3 /03

"A solution is a product in search of a problem": A History of Foundations and Evaluation Research

Peter Dobkin Hall
Hauser Center for Nonprofit Organizations
Kennedy School of Government
Harvard University

Writing in 1987, management scholars Rosabeth Moss Kanter and David V. Summers framed the essentially problematic nature of evaluation in nonprofit settings. Nonprofit organizations, they wrote

> have defined themselves not around their financial returns, but around their mission, or the services they offer. And services, of course, are notoriously intangible and difficult to measure. The clients receiving them and the professionals delivering them may make very different judgments about their quality, and donors may hold still another standard. And "doing good" is a matter of social values about which there may be little or no consensus. It is this factor, the centrality of social values over financial values, that complicates measurement for nonprofit organizations (1987, 154)

As ownerless organizations with fluid goals, unclear technologies, and multiple constituencies, the possibility of devising a single satisfactory way of evaluating performance of nonprofits seems difficult if not impossible. Invariably, as Peter Drucker wrote in 1968, any measurement of organizational performance is a value judgment:

> It may sound plausible to measure the effectiveness of a mental hospital by how well its beds -- a scarce and expensive commodity -- are utilized. Yet a study of the mental hospitals of the Veteran's Administration brought out that this yardstick leads to mental patients being kept in the hospital -- which, therapeutically, is about the worst thing that can be done to them. Clearly, however, lack of utilization, that is, empty beds, would also not be the right yardstick. How then does one measure whether a mental hospital is doing a good job within the wretched limits of our knowledge of mental disease? (Drucker 1968, 196-197)

These difficulties are compounded by the fact that, as ownerless organizations, it is unclear to whom the accountability inherent in evaluation is due: To donors -- to demonstrate the efficient and effective use of their contributions? To clients -- to vouch for quality of service? To trustees and directors -- to show the competence of managers? To government agencies and other institutional funders -- to give evidence that the effectiveness of the organization's programs? To the general public -- to prove the value of the subsidy offered through tax exemption and deductibility of donations? Each of these constituencies, as Kanter and Summers suggest, is likely to prefer a different performance standard. And none, except an organization's governing board (which is the organization in a legal sense) and the Attorney General (who represents the interests of the public) has legal standing to demand accountability. In contrast, stockholders, customers, voters, and legislators can demand -- and in some instances have the authority and legal standing to require -- accountability from business firms and government agencies.

If nonprofits are generically difficult to evaluate, grant making foundations pose the knottiest problems of all. Often framed to serve broad and barely specified purposes (like the Rockefeller Foundation's mandate to "serve mankind," their trustees have enormous discretion to define and change their goals and purposes. Unlike universities and hospitals, they are not engaged in selling goods and services of any kind -- so there are no clients or customers to whom they must be accountable. And, since they are not in competition with other firms, they have no obligation to offer themselves or their activities for comparison with other grant makers. Because foundations' main business is managing their endowments and distributing its income, they neither seek support from individual donors, other grant makers, or government agencies, nor, for that reason, have any obligation to account to them except with regard to compliance with tax, regulatory, and fiduciary obligations.

Although in many instances established by individuals and families, donors cannot hold foundations to account for the use of their gifts. Once a charitable gift is completed, donors have no legal standing to demand that their wishes be followed. While donors can seek to increase the likelihood that the foundations they establish remain faithful to their wishes, by serving on their boards and appointing as trustees family members and representatives, the law discourages donor control by subjecting such entities to more searching regulatory scrutiny and by reducing the tax benefits that accrue from charitable gifts made under such conditions. In any event, no matter how carefully donors may proceeded in their efforts to ensure fidelity to their intentions, foundations are generally perpetuities and, with the passage of time and the procession of generations, their is no assurance that a donor's descendants will remain faithful to his original intentions.

Businesses exist to fulfill their charter purposes and to maximize returns on capital. Government agencies exist to fulfill the legislative mandates on which they were founded. Nonprofit producers of goods and services exist to fulfill their missions. All of these entities, depending on the industry in which they are located, are likely to be subject to professional or industry standards. For example, all hospitals, whether public, proprietary, or nonprofit, must be licensed and must operate in accordance with pertinent government regulations. But most grant making foundations operate under no comparable public or private constraints. Beyond compliance to a generic set of fiduciary and regulatory obligations, they are free to define their goals and purposes as they see fit with no substantive accountability to anyone.

Why, then, would foundations want or need to subject themselves and their grantees to evaluation? -- a procedure that even its proponents concede is expensive and time-consuming, and onerous and that its critics argue yields results of questionable value. This essay will offer an answer to that question by exploring the history of grant making foundations and their use of evaluation over the past half-century. It will suggest, first, that the willingness of foundations to use evaluation techniques stems from the political milieu of the decades following the second World War -- an era in which an increasingly tax sensitive public was given to episodic outbursts of concern about loopholes available to wealthy individuals and institutions. Secondly, it will explore the peculiarities of foundations as complex organizations, pointing to function of evaluation in settings in which decisions are made under conditions of ambiguity. Thirdly, it will look to the development of evaluation techniques and the academic professionals and consulting firms that championed their use.

Foundations under Fire, 1953-1969

Foundations have always kindled public and legislative suspicion. John D. Rockefeller's efforts to obtain a congressional charter for the Rockefeller in 1910 set off a firestorm of protest from conservatives and liberals alike. A few years later, in the high tide of the Progressive Era, the Walsh Commission, which was created by Congress to investigate industrial relations, turned its attention to foundations to investigate the possibility that their creators used their considerable assets to manipulate the economy and their grants to influence public opinion.

Because foundations were not established in large numbers until after the second World War, however, legislative interest in their activities did not manifest itself with any real force until the late 1940s, when Congress began a comprehensive overhaul of the Internal Revenue Code. This overhaul was necessary because America's assumption of wartime and post-war leadership required new governmental capacities, including the creation of tools of social and economic management.

Steps had been taken in this direction as early as 1935, when Congress imposed steeply progressive taxes on personal and corporate income and estates. The effect of this legislation was largely symbolic, since fewer than ten percent of Americans earned enough annual income to be subject to the tax. This changed in 1942, when, to finance the war effort, Congress made virtually all incomes subject to taxation and ordered employers to withhold estimated taxes owed from all pay checks. In other legislation, in order to prevent war profiteering, it imposed taxes on the excess income of corporations. Suddenly, tax and estate planning -- devising legal ways of lowering in individual or corporation's tax liability became a big business. The exemption of charitable corporations from federal corporate income taxes, which became part of the tax code in 19__ and the deductibility of individual and corporate donations (which became law in 1913 and 1935), were quickly recognized as major loopholes that could enable individuals and businesses to lower their taxes.

Charities were virtually unregulated before the 1950s. While the Internal Revenue Service had procedures for registering organizations as charitable tax-exempt entities, there was little incentive for any charities but the largest charities to apply for charitable status since only those had donors wealthy enough to need to use the charitable deduction. Today charitable tax exempt status frees charities of state and local tax liabilities, but before the late 1940s, few states imposed sales or other taxes affecting corporations. Because they didn't pay taxes, the IRS took little interest in charities and had no enforcement or oversight mechanisms in place to supervise their activities. This made them especially attractive to tax planners and their clients. In the years following the war, the number of foundations skyrocketed.

While having ostensible charitable purposes, some of these, on closer examination, were little more than tax shelters. Among the more notorious schemes was the _____ Foundation, established by New England textile tycoon Royal Little, who turned it into a tax free holding company for his industrial empire. Another was the New York University macaroni monopoly, under which the university's law school acquired the largest noodle manufacturer in the world and claimed it as an exempt charity because all of its profits were devoted to educational purposes. Still another was the Ford Foundation, established as a part of automobile pioneer Henry Ford's estate plan. By the terms of Ford's will, the company, which had been privately held, was reorganized as a joint-stock company with two classes of stock. The voting stock was distributed to family members. The non-voting stock was given to a new charitable entity, the Ford Foundation. Because of the tax savings resulting from the establishment of the foundation, the vast Ford fortune passed to his descendants without paying a penny in estate taxes.

None of these practices was illegal at the time. No laws compelled charities to make pay outs to beneficiaries, restricted excess business holdings, prevented the operation of unrelated businesses, or curtailed donor control. But to average citizens chafing under the unaccustomed burden of income taxation, it seemed patently wrong for the country's wealthiest individuals and largest corporations to avoid paying their fair share. It was not long before Congress turned its attention to laws affecting "foundations and other exempt entities," revealing a long string of abuses.

As Congress was concluding its work on the landmark 1954 Internal Revenue Code, another set of issues focused public and legislative attention on foundations. Rising tensions with the Soviet Union and its allies -- the Cold War -- fueled concerns about communist subversion. Foundations became suspect in the search for subverives not only because Alger Hiss, an executive with the Carnegie Endowment for International Peace, was a principal in a high-profile spy case, but because a number of foundations had made grants or supported programs that some congressmen deemed suspect. In quick succession, two select House committees conducted widely publicized investigations into foundations. While they came away relatively unscathed. the more far-seeing foundation leaders knew that they would likely be subject to future bouts of regulatory and investigatory enthusiasm.

Chief among these were the officials of the Russell Sage Foundation. Founded in 1907, Sage was the first of the modern grant making foundations. Although it operated under a broad mandate to devote its resources to "the improvement of social and living conditions in the United States of America," by the 1940s it had become particularly interested in the tracking the development of philanthropy and the proliferation of foundations. It sponsored the first directories of foundations (the ancestors of today's standard compendium, *The Foundation Directory*), the first of which appeared in 1920. Beginning in 1946. it initiated a series of major analytic publications identifying foundation trends and practices (Andrews & Harrison 1946; Andrews 1950, 1952, 1953 1956a, 1956b, 1958). Sage was the moving force in organized the 1956 Princeton Conference Conference on the History of Philanthropy, which sought to stimulate scholarly interest in organized charity and philanthropy (Russell Sage 1956). It was a leader in the effort to establish the Foundation Center, an enterprise devoted to collecting and disseminating information about foundations. The Foundation Center began issuing the *Foundation Directory* in 1960 and initiated publication of the periodical, *Foundation News* (which is now published by the Council on Foundations).

