affairs, Vol. 50, No. 4 (July 1972), pp. 661-668, and H. S. D. Cole et al. (eds.), Models of Doom: A Critique to the Limits of Growth (New York: Universe Books, 1973).

The charter for metaevaluation tasks concerning the evaluation of the social implications of technologies currently in the planning and R&D stages is almost unlimited, both in magnitude and criticality. Three early tasks present themselves. The first is to specify a few relatively concrete areas for initial inquiry rather than attempt a premature holistic approach.[35] A second immediate task is to structure (or at least propose) a long-term research plan so that progress can be directed, accumulated, and measured. A third metaevaluation task might be the systematic institutionalization of multidisciplinary research across technology issue-areas, skills and capabilities which will be in great demand.

TECHNOLOGY AND EVALUATION: A NECESSARY UNION

This paper has taken a somewhat unorthodox view of evaluation and an untypical perspective on technology. The former has generally been characterized as a retrospective (as opposed to prospective) assessment of methodology. Evaluation has, almost by definition, treated past incidents; it has explicitly emphasized an historical rather than a predictive mode of analysis, even in matters of public policy and planning programs.[36] Technology has, until quite recently, been viewed as a beneficial, almost benevolent element of the society. Its developmental process and end product, although not easily predictable,

---

[35] Abraham Kaplan, The Conduct of Inquiry: Methodology for Behavioral Science (San Francisco: Chandler, 1964), p. 279.

[36] This emphasis is stressed in the recent EPA evaluation of DOE's Conservation and Solar energy programs. See U.S. Environmental Protection Agency, "Assessment of Program Evaluation," Review of the Department of Energy's Conservation and Solar Energy Programs (Washington, D.C.: Government Printing Office, EPA-60-81-001, January 1981).

have been seen as delivering a universal good. Therefore, technologists have rarely been asked to submit their product or process to careful scrutiny beyond schedule and cost accountings.

In many ways, these traditional outlooks are quite correct. One must, however, question if they are satisfactory for future exigencies given that newly-emerging technologies will have a growing--perhaps dominant--influence on our lives. The issue of adequacy is predicated on two assumptions: that evaluation has a viable and consensually-recognized role in public policymaking; and that technology is indeed Janus-like, that it could be either beneficial or dangerous in its social applications and ramifications.[37] If these are valid assumptions, then the traditional characterizations of evaluation and technology are no longer sufficient for public policy objectives and a necessary union must be proposed and consumated between the two. The purpose of that union is to predict during the R&D process with some confidence and methodological clarity the future effect of developing technologies on society. This calls for a sharp focus on the technology R&D stage, for it is there where the evaluations and the resulting policy choices must be made.

This proposal basically violates the concept of evaluation as a retrospective method of inquiry, for it asks the evaluator to deal with future contingencies. This is, perhaps, an unfair burden better left to one's more speculative futurist associates. I suggest, however, that there is scant alternative but for the evaluation community to accept

[37] deLeon, "Technology and Public Policy: Whither Side of Janus?"

the challenge if it wishes to remain active in the public policy arena or even intellectually invigorating. Three reasons motivate this conclusion. First, the problems potentially posed by advancing technologies cannot be alleviated by conventional forecasting methodologies; more stringent, methodologically-reliable approaches which can be developed by evaluation experts are needed. Second, the problems are now significant and will only grow in importance. For evaluators to ignore them because the community has predominantly been historical in its analytic modes would imply professional stasis or an unwillingness to meet new demands. And third is the necessary union: public policymakers will require an assessment of developing technology and will therefore need--and hopefully seek--the skills of the evaluation community. For the latter to appear intellectually unsuited or professionally impotent to meet the emerging task would force public policymakers to treat the problems as best they can, which in most cases means something like "by guess or by golly" speculation. My personal suspicion is that formal, prospective evaluation could do much better, especially if its practitioners consciously recognized the attendant issues and develop the requisite methodologies to inform the policymakers.[38]

One should not pretend this is an easy task or set of tasks, but it is important enough that considerable efforts over the next decade be devoted to it from both the perspective of specific evaluation

[38] Peter deLeon, Solar Energy Program Evaluation: An Introduction (Golden, Colo.: Solar Energy Research Institute, SERI/TR-51-294, Sept. 1979), is one attempt to reconcile the evaluation and policymaking communities.

methodologies (e.g., risk analysis) and more general metaevaluation exercises. Failure to accept this challenge will not only deprive the evaluation community of a critical new growth area (in both the professional and intellectual senses) but also isolate the discipline from a crucial set of public policy issues. Neither should be readily abandoned, even if it means the deliberate rejection of the traditional role and approach of public policy evaluation, for they would represent lost opportunities the analytic community could not easily regain. Furthermore, although with somewhat less confidence, I suggest that the insights to be possibly gained from the evaluation of technology R&D are of such importance to society that the public, acting through its policymakers, should be reluctant to permit evaluators to dodge these responsibilities. All parties could lose should rigorous, formalized program evaluation not be applied to technology R&D. The converse--that every party could win with the careful application of evaluation approaches--is scarcely guaranteed but the potential symbiosis is certainly more attractive.