Clearly Sage had more in mind than disinterested information gathering and analysis of foundations. Its concern with promoting transparency, accountability, and best practices as defenses against abuses and public exposure was evident from the contents of the earliest issues of *Foundation News.*

The efforts of Sage and a handful of other major grant makers could not stem the public's growing skepticism about the value of foundations and its suspicion of the motives of their founders and executives. Through the 1960s, Congress and the Department of Treasury focused attention on questionable foundation practices and priorities. Congressional hostility culminated in the hearings held by the House Ways and Means Committee and the Senate Finance Committee on the 1969 Tax Reform Act, where the leaders of the nation's major foundations failed to provide convincing arguments for perpetuating the tax benefits and minimally regulated status that these institutions had enjoyed. The Tax Reform Act included stringent reporting requirements, limitations on donor control and excess business holdings, and provisions requiring that foundations pay out a specified percentage of their annual revenues.

Some philanthropic leaders, notably John D. Rockefeller 3rd, shared Congress's doubts about the practices and performance of foundations. While the Tax Reform Act was still being debated, he convened a national task force, the Commission on Foundations and Private Philanthropy, chaired by industrialist Pete Peterson, to study " all relevant matters bearing on foundations and private philanthropy" in order to make "long term policy recommendations" (Commission 1970). Among the questions the commission addressed was one of the most fundamental:

> It can be agreed that foundations have contributed to our society in past decades. But this does not answer the questions that are rightly asked about the future. With the future in mind do the benefits of such tax-favored institutions outweigh the risks of abuse? Do foundations have

> advantages over other means for promoting the general welfare? (Commission 1970, 117).

While the Commission did not find it possible to make a cost-benefit analysis of tax revenues foregone by government versus the social benefits produced by foundations it did suggest that they could be assessed both quantitatively and qualitatively. Among its final recommendations, it included a section on the "evaluation of programs -- government and non-government" which suggested the value of "impartial appraisals" not only of government policies and programs, but also of other institutions -- including foundations (Commission 1970, 129-130). In addition to the concrete benefit that such assessments would produce for program planners and policy makers, "periodical appraisals" of foundation performance, "shared with the American public. . . would help reduce some of the mystery from foundation activities and convey a sensitivity to the public that had not always been present" (Commission 1970, 130-131). The panel also made specific suggestions about the ways in which " more thoughtful and consistent monitoring and evaluation of foundation grants could not only reduce the possibly of activities by grantees that reflected unfavorably on grant makers, but also for improving the quality and effectiveness of grant-funded projects (Commission 1970, 132-133).

Though the Peterson Commission's report had no impact on the substance of the Tax Reform Act and little discernable influence on foundation themselves, it undoubtedly encouraged the staff at Russell Sage to begin investigating the potential of evaluation techniques and their applicability to foundations and their activities. As an institution whose primary focus was social welfare policy and which worked more closely with academic social scientists than most other grant makers, Sage was uniquely well situated to tap into the emerging field of evaluation research (Glenn, Brandt, & Andrews 1947; Hammack & Wheeler 1994).

Evaluation of one kind or another had an historic association with social reform movements. As early as the mid-nineteenth century, public health advocates and physicians promoting "moral treatment" of the mentally ill used statistical methods to demonstrate the efficacy of the interventions they espoused. The Civil War era reports of the United States Sanitary Commission, the private body to which government had entrusted responsibility for military sanitation and relief, were replete with statistics vindicating the value of professionally-administered programs. In the decades following the war, state charities commissions and private charity organization societies used statistics to target problems and prove the economic and social benefits of "scientific philanthropy" -- charitable activity intended to attend to the root causes of social problems rather than merely providing relief to their victims (Frederickson 1965).

As a product of the effort to reform private charity and government social welfare policy, the Russell Sage Foundation promoted this approach to identifying and solving social problems. So, in the wake of the passage of the 1969 Tax Reform Act and the publication of the report of the Peterson Commission, it is hardly surprising that it should be the first foundation to give serious attention to evaluation. In 1973, sociologist Orville G. Brim, Jr., who served as president of the foundation from 1964 to 1972, contributed a forthright essay titled "Do We Know What We Are Doing" to an edited volume, <u>The Future of Foundations</u> published by the American Assembly, one of the era's most prestigious public policy groups (Heimann 1973). In it, he not only sharply criticized foundations for the uninformed quality of their decision making and their isolation from the realities of public life, but also provided a detailed account of Sage's own development and use of evaluation. These were framed by a reiteration of the foundation's overall concern with the future of foundation philanthropy. The public, Brim wrote,

> through its institutions, has a stake in the success or failure of foundations. The Tax Reform Act of 1969 challenged foundations to defend their position as contributors to pluralism, socially and politically, in a way distinguished from those activities of government and business (Brim 1973, 220).

Noting that over the previous two decades there had been "substantial expressions of public interest as to whether the public was getting its money's worth from the tax exemption granted to private foundations," Brim argued that "knowledge gained from evaluation" would have "more than just personal interest to foundation administrators." It might "help them to sustain the institution in American life" (220).

Brim set forth five categories of information that might serve those purposes: assessments of the success or failure of specific projects by foundation executives, grant recipients, and affected constituencies; administrative comparisons of similar projects; inter-program comparisons intended to assess areas of activity for foundation program planners; comparisons of foundations in terms of their relative success in targeting problems and dealing with them effectively; and comparisons of foundations' goals and activities with "those of other institutions, public and private, pursing the public good" (222).

Brim went on to give an overview of the basic concepts of evaluation research as they existed at the time.[1] "Evaluation research," he wrote

> is the application of social research to provide the administrator with accurate information on the consequences of his action. The main aim is to measure the benefits received from the program, particularly with repect to objectives that have been set by the program, and to relate these to the cost incurred. More specifically, evaluative research must determine the extent to which a program achieves its goals, the relativ empact of key program variables, and the role of the program itself versus external variables or influences (226).

Citing Edward A. Suchman, whose pioneering monograph, Evaluative Research (1967), Sage had funded and published, Brim distinguished the "impressionistic evaluative procedures" used by administrators, politicians, and journalists" from "hard-headed" and "specific" assessments based on

> (1) extensive examination of objectives of a program (including uderlying assumptions); (2) development of measurable criteria specifically related to these objectives; (3) establishment of controlled situations to determine the extent to which these objectives and negative side effects are achieved" (226-227).

Before turning to Sage's evaluation efforts, Brim gave his own assessment of the "state of foundation evaluations." Noting the "continuing exhortation" from both within and without the foundation field for evaluation, he found that there was little real interest in or understanding of formal evaluation. "The administrators of foundations seem trapped in foundation lore about how decisions should be made," he wrote. "Much of the traditional wisdom consists of unexamined maxims which have come to be rather comfortable substitutes for hard-headed evaluative information" (228). He called attention to the handful of serious evaluation efforts: Suchman & Ricker's study of the Maurice Falk Medical Fund, Eugene Struckoff's study of the New Haven Foundation, Winston Franklin's evaluation of the Kettering Foundation, and evaluations of specific programs at the Rockefeller, Sloan, Ford, and Markle foundations (230).

Evaluation efforts at Russell Sage consisted of two major initiatives. The first was a study of projects supported during the 1960s. The second was a year-long analysis and critique of the foundation's work as a whole conducted by "several radical sociologists" (231). The project evaluation was conducted by an outsider, CUNY sociology professor Lindsey Churchill (an ex-Sage staffer), assisted by Brim and two other foundation employees. The projects evaluated included 110 research projects, 11 research and training programs, 10 fellowship programs, 16 dissemination projects (support for publications and conferences), and support for 15 visiting scholars. Data considered included documentary information about projects and grant recipients from foundation files and a survey project directors in which they were asked to assess the impact of their own Sage-funded projects. As Brim describes it, the study seems to have raised more questions than it answered. The evaluators came up with nicely quantified ratings, but they weren't sure what they meant:

> if we find no strong positive or negative correlations between the various measures of success and the characteristics of projects and methods of operation, then we must consider at least three instructive reasons for the lack of findings. First, the cause may lie in unreliable or invalid ratings of success. Second, we may not have in our batter of potential predictors the truly significant factors that influence differential success of outcomes -- we may not have the information in our records, and we did not think to ask for it. Or, third, perhaps both the ratings of success and the predictive factors are satisfactory and the reason there is no significant relationship between project outcome and our antecedent predictive characteristics of the project or administration is that success and failure are much too

> dependent on chance factors, that is, a number of unsystematically distributed effects on the course of a given progam that cannot be anticipated. We can speculate about what this wold imply for guidelines of 'rational' foundation management (238).

It was one thing, as Brim discovered, to charge into evaluating complex settings and organizations with "hard-headed" and "specific" positivist assumptions. It was quite another thing -- regardless of how quantifiable the results -- to come up with meaningful or reliable findings that could provide any real guidance to foundation decision-makers.

. Though Brim concluded his essay by endorsing continuing development of evaluation in foundation settings and with a long laundry list of ways foundations could inform themselves about their work and how the public perceived it, his failure to demonstrate persuasively that evaluation could be applied effectively to foundation activities may account for why the second major national panel on nonprofits, the Filer Commission (the 1974-1979 Commission on Private Philanthropy and Public Needs, also organized at the behest of John D. Rockefeller 3rd, had very little to say about the subject -- and why evaluation was not docketed as a major concern of academically-based nonprofits research that began to emerge after 1978.[2] It appears that by reframing foundations as part of a more inclusive domain of organizations -- the "nonprofit sector --, grant makers had enfolded themselves in the warm blanket of public approval enjoyed by other charities (on this, see Karl 1987). Having done so, foundations no longer felt the need to prove their right to exist.

The Great Society, the Conservative Revolution, and the Growth of Evaluation Research.

Had the foundation world been a major patron of evaluation research, the field would have died stillborn. As it happened, in these waning days of the Great Society, the same factors that had led Congress to question the public benefit produced by foundations also fueled demands that federal social programs demonstrate their effectiveness. Tax sensitive Americans wanted to know whether their tax dollars were being well spent. Brim noted this fact in his essay, referring to the role of the Office of Management and Budget, the General Accounting Office, and Congress in insisting "that agencies produce real evidence of the effectiveness of their programs" (Brim 1973, 239). Major Great Society legislation, including the Economic Opportunity Act of 1964 and Elementary and Secondary Education Act of 1965, both required and provided funding for evaluation. According to Brim, "federal expenditures for evaluation research in fiscal 1970 provided approximately $45 to 50 million in the general areas of social welfare and social services, housing, education, etc., with the largest portion being expended by the Department of Health, Education, and Welfare" (239).

The establishment of publications like <u>evaluation: a forum for human services decision-makers</u> in 1972 affirm the fact that the public not the private sector was the real source of energy and innovation in evaluation research. The magazine was produced by the Minneapolis Medical Research Foundation, with funds from the National Institute of Mental Health. Its first issue featured a lead article by Minnesota Senator Walter Mondale titled "Social Accounting, Evaluation, and the Future of Human Services" and an interview with Nixon administration HEW secretary Elliot Richardson. In its lead editorial, "What Is Evaluation Good For?" the journal's editor, Susan Salasin, set forth its belief in the possibilities of evaluation. "In every article or discussion about evaluation these days," she wrote,

> one question, implicitly or explicitly, gets raised. "What good is it?" Some decry the current methods available for use, other bemoan the fact that findings are not utilized, and still others recoil from the insights evaluations have provided regarding the inadequacy of human service program (Salasin 1972, 2).

More fundamental than the question of the effectiveness of evaluation, she argued, was the question of our expectations of the procedure.

> If instead we try to ask ourselves what evaluation has given us, aside from a lack of answers, we can see a new role for evaluation. Instead of talking about what evaluation doesn't do, we can talk about what it does do. The experiences reported herein seem to indicate that evaluation does lead to a discovery of what some of the problems with human service programs are. Evaluation does highlight the fact that we think we can accomplish

> more for people than we can, in fact, accomplish. Viewed from this perspective, evaluation is a valuable and effective tool that can help sense out the problems in our programs, and keep us honest in making promises to people about how we can help.

In defining evaluation "as a tool for problem identification and realistic sight-setting" and as "a social movement dedicated to finding out how well be are doing with our human services programs, and discovering better ways of doing," Salasin was clearly endeavoring to embrace evaluation as an instrument for maintaining public commitment to human services. In doing so, evaluation seemed to clearly place itself in the service welfare state liberalism. At the same time, her editorial suggested that because of the huge political and financial stakes involved in how evaluation was used, by whom, and for what purposes, the field and its methods were bound to develop not in the disinterested remoteness of the Ivory Tower, but in the politically charged and turbulent environment of an America that was becoming increasingly polarized over what, if anything, the public owed to its young, poor, and disabled citizens.

Problematic findings, contested methodology, and political conflict do not appear to have been obstacles to the growth of evaluation research. The continuing expansion of the field was signaled by the establishment of two new journals in 1978. Evaluation and Program Planning - An International Journal, based in the Department of Mental Health Sciences at Philadelphia's Hahnemann Medical College and Hospital, boasted an editorial board that included economists, sociologists, psychologists, psychiatrists, accountants, specialists in criminal justice, state and federal officials, and management consultants (though, significantly, no one from the foundation world). The lead editorial in its first issue proclaimed its editors' belief that evaluation was more than an academic field -- it was an emerging profession. " Evaluation and Program Planning," they wrote

> is predicated on the principle that the techniques and methods of evaluation and planning transcend the boundaries of specific fields, and that an interdisciplinary forum of communication is necessary. Thus, articles will be published which explore the evaluation and planning of programs in a wide range of areas: public health, mental health, organizational development, social service, public action, mental

> retardation, drug abuse, corrections, criminal justice, personnel, and education ("To the Reader" 1978).

The two lead articles in its initial issue suggested that one of the primary goals of the editors was the affirmation of professional identity. Both, "Evaluation: Manifestations of a New Field" and "The Development of Evaluatino as a Profession: Current Status and Some Predictions," were written by health Philadelphia care managers Eugenie Walsh Flaherty and Jonathan A. Morrell.

While acknowledging that the term "evaluation" was used "to describe a growing body of efforts sharing certain common characteristics," they dodged setting forth any definition of evaluation, claiming that it was "premature to seek definitions in this still evolving field" (Flaherty & Morrell 1978, 1). Instead, they looked at "the empirical growth of whatever it is that people call evaluation" in order to identify "the pieces which may fit into a conceptual framework in the future." They began by outlining three "characteristics" of evaluation: "the role of the scientific method, the goal-attainment approach versus the systems model, and the use of evaluation information" (2). The role of the scientific method was contested, one camp embracing positivistic standards of experimental research that, to its critics, seemed not only inapplicable to real world situations, but also, to the extent that its findings were useful at all, applied only to decision making after the completion of a project.

The second characteristic involved debates over "the kind and timing of information which evaluation should provide: the goal attainment approach and formative evaluation." Those favoring the former approach, summative evaluation, sought findings that would guide decisions about renewing support for completed projects based on their success in achieving predetermined goals. Others favored formative evaluation, a far more contextualized approach

> founded on the belief that the users of evaluation information are best served by (1) a comprehensive description of how the program actually functioned and (2) feedback about the attainment of program objectives, about unexpected results of program efforts, and about the relationship between program functions and the attainment of objectives at continuous

> points during the program's course, rather than after program termination" (2).

Formative evaluation was intended to guide on-going management intervention and decision making by project administrators -- in contrast to summative evaluation, which was primarily intended to serve the needs of funders.

The third characteristic "about which evaluators differ" involved the use of evaluation information. One camp believed that evaluation information not generated for the purpose of guiding decision making was not evaluation -- it was just research. The other camp took a less utilitarian view.

Consistent with their desire to promote the professionalization of evaluation, Flaherty and Morrell conjured up a history for the field.[3] To their credit, they avoided overly grand claims, locating the origins of the field in the recent past, in the work of psychologist Kurt Lewin and in pioneering educational and welfare progam evaluations. They also found evidence of extensive evaluation activity in the 1950s in the federally-funded mental health field. It was only in the 1970s, however that books, articles, and academic programs began to proliferate. They attributed this to political pressures on publicly-funded programs, a growing desire for relevance among social scientists, academic resource scarcity, and "an expansion in the social science methodologies appropriate for research in applied setting" (3). What they seemed to be describing was the emergence of a field of practice that had no clearly defined conceptual core or generally accepted methods that was both developing in response to existing funding opportunities (evaluation mandates in federal programs) and was at the same time creating new opportunities by identifying new problems to investigate and new methods with which to investigate them.

In the second article, Morrell and Flaherty sought to assess the progress of evaluation in terms of existing models of professionalization. Discounting the lack of consensus in the field over concepts, definitions, and mehtods, they focused on the development of the knowledge base, on perceptions of the legitimacy of evaluation as an enterprise, and on the politics of the field. "Clearly there is a good deal of prestige, power, and income involved in the evaluation enterprise," they conceded (Morrell & Flaherty 1978, 14). The institutional challenges involved how to codify the knowledge, certify expertise, and exercise control over who performed evaluation. They viewed the establishment of evaluator accreditation programs in states like New York as a promising sign that evaluation was professionalizing.

The journal's preoccupation with professionalization -- manifested by an almost obsessive introspection -- continued in future issues, with such articles as Conner & Dickman's "Professionalization of Evaluation research: Conflict as a Sign of Health" (1979), Dunn, Mitroff & Deutsch's "The Obsolesence of Evaluation Research (1981), and Nick L. Smith's "Classic Articles in Educational Evaluation" (1981). At the same time, the practical concerns of professional entrepreneurship were addressed by pieces like Joan E. Silber's "Purchasing and Selling Evaluation" (1981), an article that candidly addressed the situation of evaluators as individuals who need "not only the creative, critical, and technical competencies of a scientist, but also the ability to work responsibly within a political and business context" (173). Professionalized or not, evaluation had clearly become a big business which was attracting both academics and business consultants.

In 1978, the first genuine academic journal devoted to evaluation, <u>Evaluation Quarterly</u> appeared. Published by Sage Publications, a respected publisher of journals in the applied social sciences, it was edited by University of California-Santa Barbara sociologist Richard A. Berk and Howard E. Freeman, an individual whose career interconnected the various domains out of which evaluation was emerging. After receiving his doctorate in sociology at New York University, Freeman worked at the Rand Corporation (a major evaluator of government programs), taught at Harvard, Brandeis, and UCLA, and served as consultant to the Ford, Russell Sage, and Robert Wood Johnson foundations -- the three most energetic supporters of evaluation in the grant making world. The forty-three member editorial board was an assemblage of the best and brightest academicians and "pracademicians" including executives from the Robert Wood Johnson Foundation, the Urban Institute, the U.S. Office of Education, and the National Opinion Research Center, and senior professors from Harvard, Columbia, Johns Hopkins, Stanford, Northwestern, Indiana, and other first rank research universities. Like the other journals, the editors represented the broad range of social disciplines, including anthropology, economics, education, public administgration, psychology, psychiatry, sociology, and urban planning.

The lead article in the first issue of the journal was an assessment of evaluation research by Peter H. Rossi and Sonia R. Wright. Rossi, who had taught at Harvard, Johns Hopkins, the University of Chicago, and the University of Massachusetts, was one of the most productive and admired scholars in his field. Author of a dozen books, he served as editor of the three major sociology journals and would soon be elected president of the American Sociological Association. Rossi and Wright reviewed the growth of the field and took note of what they regarded as the most important evaluation efforts: the 1969 evaluation of Head Start, a major national compensatory education program by Ohio University and the Westinghouse Learning Corporation; the famous Coleman Report, an assessment of the availability of educational opportunities and the impact of school quality on educational achievement; the 1972 evaluations of federal employment training programs; the Educational Testing Service's 1971 evaluation of the impact of Sesame Street on children's cognitive skills; as well as several smaller efforts. Rossi and Wright concluded that while some of the evaluation efforts were exemplified best practice, most fell far short. "The current state of the art is, on the average, considerably below the level illustrated here." In closing, they affirmed the value of evaluation research and called attention to the need for improvements in the field.

Quite clearly, by the 1980s evaluation research had established itself as a credible enterprise. Despite continuing disagreement over objectives, concepts, and methods, perhaps best exemplified in Guba and Lincoln's polemic <u>Fourth Generation Evaluation</u> (1989), it had moved from the margins of applied social science to full institutionalization (if not legitimacy), with its own academic journals, courses, and training programs. As this occurred, it began to attract the attention of foundation executives and trade associations.

Robert Wood Johnson Pioneers Foundation Evaluation, 1972-1980

There can be no doubting that the impact of the Robert Wood Johnson Foundation's commitment to evaluation on nonprofits' growing interest in the field during the 1980s. Established by General Robert Wood Johnson, head of the medical supply giant, the foundation, when it began operating in 1972, was the second largest in the United States (Nielsen 1985, 121; Rogers 1987). The foundation's trustees were given broad discretion to define its mission and programs. Approaching the responsibilities of philanthropy from a business perspective, they determined "that the Johnson Foundation should be 'productive': that it have clear goals and priorities, meet its deadlines, show results, and stay out of trouble (Nielsen 1985, 123). Working with a staff of physicians and health policy experts headed by Dr. David Rogers, they committed themselves to an activist reform-oriented agenda:

> It was agreed that the foundation's grants would not be used for such activities as funding cancer research or building wings on hospitals, but would go to projects to improve the availability of primary health care. It would not be concerned with training surgeons in open-heart procedures but family doctors in general practice. It would help set up not new biomedical laboratories but group practices to coax doctors back to inner-city and rural areas. it would fund research not on gallbladder removal procedures but into ways to organize health-care delivery more equitably and efficiently (Nielsen 1985, 124).

This orientation to experimental/"field trial" interventions was coupled with a fundamental commitment to "objective, third-party evaluations of the results of these actual pilot models" (125).

An account of Johnson's use of evaluation research was the subject of the first article in <u>Evaluation and Program Planning</u> to address the question of evaluation research in the foundation field, an article co-authored by foundation president David Rogers, staffers Linda Aiken and Robert Blendon, and UCLA sociologist Howard E. Freeman. The article was a notable departure from the navel-gazing, self-congratulation, and polemics that characterized much of the evaluation literature. And it was especially significant because of the care its authors took in highlighting the unrealized potential of evaluation in the field of grant making.

The article begins by quoting foundation critic Waldemar Nielsen's assertion that "not one-tenth of foundation grants have any measurable impact upon the major social problems confronting the nation" (1972). It went on to refer to the views of John W. Gardner, former head of the Carnegie Corporation and to quote the Peterson Commission's recommendations regarding the need for evaluation of grant making.

The article took pains to distinguish what Robert Wood Johnson was doing from "'evaluation' as the term is typically used in the foundation world" (Rogers, Aiken, & Blendon 1980, 120). "Many of the larger foundations 'evaluate' their grants," they wrote.

> In this sense, evaluation is commonly employed to describe the general process of determining how well a foundation-funded project was implemented by a grant recipient. While such an assessment implies an explicit and rational basis for judging the "worth" of an effort, it usually doesnot involve the marshalling of objective facts and evidence to impartially document whether the project improved the lives of people, or the performance of an organization in carrying out its mission and reaching its objectives. Without being pejorative, they are usually impressionistic judgments based on unsystematic information about what went on during the period of the grant (120).

In contrast, Robert Wood Johnson's approach involved "the use of social research procedures, including systematic collection of factual data and use to appropriate analytical methods to arrive at a reproducible conclusion from which a decision can be made regarding success or failure in implementation and outcome of a particular program." These efforts had two foci: one on the "extent of the problems it is addressing (who is affected, to what extent, and for how long)"; the other, using monitoring and impact studies, on the extent to which "projects significantly affect health care problems in measurable ways."

Since there was no reliable data on problems relating to access to health care, the foundation began its work by supporting a series of baseline studies which sought to measure "the extent to which persons in the U.S. were experiencing problems in gaining access to desired medical care, and to what degree this was attributable to the absence of health resources or to imbalances in the mix of physician specialists and other health care providers." These would serve as the yardsticks for assessing the impact of foundation programs. Through a national survey of households, the foundation discovered that while the majority of those questioned found medical care to be reasonably accessible, some 24 million Americans -- poor Hispanics in the Southwest, certain inner city poor, and poor people in the rural South -- had the greatest difficulty in obtaining medical care. The findings suggested that "highly targeted programs designed to specifically reach these groups" rather than a national effort would be the most effective. They also suggested that such programs were "likely to require long-term subsidies of public and private funds" (123).

The second baseline study focused on the question of the balance between primary and specialized medical care as it related to health care access. Its findings suggested a need for "increasing the number of generalist physicians" and encouraging greater "specialist participation in primary care." Accordingly, the Foundation made major investments in "special training programs for generalist physicians" and in shifting the focus of specialized residency programs from hospital-based subspecialty care to general ambulatory care. In addition, it underwrote the creation of a new specialty called "Family Practice."

As described, these evaluation activities appear to have been formative rather than summative -- producing on-going insights into programs in ways that permitted changes of direction and emphasis as programs developed. This proved to be crucially important because it enabled the foundation to respond to unexpected changes in the areas in which they had intervened. When, for example, federal funds for training minority physicians were slashed in the mid-1970s, the foundation was able to make a strategic shift in its grant making from awards to medical schools' minority recruitment programs to central agencies that funded scholarships for particular students. This enabled the foundation to get more "bang for its buck," supporting larger numbers of students without having to significantly increase its investment in training.

The foundation drew primarily on academia for its evaluation expertise, using evaluators from University of Chicago's Center for Health Administrative Studies, the University of Southern California, and UCLA's Institute for Social Science Research. It also used such non-academic bodies as the Rand Corporation and the National Academy of Sciences.

While Robert Wood Johnson remained strongly committed to evaluation, reporting that it had more than twenty such efforts underway as of 1980, it was also candid about the problematic nature of the enterprise. "As our evaluation program has evolved," Rogers and his co-authors wrote,

> the reasons such studies are rarely undertaken by foundations has become increasingly evident to us. The answer to the question: Should a private foundation make major investments to "evaluate" their programs?, is not as clear as the Peterson Commission report might suggest. Evaluations are expensive and require extraordinary care in their design and implementation if one is to answer generally significant questions. The tools used in assessment of complex programs involving many people and institutions are i mperfect. Data collection is enormously time-consuming, and such studies require continuing attention and patience to keep them on track (127).

Because the cost of a multi-site evaluation often exceeded a million dollars, such efforts were clearly beyond the funding capacity of most foundations at a time when only 50 of the nation's 21,000 grant makers had annual expenditures exceeding $4 million. "Thus," they concluded, "the cost of just one major independent evaluation study of three- or four-grant projects exceeds the grant-making capacity of 95% of the nation's private foundations" (127). Evaluation efforts also required "trained research staff who spend their time almost exclusively on the design and monitoring of such programs" -- at a time when only 2% of foundations had professional staffs. Finally, unlike Johnson's programs, which were designed to produce measurable outcomes, most foundation initiatives were "difficult to evaluate objectively" (128). While concluding that evaluation had served it well, the article closed by declaring that the costs of evaluation were "likely to be prohibitive for 99% of the nation's foundations" (128).

Evaluation and Organizational Behavior

Looking back on his accomplishments as president of the Robert Wood Johnson Foundation on his retirement in 1987, David Rogers affirmed his belief in the value of evaluation in an article in Foundation News, a magazine published by the Council on Foundation, the major grant makers' trade association. Interestingly, rather than suggesting that evaluation had produced reliable assessments of the impact of its programs and projects, Rogers indicated that its major contribution was organizational. By focusing on the outcome of grants rather than on the process by which "we, or our grantees, might get there," evaluation permitted the foundation's staff -- a "diverse lot of individuals of widely differing value systems and political persuasions" -- to agree on its "major programmatic thrust" (Rogers 1987, 49). This revealing admission illuminates an aspect of evaluation that few evaluation researchers at the time were willing to acknowledge: that rather than serving the positivist goal of producing "objective" measurements of the impact of foundation interventions, its primary value was to reduce uncertainty and decensus within grant making organizations.

Carol H. Weiss, the Columbia University-based evaluation research pioneer, had called attention to this issue as early as 1973 in a cautionary article in Evaluation titled "Where Politics and Evaluation Research Meet." While "evaluation research is a rational enterprise," she wrote, it "takes place in a political context" (Weiss 1973, 37). She went on to outline the three major ways in which politics "intruded" on evaluation. First, she suggested, "the policies and programs with which evaluation deals are the creatures of political decisions" -- "proposed, defined, debated, enacted, and funded through political processes." Second, because evaluation was undertaken to inform decision-making by political actors, findings inevitably had political consequences.

> Third, and perhaps least recognized, evaluation itself has a political stance. By its very nature, it makes implicit political statements about such

> issues as the problematic nature of some programs and the unchallengeability of others, the legitimacy of program goals and program strategies, the utility of strategies of incremental reform, and even the appropriate role of the social scientist in policy and program formation (37).

"Only when the evaluator has insight into the interests and motivations of other actors in the system," she wrote

> into the roles that he himself is consciously or inadvertently playing, the obstacles and opportunities that impinge on the evaluative effort, and the limitations and possibilities for putting the results to work -- only with sensitivity to the politics of evaluation research -- can the evaluator be as creative and strategically useful as he should be (38).

Weiss was keenly aware of the likelihood that, in these political settings, evaluation might have little to do with the ostensible goal of measuring success or failure of programs. Policymakers, she wrote

> are members of a policy-making system that has its own values and its own rules. Their model of the system, its boundaries and pivotal components, goes far beyond concern with program effectiveness. Their decisions are rooted in all the complexities of the democratic decision-making process: the allocation of power and authority, the development of coalitions, and the trade-offs with interest groups, professional guilds, and salient publics. How well a program is doing may be less important than the position of the congressional committee chairman, the political clout of its supporters, or other demands on the budget. A considerable amount of ineffectiveness may be tolerated if a program fits well with prevailing values, if it satisfies voters, or if it pays off political debts (40).

Because of this, she noted, evaluation studies seemed to have "little noticeable effect on the making and remaking of public policy."

The one situation in which evaluation research seemed most likely to affect decisions was when the researcher accepted the values, assumptions, and objectives of the decision-maker. "This means, obviously," Weiss noted, "that decision-makers heed and use results that come out the way they want them to. This suggested, as David Rogers would later conclude in weighing the impact of the Johnson Foundation's efforts, that evaluation had more to do with organizations' political and psychological equilibrium than with a truly disinterested search for objective measurements of impact. The very willingness of social scientists to evaluate programs, she observed, "gives an aura of legitimacy to the enterprise" (41).

Ultimately, Weiss concluded, "for the social scientist who wants to contribute to the improvement of social programming, there may be more effective routs at this point than through evaluation research" -- most particularly "research on the social processes and institutional structures that sustain the problems of society and closer social science involvement in the application of that research" (45).

Given the evident shortcomings of evaluation research, particularly as it applied to foundations, it is curious that, by the mid-1980s, interest in the subject began to intensify. If publications are any evidence, evaluators themselves had shown little interest in foundations since the Russell Sage and Robert Wood Johnson initiatives of the early 1970s. But in 1985, Nick L. Smith of the Northwest Regional Education Laboratory, published the results of a National Institute of Education study of foundation support of evaluation. At the same time, articles on evaluation began appearing with increasing frequency in foundation trade publications and in nonprofit scholarly journals.

The Smith article, which appeared in <u>Evaluation Review</u> in the spring of 1985, began by calling attention to the fact that "resources usually available for evaluation have begun to dwindle" as the federal government, now in the full grips of the conservative revolution, was "encouraging increased private sector support for social programs. Evaluation researchers are looking more widely for financial support," he noted, "and are asking whether foundations might be a source of evaluation funds" (Smith 1985, 215). With federal funds drying up, evaluation researchers were looking for new markets and, Smith suggested, "foundations are a multi-billion dollar a year enterprise with vast potential for contributing to the improvement of evaluation methods and practice" (216).

While Smith offered his readers a good overview of the types and sizes of foundations and the kinds of evaluations they had used, he curiously omitted any reference to the *caveats* that David Rogers and his colleagues from Robert Wood Johnson had issued five years earlier in their assessment of the foundation's own evaluation efforts. The emphasis in the article was clearly on the opportunities that foundation evaluation offered, not the obstacles it presented -- including, most notably, the small number of foundations that could afford to support evaluation research, the small number of staffed foundations, and the problems of evaluating outcomes that eluded quanitification.

Of the nation's 450 major foundations, Smith found only 205 evaluation grants awarded by 76 grant makers. Most of these were by a handful of foundations: Ford (22), the Lilly Endowment (14), Kellogg (12), Robert Wood Johnson (9), the Cleveland Foundation (8), Charles Stewart Mott (6), and Markle (6). The mean size of awards ranged from a high of $338,026 for the nine Robert Wood Johnson grants to $34,793 for the 8 awards by the Cleveland Foundation. Evaluation grants and contracts went overwhelmingly to research universities, though one consulting firm -- Television Audience Assessment -- was noted as a major awardee.

Smith's findings were not particularly encouraging. He noted that, relative to the resources available, foundation use of evaluation research was minimal and that most of the evaluation that had been done had produced only internal documents, not any material that was publicly available. (This reinforces the notion that foundations used evaluation to address internal conflicts rather than to generate public support for their work). Smith's article ended with an epilogue instructing potential applicants for foundation support on how to do grants research and how to approach foundations. "It takes effort," he noted, "to match the interests of an evaluation researcher with the proper foundation."

It can't have been an accident that, within months of the publication of the Smith article in Evaluation Review, a major feature on evaluation, "Getting to Know You: Post-Grant Evaluation Is Enabling Foundations to Discover Many Techniques that Help the Grantee -- and the Grantmaker -- Achieve Common Goals" appeared in Foundation News, the bi-monthly magazine of the ouncil on

Foundations. Its author was evaluation researcher Martha G. Butt, vice president of the Northwest Area Foundation, a regional grant maker whose activities covered the states of Minnesota, the Dakotas, Montana, Idaho, Oregon, and Washington.

This regional philanthropic subculture had been unusually active in promoting interest in evaluation. Mary Stewart Hall, philanthropic advisor to the Seattle-based Weyerhauser family and director of the Weyerhauser's corporate giving programs, had both funded evaluations and promoted them in a series of speeches to groups like the Conference Board and articles in edited volumes on corporate philanthropy (Hall 1977, 1979a, 1979b,1982). In the 1970s, Humphrey Doermann, head of the Minneapolis-based had funded evaluations. And, of course, <u>evaluation</u>, the first of the journals devoted to evaluation research, was produced by the Minneapolis Medical Research Foundation. Butt's article referred to evaluation activities by the Model Cities Health Center and the Northwest Area Foundation, both based in St. Paul, the Seattle-based Weyerhauser Company Foundation, and the Bush and Jerome foundations, both based in Minneapolis.

Butt's article offered an enthusiastic endorsement and a detailed overview of the types and uses of evaluation. She recommended the use of independent consultants who could provide "the highly valuable characteristics of objectivity, autonomy, and credibility" (29). Although the piece included cautionary sidebars on the limitations of evaluation, such as Robert Bothwell's urging that funders consider grantees as "partners, not 'second-class citizens,' in philanthropic endeavors," the thrust of the article was positive. It was clear an effort to build a market for evaluation researchers in the foundation world. Clearly, evaluation was a front-burner issue for corporate and foundation grantmakers in the Northwest. And, as one of the nation's centers of philanthropic innovation, the interests of northwestern grant makers was bound to have a national impact.

Ubiquity

Beginning in the late 1980s, evaluation began to become ubiquitous. It was energetically promoted by the major philanthropy trade groups, the Council on Foundations and INDEPENDENT SECTOR, which published evaluation resource guides and hortatory memoranda and ran evaluation sessions at their annual meetings. Nearly every year after 1987, Foundation News featured at least one major article on evaluation and the industry press -- Nonprofit World (published by the practitioner-oriented Nonprofit Management Association), the Chronicle of Philanthropy, Philanthropy, and Nonprofit Times between them ran a total of 21 articles on evaluation between 1987 and 2002. The nonprofit scholarly journals -- Nonprofit & Voluntary Sector Quarterly, Nonprofit Management & Leadership, and Voluntas -- which had ignored evaluation before 1990 -- began publishing on the subject. Between 1995 and 2002, 15 articles on evaluation appeared in these journals, most of them by scholars who had not previously written on the subject. Articles on evaluation in trade and scholarly publications touched on virtually every industry in which nonprofits operated -- the arts and culture, education, health care, and human services.

Why did interest in evaluation grow so rapidly in the 1990s? Had the shortcomings of evaluation which foundation leaders like Orville Brim and David Rogers and evaluation researchers like Carol Weiss had set forth in the 1970s and 1980s been overcome by significant improvements in technique? Had there been a transformation in some fundamental aspect of the nonprofit sector which led organizations embrace a problematic and expensive activity that it had once generally scorned? Or were there forces at work involving the internal dynamics of nonprofits, particularly the professionalization of nonprofit management?

It may be that all three of these factors have helped to encourage foundations and other nonprofits to accept evaluation as a part of their work. Carol Weiss believes that evaluation has come a long way from the naive positivism of its early days and that evaluation methods are far more adept at addressing qualitative issues than they were. Certainly her own contributions to a richly contextualized approach to evaluation and her role in helping evaluators appreciate the complexity of complex organizations have helped the field progress.[4] The ability to engage qualitative issues undoubtedly made evaluation more user-friendly, as did the expansion in the range of types of evaluation (Alie & Seita 1997, 40-49). The turn from the rigorous quantitative positivism of the Robert Wood Johnson Foundation style of evaluation to the "soft," qualitative, and less comprehensive types of evaluation now used by nonprofits undoubtedly made the exercise less demanding and expensive.[5]

Undoubtedly two major trends in the nonprofit sector helped to promote the acceptance of evaluation research -- or at least the rhetoric of evaluation. One was the huge increase in the number of new foundations established during the 1990s. As Foundation Center president Sara Engelhardt pointed out in a recent interview, half the foundations in existence today were created within the past twenty years (Saidel 2002, 489). The proliferation of foundations brought large numbers of donors and trustees, most of them from the results-oriented world of high-tech business, into the world of philanthropy. They expected, realistically or not, to see the results of their generosity. Another factor was the professionalization of nonprofit management, in particular the training of managers in schools of business and public administration, many of which offered courses on evaluation as part of their curricula. Familiar with the rhetoric of evaluation, these managers were adept at giving their donors and boards what they thought they wanted.

Whether there was any real substance to this evaluation activity is far from clear. "Many foundations," Engelhardt says,

> fall into a trap when they decide they want to do evaluation. They aren't really serious about what it would take and what it would mean, but they feel the pressure to do it. They feel they're helping the grantee by doing it. I don't think that's helpful. In fact, in the [mid-1980s], I was chairing the Council on Foundations Research Committee, and its foundation members wanted help in how to do evaluation, so we decided to do an evaluation handbook. It would take good evaluation models from foundations and get them out there. So we made a call for evaluations that

> foundations thought had been helpful. We got the worst, unbelievable things, like counting how many proposals they got and worse. It was clear to us that there was not a very high understanding of what evaluation was, what it would take, and what good evaluation might tell you (Saidel 2002, 493).

Certainly <u>Nonprofit World</u>'s 1997 survey, "Who's Using Evaluation," does not suggest that things have improved much since the mid-1980s. The article defines "formal evaluation" as "the systematic investigation of the worth or merit of an object; for example, a program, project, or grant proposal" (Alie & Seita 1997, 41). This kind of soft approach included such data as written reports, site visits, indirect contacts, financial reports, numbers of clients served, accreditation and licensing, and assessment of management practices -- a far cry from the kind of systematic assessment of impact that Brim of Russell Sage and Rogers of Robert Wood Johnson had in mind.

Engelhardt herself, looking back on thirty years of watching foundations trying to use evaluation, remained skeptical of its value. She saw internal pressure from foundation boards as having played an important role in prompting grant makers to use evaluation:

> On the larger-staffed foundations, often the trustees were not in the field of philanthropy; they were in business or some other areas, and they too needed to understand what is success, often again because you don't see the huge leverage, the huge outcomes that you might have seen before. Several foundations began experimenting even then [in the 1970s] with evaluation and discovered the hard truth that you don't often find out anything useful, except that you bet on a good or a bad person to run the program. Usually, the success or failure couldn't be pinned on the strategy but had more to do with some other things, and often it cost more to study the results of a project, to get some really good useful information, than the project itself cost. And they found evaluations were often not used very faithfully, even if they did give you somr answers, because people didn't trust them (492).

Though she thought that interest in evaluation had subsided, she noted the upsurge of attention it was receiving in the late 1990s, which she viewed (incorrectly in my view) as "actually coming from the new research in the evaluation field, external to foundations and external to the specific research programs that they're supporting" (492). "My own view," she continued

> is that evaluation as part of the grantmaking cycle and process has still not been perfected in a way that makes it terribly useful. Every renewal request, every renewal grant entails some kind of evaluation. Someone is making a value judgment about whether it worked and how and why and often making adjustments to the strategy or the staffing or whatever they think will make it better. It's not as though there's no evaluation, but I think we're still faced with some resource issues and some timeliness issues about the degree to which you can use evaluation effectively to operate a grant program (492).

Engelhardt's interviewer, management scholar Judith Saidel, shared Engelhardt's skepticism, responding that if she were to take a poll among her colleagues she'd find a "fairly high level of skepticism" about the usefulness of evaluation research.

Engelhart reiterated a view that had been voiced as early as the 1970s by Carol Weiss: "evaluation is not meant to inform grantmaking so much as to help the organization." More to the point, as Weiss described it, evaluation as used by major foundations like Kellogg and Lilly appeared to be more geared to aligning grantor-grantee expectations and to protecting foundation staff from board pressure than to yielding measurable results regarding outcome and impact.[6]

A 1998 Kellogg Foundation-commissioned study of the use of evaluation by twenty-one grant makers seems to confirm the notion that evaluation had more to do with maintaining organizational equilibrium than with assessing impact (Patrizi & McMullan 1999).[7] According to the study, the increased importance of evaluation since the mid-1990s "in most cases reflects rising concern among foundation board members about accountability and outcomes" (30). Participants in the study identified five important purposes for evaluation: improving grantee practice, improving foundation practice, providing additional feedback to foundation board members, understanding how grantmaking affects public policy, and gaining insight toward best practices -- purposes that Patrizi and McMullan, the researchers conducting the study, considered problematic. "The wide range of purposes in the 'top five' raises significant questions about the ability of evaluations to meet successfully the broad array of demands placed on them" (32). When they asked participants "to rank the relative priorities they identified as 'very important,'" they found that public policy planning and development -- the ostensible chief end of evaluation research -- was rarely mentioned. Most gave highest priority to strengthening grantee, foundation, and field practice.

What this seems to mean in practice is that "increasingly outcomes-oriented boards" are putting pressure on management and staff for information "regarding the bottom line and efficacy of foundation strategies" (32). Curiously, Patrizi and McMullan report, "management is minimally involved in evaluation," though staff were interested in it as a way of helping them monitor grants in their portfolios. Although they found lip service given to evaluation as a way of improving grantee practice, they could discover little in the way of substance. "We say we want to improve grantee practice and learning," reported on foundation evaluation director, "but we're unwilling to pay for some of the things that might help, like data systems or research, since we're not going to use them ourselves" (33). More than anything, evaluation seemed to serve political putposes. "Evaluations produce information, an important type of currency within any organization," Patrizi and McMullan observe.

> Consequently the role, strategic importance and use of evaluation are often critical junctures of compromise among foundation staff.
>
> We heard from numerous respondents that high-stakes evaluation questions can set program and evaluation staff at odds and complicate the agenda for evaluation (32).

Though the study concluded with a lukewarm endorsement of evaluation, it offered far more detailed descriptions of its shortcomings. Several of their respondents "questioned whether learning from the past was valued when every indication from management and the board was that only the future mattered to them (35).

"Do We Know What We Are Doing?"

Three decades ago, as foundations faced what seemed to be a real possibility of extinction in the face of public and congressional hostility, this question was asked by the head of one of the nation's leading grant makers. He offered evaluation as the answer. Today, when public sentiment is probably more favorable to foundations and other nonprofit organizations than it has ever been, we are still asking the same question -- and still, apparently, offering the same answer.

The difference is that evaluation, while framed with the same rhetoric of rationality and purposiveness, in practice has taken on a very different function. Results-oriented boards demand proof of foundation efficacy, but are indifferent to evaluation findings. Foundation management pressures staff to do evaluation, but does not use the information it generates in planning. Foundation staff do evaluation, but generally lack the resources or the competence to do it with any rigor. Grantees are compelled to participate in evaluation, but -- in instances where they have access to its products -- seldom find them useful.

Formally rational, but substantively ceremonial and symbolic, evaluation as used in foundation today generally appear to have more to do with managing legitimacy than with any genuine concern with efficiency. As Meyer and Rowan note in "Institutionalized Organizations: Formal Structure as Myth and Ceremony" (1991),

> in institutionally elaborated environments, organizations . . . become sensitive to and employ external criteria of worth. . . . For example, the conventions of modern accounting attempt to assign value to particular components of organizations on the basis of their contribution -- through the organization's production function -- to the goods and services the organization produces. But for many units -- service departments, administrative sectors, and others -- it is utterly unclear what is being produced that has clear or definable value in terms of its contribution to the organizational product. In these situations, accountants employ shadow prices: they assume that given organizational units are necessary and calculate their value from their prices in the world outside the organization. Thus modern accounting creates ceremonial production functions and maps them onto economic production functions: organizations assign externally defined worth to advertising departments,

> safety departments, managers, econometricians, and occasionally even sociologists, whether or no these units contribute measurably to the production of outputs (Meyer & Rowan 1991, 51).

"Ceremonial criteria of worth and ceremonially derived production functions are useful to organizations," Meyer and Rowan continue.

> They legitimate organizations with internal participants, stockholders, the public, the state, as with the IRS or the SEC. They demonstrate socially the fitness of an organization. The incorporation of structures with high ceremonial value, such as those reflecting the latest expert thinking or those with the most prestige, makes the credit position of an organization more favorable. Loans, donations, or investments are more easily obtained. Finally, units within the organization use ceremonial assessments as accounts of their productive service to the organization. Their internal power rises with their performance on ceremonial measures (52).

Meyer and Rowan were describing the power of rituals and ceremonies of rationality in organizations that actually had production functions -- business corporations and government agencies that actually produced goods and services. The situation for foundations, which produce nothing and which are accountable neither to external stakeholders (customers, investors, donors) or internal ones (members, stockholders) nor held to any standard of efficacy or efficiency (other than those that boards and staffs may set for themselves) surely increases the functional significance of these rituals. The only maximand for foundations is legitimacy. And the rhetoric -- but not the reality -- or performance evaluation, drawn from the worlds of science and business, is powerfully legitimating.

What explains the proliferation of program evaluation in the foundation world in the face of almost universal skepticism about its value? One reason is certainly the fact that the most energetic promoters of evaluation research in foundations were the three most generous funders of nonprofits research during the 1990s -- Atlantic Philanthropic Service, the Lilly Endowment, and the W.K. Kellogg Foundation. Whatever their particular agendas, the sheer weight of their influence on ARNOVA (Association for Research on Nonprofit Organizations and Voluntary Action), the major association of nonprofits researchers, and their demonstrated capacity to incentivize areas in which they wanted research done, assured that evaluation would receive attention from the scholarly community. None of the three major scholarly journals -- <u>Nonprofit & Voluntary Sector Quarterly</u>, <u>Nonprofit Management & Leadership</u>, and <u>Voluntas</u> had considered evaluation to be of sufficient interest before 1990 to represent in an article or review. All 17 articles on the subject in these journals appeared after 1990 -- 15 of them between 1996 and 2002. Not coincidently, ARNOVA's 2002 lifetime achievement award was given to an evaluation researcher.

While patronage of scholarship that helped to give evaluation research credibility in the nonprofit world undoubtedly played a role in its proliferation, it seems even more likely that institutional factors peculiar to foundations impelled foundation staff to embrace it. In 1991, sociologists Joseph Galaskiewicz and Ronald Burt conducted a study of decision making in corporate philanthropy, focusing on grant makers in Minneapolis-St. Paul. They wanted to know what factors influenced corporate contributions officers' evaluations of grantees -- specifically, why "nonprofits that were recognized and evaluated positively by contributions officers in Minneapolis-St. Paul received more money from corporate donors, irrespective of their size, activities, and fund-raising capacity" (Galskiewicz & Burt 1991, 88). To do this, they asked whether similarity of evaluations was due to communication between contributions officers who would have opportunities to exchange information about grantees through informal contacts at meetings of the professional associations (i.e., regional associations of grant makers, state associations of nonprofit agencies, and the like) or due to symbolic communications in which less prominent contributions officers in less centrally situation organizations legitimated themselves by emulating the judgments of higher status officers. They found that

> direct and indirect contacts had little or no impact on the officedrs perceptions and evaluations. Officers close to one another in the contact network did not recognize the same organizations, nor did they think that the same nonprofits had achieved extraordinary accomplishments. This leads us to believe that knowledge of and opinions about donees does not spread trough direct contacts among giving officers within this occupational subculture. The contributions officers are guided in their

> evaluations by the opinions of their peers in the social structure of other officers rather than the opinons of officers with whom they have personal contact (103).

This contagion effect was stronger, they discovered, for nonprofit organizations providing cultural services, most likely, they believed, because culture is more difficult to evaluate than health and welfare. In other words, having to evaluate cultural services proposals increased the ambiguity of an already ambiguous decision making situation.

The notion of inter-organizational contagion seems a persuasive explanation for the proliferation of evaluation research in the foundation world in the 1990s -- a decade in which the number of foundations increased by 47% and their assets nearly quadrupled. Unlike earlier foundations, which were concentrated in the Northeast and upper Midwest, the new foundations were more likely to be found in areas of the Southeast and West which had less well-developed professional cultures of grant making. As the national philanthropy trade associations -- the Council on Foundations, Independent Sector, National Center for Nonprofit Boards, the National Society of Fund-Raising Executives, and others -- made particular efforts to reach out to new foundations, officers and directors of new philanthropies were, through the course of the 1990s, ever more likely to be exposed to information about evaluation. (In the course of the 1990s, the major non-scholarly philanthropy periodicals published more than 30 articles on evaluation, most of them after 1995). Like the lower status and professionally isolated contributions officers in Galaskiewicz and Burt's Twin Cities study, the new cohorts of philanthropoids are more likely to be influenced by the values and practices of more prominent and well integrated organizations and individuals. "For officers on the periphery of the [grant making] community," Galaskiewicz and Burt write,

> information is more often rumor -- and it is among them that recognition evaluations show evidence of social contagion. In contrast to merely recognizing a nonprofit organization, claiming that it provides extraordinary services is much more a matter of subjective judgment. It is this more ambiguous evaluation that shows consistent evidence of contagion for central as well as peripheral officers (104).

It is easy to understand why evaluation as by most foundations today is little more than conventional wisdom cloaked in a rhetoric of evaluation and planning. The same conditions prevail that Brim identified in his 1973 essay on evaluation: foundations continued to be institutionally isolated by the lack of accountability to external stakeholders -- or as Brim put it, "they don't know whether they are doing what they think they are doing -- or whether what they are doing makes any difference to anyone or not. Institutional isolation breeds narcissism and illusory feelings of power, and separates administrators from the frontiers of thought" (Brim 1973, 223). There is still "no readily accessible hard-headed impersonal way to evaluate the performance of foundation administrators. There are no performance statistics. . . , no bottom line showing profit or loss" (223). The best contemporary evaluation can do is raise subjective standards, based on the aligned expectations of grant makers, grantees, and relevant stakeholders about purpose and process. Though foundation executives and trustees are not as "socially encapsulated" as they were thirty years ago -- boards are no longer overwhelmingly white, male, and Protestant -- to say this is a little like saying that Harvard is no longer an elite institution because it recruits most of its students from public rather than private schools, as it did half a century ago.[8] Foundation executives are no more likely to "deal in truth in exchanges with their primary constituency -- the grant applicants" than they were in Brim's time. Because few efforts have been made to address -- or even study -- the problematic "social relations of philanthropy," as Susan Ostrander calls them (1995), foundation officers are unlikely to hear criticism from grantees (including, of course, evaluators, a substantial number of whom do their work under foundation grants and contracts). Finally, then as now, "foundations lack natural enemies in our society" (Brim 1973, 224). The have no competitors. They are not regularly monitored by journalists, public agencies, or trade associations.

The situation is perhaps best captured by the title given by a task force of nonprofit executives to a 1991 report produced by Independent Sector, Obedience to the Unenforceable: Ethics and the Nation's Voluntary and Philanthropic Community. Foundations are compelled to obey basic fiduciary duties of loyalty and care, to operate within the law, and to make certain required periodic reports to federal and state authorities. Beyond those minimal obligations, the law has little to say about what foundations do and how they do it. And the public, donors, beneficiaries, and other affected have little or no recourse under the law if they are unhappy with what foundations do.

Conclusion

It may well be that the basic premise of evaluation as it applies to foundations was mistaken. Standards of efficiency and efficacy are appropriate to organizations that have constituencies that can not only demand but legally compel accountability. No one has the power or the standing to set such criteria for grant making foundations. The law is largely silent as to what constitutes the substance of charity or the best ways of administering it (Brody in press). And, the episodic fears of foundation leaders about "outbursts of regulatory enthusiasm" notwithstanding, it is inconceivable that government would every seriously attempt to eliminate foundations from the American scene.[9]

Evaluation research was the product of modern welfare state liberalism, premised on the belief that people's lives could be changed through changing the conditions under which they lived. Though often associated with efforts to debunk social change initiatives, the early evaluation literature makes clear the extent to which is liberal policy makers and academics were as eager to find ways of persuasively demonstrating the efficacy of government interventions as conservatives were to find ways of showing that they were a waste of taxpayers' dollars. The conservative revolution that began in 1980 not only succeeded in killing off the liberal welfare state, it also shattered the faith, shared mid-century liberals and conservatives alike, that policy propositions could be tested and were amenable to disinterested, objective analysis. The conservative revolution substituted moral yardsticks for social scientific ones. The utilitarian commitment to the greatest good of the greatest number was replaced by an evangelical commitment to saving the world one soul at a time.[10] Careful methodology and quantitative proof was supplanted by policy based on scripture and anecdote (i.e., the alleged efficacy of faith-based substance abuse programs).

In the current political climate, it hardly matters whether foundations are efficient or whether they can demonstrate the effectiveness of their interventions, since public policy no longer even makes a pretense of being based on scientific knowledge. And, ironically, there is considerable evidence to suggest (see Weiss 1973b) that the belief that it was had more to do with yearnings of liberal academics for engagement in policymaking than with any extant political reality.

One thing is clear: that foundations are free to seek or to avoid efficacy and efficiency; to use or not use evaluation. The existence of real needs -- hunger, suffering, ignorance -- is relevant to what they do only to the extent that their officers, directors, and staff choose to respond to them.

But it is important to keep in mind that the <u>instrumental</u> aspect of foundations -- their demonstrated capacity to initiate and sustain certain kinds of social and political change -- is only one aspect of what they are. The other aspect -- the <u>expressive</u> -- cannot and should not be ignored. The conception of charity as a results-oriented enterprise intended to produce social change is barely a century old. And even in its twentieth century heyday, "scientific" philanthropy comprised only a small portion of the funds given by Americans for charitable purposes (the vast majority was devoted to -- and continues to be devoted to expressive activities, primarily religious ones).

The true genius of American philanthropy may be its capacity to effect change -- not intentionally or instrumentally -- but expressively and often accidently. An early nineteenth century bequest to a college for the teaching of Middle Eastern languages and the acquisition of books about the Middle East languages had almost no relevance to the current needs of American living at the time. The creation of such an endowment could have been no more than an expressive gesture by an individual who happened to be enamoured of Arabic culture. Certainly, over the years, college administrators -- always searching funds to meet current needs -- looked on the growing endowment, the poorly attended classes, and the seldom-read books it paid for with eye to diverting it to more relevant purposes. The day would eventually come when the resources created by this endowment -- the network of scholars, the library resources, and the human capital represented by a pool of people who could speak and read these languages -- would prove to of enormous value, as the United States joined the world community.

The charitable universe in the United States undoubtedly has more than its share of charitable endowments created for expressive rather than instrumental reasons. But the fact that they may not address current needs as defined by some scheme of evaluation does not mean that they serve no social purpose. Who could have guessed that an enterprise as seemingly irrelevant to the needs of a nation in the grips of the Depression -- the Institute for Advanced Study, an enterprise devoted to supporting theoretical work by physicists and mathematicians -- would not only help the United States achieve decisive victory in the second World War, but profoundly shape the character of America's actions as the leader of the "free world"?

Expressive philanthropy is a roll of the dice. It allows individuals to dedicate money in perpetuity to any legal purpose without reference to any measurable standard to usefulness or charitableness. While donors and their trustees are free to seek such standards, they are not required to do so -- nor should proponents of performance evaluation in the foundation world seek to require them to do so. Even if they attempted to hold forth evaluation as a requisite of philanthropic best practice, it seems unlikely, given the nature of foundations as organizations, that they could succeed.

ENDNOTES

1. In this section of his essay, Brim provided a bibliography of key works of evaluation scholarship by such authors by such authors as Donald T. Campbell, Francis H. Caro (whose edited volume of readings on evaluation research Sage had published in 1971), Jack Elinson, Gerald Gordon, Caroline Hodge Persell, Alice Rivlin, Peter H. Rossi, Eleanor Sheldon and Howard Freeman, Carol Weiss, and Walter Williams. The works came from a variety of disciplines and professions (psychology, education, medicine, public health, public policy, sociology, social work) and addressed themselves to a variety of issues (social action programs in health and welfare, psycho-social aspects of disease, education and welfare reform).

2. Privately funded and organized but sponsored by the Department of Treasury, the Filer Commission addressed itself not only to foundations, but also to the entire domain of charitable tax-exempt enterprise. The commission supported a broad range of research projects, the findings of which were published in six weighty volumes in 1977 [Commission 1977; for accounts of the commission and its work, see Hall 1992 and Brilliant 2000).

   In contrast to the Peterson Commission's recommendation, the Filer Commission and its researchers had little to say about evaluation (Commission 1975). Rather than attempting to assess the overall value of tax-exempt enterprise, its focus had shifted to mapping its scope, scale, role, and function. This shift of emphasis was echoed in the 19 75 "Proposal for Study of Independent Institutions," the prospectus for what would become the first academic center devoted to the study of nonprofits. The term evaluation did not appear in its 41 pages. Evaluation would not become a topic of significant interest for nonprofits researchers until the 1990s.

3. On the creation of the history of emerging professional communities as a component of the process of professionalization, see Bernard Bailyn Education in the Forming of American Society (1960). This kind of history, constructed in "an atmosphere of special professional purpose," overemphasizes the distinctiveness of the enterprise and obscures its connections to other institutions and centers of power.

4. I am grateful to my Harvard colleague Carol H. Weiss for her insights into the growth of the evaluation field. Her observations about the current ubiquity of evaluation in the nonprofit world were particularly valuable. She was interviewed by telephone on February 27, 2003.

   According to Weiss, the field fell into disarray in the late 1980s, when Lincoln and Guba's polemical critique of the field, Fourth Generation Evaluation (1989) attacked quantitative positivist approaches. (Even thirteen years after its publication, the book has the capacity to elicit angry comments from veteran evaluation researchers). The two factions eventually mended fences and in doing so helped move the field forward.

5. In fact, the range of activities now considered to constitute evaluation -- often grouped under the rubric "informal evaluation tools" -- so far depart from the concepts and methods of evaluation set forth in the 1970s as to raise questions about whether they really can legitimately be called evaluation research.

6. My observations of Lilly's evaluation activities are based on my serving as a consultant to the Endowment when it was preparing its evaluation guidelines in the late 1980s and, during the 1990s, as a planning and evaluation consultant working with the staff of the its religion division. As the principal investigator on four major grants to Yale's Program on Non-Profit Organizations during the 1990s, I also experienced Lilly's evaluation activities as a grantee. My experience with Kellogg's evaluation process was as principal investigator for its Building Bridges initiative grant to Yale's Program on Non-Profit Organizations. On the formal evaluation guidelines used by these foundations, see W.K. Kellogg (1998) and D. Susan Wisely (1990).

   Lilly's evaluations used a combination of informal pre-grant assessment (usually in the form of extended conversations between program officers and applicants), written reports, site visits, and meetings of grantees participating in particular "initiatives" to report on their work and engage with foundation staff and consultants.[6] Kellogg's approach to evaluation was similar. Applicants under a cluster of grants were convened for focus group discussions. Once grants were awarded, grantees were required to retain outside evaluators who reported to foundation staff. Foundation staff made site visits and periodically convened grantees to report on and discuss their progress. Written reports were also required. Although the foundation had its own evaluation staff, the evaluation process itself was contracted out to a consulting firm that specialized in program evaluation.

   My distinct impression, both as evaluator and evaluatee, was that these approaches to evaluation, while they did afford grantmakers opportunities to identify grantees who weren't performing well and, in the case of research grants, a capacity to make crude assessments of productivity (how many books, articles, conference panels, courses, &c.), really did not enable foundations to assess the impact of their grantmaking. To take one example, the Lilly Endowment invested millions of dollars over the course of more than a decade to address the financial crisis in mainline Protestant denominations. An impressive amount of very fine scholarship was produced. But there is no evidence whatsoever that any of this research had the slightest impact on the continuing decline of the mainlines other than to clarify understanding of why it was happening.

More seriously, neither foundation seemed to have the slightest interest in any kind of critical examination of their own grant making processes. My own contribution to the evaluation of the historical research supported by Lilly's financing of religion initiative was an assessment of the history of the initiative itself. On receiving the paper, the religion division called a meeting of its top staff, called me on the carpet, and consigned the paper to some archival *oubliette*. (It's worth noting in this connection that, back in the 1950s, when Lilly commissioned Canadian sociologist John Seeley to study Indianapolis's Community Chest, the results so displeased the Endowment that it endeavored to suppress what is now regarded as a classic of philanthropic scholarship. The Seeley study was so discredited within the Endowment that it didn't even have a copy of it in its archives -- a defect that I was happy to remedy when I found out about it).

The Kellogg Foundation showed a similar rigidity in its approach to evaluation. In 1997, it launched the "Building Bridges" initiative, an effort to stimulate cooperation between nonprofits researchers and practitioners. The foundation had a peculiarly out-of-date view of the nonprofit sector, insisting that it was a donative and voluntary enterprise, despite compelling evidence to the contrary. For its pre-grant evaluation, intended to define the need for the effort and target the most potentially productive grantees, the foundation summoned to Battle Creek the leading figures in nonprofits scholarship. In the focus group discussions, when a number of researchers questioned the foundation's understanding of nonprofits, they were told in no uncertain terms that if they insisted on voicing these views, they would be asked to leave the project. The subsequent formative evaluation components required all the grantees to attend yearly meetings meetings -- some in places as remote as Argentina. At these, participants were compelled to join in games and role-playing exercises whose purposes can only be guessed. Overall, I came away with the impression that the main objective of Kellogg's evaluation activities was to align grantees' beliefs with those of the grantor -- by force, if necessary

For a number of the Building Bridges initiative participants, Kellogg's coerciveness was so troubling and its evaluation activities so distracting and intrusive that they vowed never to seek funding from the foundation again. Adding to this disenchantment was the feeling of betrayal resulting from the fact that the foundation, having promised funding at a certain level, underfunded all the recipients, yet expected them to do the same amount of work to which they had committed in their proposals.

7. Grant makers participating in this study included the Atlantic Philanthropic Service Company and the following foundations: Annie E. Casey, Edna McConnell Clark, Ford, William T. grant, Robert Wood Johnson, Ewing Marion Kaufman, W.K. Kellogg, DeWitt and Lila Wallace-Reader's Digest, Lilly Endowment, Cahrles Steward Mott, David & Lucille Packard, Pew Charitable Trusts, Rockefeller Foundation, Mary Reynolds Babcock, California HealthCare, California Endowment, Heinz Endowments, James Irvine, and Skillman.

8. As Nicholas Lemann's superb study of the Educational Testing Service suggests, admissions based on standardized testing, while making the applicant poor more diverse in terms of race, gender, and geographical origin, still favored the privileged whose families could afford to live in places with good schools, to provide them with culturally enriched environments, and to purchase the test preparation courses needed for high levels of performance (Lemann 1999).

9. Any attempt to do so would surely violate the Contracts Clause of the U.S. Constitution, which forbids federal and state governments from impairing the obligation of contracts. This was precisely the ground on which the U.S. Supreme Court in the landmark 1819 Dartmouth College Case prevented the State of New Hampshire from taking over and reorganizing a private eleemosynary corporation.

10. Perhaps the best and most influential statement of this viewpoint is Marvin Olasky's The Tragedy of American Compassion (1992), a conservative evangelical critique of social welfare policy written by a born-again University of Texas journalism professor. Olasky's book profoundly influenced the thinking of Republican congressional leaders at the time of the drafting the 1996 legislation that "ended welfare as we knew it." Olasky's writings also helped to shape the thinking of George W. Bush, who incorporated many of his ideas in Texas's welfare reform while he was serving as governor.

Author's Note: I am grateful to Lisa Berlinger, Eleanor Brilliant, Erna Gelles, Laura Leviton, Robert Long, Georgia W. McDaniel, Mark Moore, Andrew Morris, Michael O'Neill, Mark Rosenman, Carol H. Weiss, and Robert Wineburg who shared their thoughts and their experiences about the development of evaluation research.

REFERENCES

Aikin, L.H., Blendon, R.J., Rogers, D.E., & Freeman, H.L. 1980. "Evaluating a Private Foundation's Health Program." Journal of Evaluation and Program Planning III, 119-129.

Alie, R.E., & Seita, J.R. 1997, "Who's Using Evaluation and How: New Study Gives Insight." Nonprofit World15(5)(September), 40-49.

Backer, T.E. 1999. Innovation in Context: New Foundation Approaches to Evaluation, Collaboration, and Best Practices. Encino, CA: Human Interaction Institute.

Backer, T.E., & Bare, J. 1999. "Scanning the Environment for Philanthropic Best Practice Systems." Foundation News and Commentary 40 (November/ December), 25-28.

Bailyn, Bernard. 1960. Education in the Forming of American Society. New York, NY: Vintage Books.

Billitteri, T.J. 1998. "Tracking the Effectiveness of Charities." Conicle of Philanthropy 10 (June 4), 35-37.

Brilliant, Eleanor L. 2000. Private charity and public inquiry: a history of the Filer and Peterson commissions Bloomington : Indiana University Press.

Brim, O.G., Jr. 1973. "Do We Know What We Are Doing?" In F.F. Heimann (ed.), The Future of Foundations, pp. 216-258. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Brody. E. In press. "The Legal Framework for Nonprofit Organizations," from W.W. Powell and R. Steinberg (eds.), The Nonprofit Sector: A Research Handbook – Second Edition. New Haven, CT: Yale University Press.

Buhl, L. 1996. "The Evaluation Paradox." Foundation News and Commentary 37 (January/February), 34-37.

Carson, E. 2000a. "Grantmakers in Search of a Holy Grail." Foundation News and Commentary 41 (January/February), 24-26.

Carson, E. 2000b. "On Foundations and Outcome Evaluation." Nonprofit & Voluntary Sector Quarterly 29 (September), 479-481.

Commission on Foundations and Private Philanthropy. 1970. Foundations, Private Giving, and Public Policy: Report and Recommendations of the Commission on Foundations and Private Philanthropy. Chicago, IL: University of Chicago Press.

Commission on Private Philanthropy and Public Needs. 1977. Research Papers. 6 volumes. Washington, DC: Department of the Treasury.

Commission on Private Philanthropy and Public Needs. 1975. Giving in America: Toward a Stronger Voluntary Sector. Report of the Commission on Private Philanthropy and Public Needs. Washington, DC: Department of the Treasury.

Conner, R.F., & Dikman, F.B. 1979. "Professionalization of Evaluative Research: Conflict as a Sign of Health." Journal of Evaluation and Program Planning II, 103-109.

Curtis, Jody. 1999. "Allison Fine: Best Practices: It's Time to Stop Talking, Start Walking." Foundation News and Commentary 40 (November/December), 30-31.

Davis, H.R. 1972. "Insights: Solution for a Crisis?" Evaluation I(1)(Fall), 3-4.

Doermann, H. 1974. "Statistics Used in Grant Evaluation." Interoffice Memorandum. St. Paul, MN: Bush Foundation.

Drucker, P.1968. The Age of Discontinuity. New York, NY: Harper & Row.

Dunn, W.N., Mitroff, I.I., & Deutsch, S.J. 1981. "The Obsolescence of Evaluation Research. Journal of Evaluation and Program Planning IV, 207-218.

Dykstra, C. 19__. "Evaluation as Collaborative Inquiry."

Fetterman, D.M., Kaftarian, S.J., & Wandersman, A. (eds.). 1996. Empowerment Evaluation: Knowledge and Tools for Self-Assessment and Accountablity. Thousand Oaks, CA: Sage Publications.

Fine, A., Thayer, C.E., & Coghan, A. 1998. Program Evaluation Practice in the Nonprofit Sector. Washington, DC: Innovation Network.

Fine, A., Thayer, C.E., & Coghan, A. 2000. "Program Evaluation practice in the Nonprofit Sector." Nonprofit Management & Leadership 10 (Spring), 331-339.

Flaherty, E.W., & Morrell, J.A. 1978. "Evaluation: Manifestations of a New Field." Journal of Evaluation and Program Planning I, 1-10.

Frumkin, P. 1999. "Evaluating for Success." Philanthropy 8 (September-October), 9-12.

Galaskiewicz, J., & Burt, R.S. 1991. "Interorganization Contagion in Corproate Philanthropy." Administrative Science Quarterly 36 (March), 88-105.

Glasrud. B. 2000. "So It's 2000 - Now What?" Nonprofit World 18 (January February), 16-18.

Glenn, John Mark, Brandt, Lilian, and Andrews, F. Emerson. 1947. Russell Sage Foundation, 1907-1946. New York, NY: Russell Sage Foundation.

Gray, S.T. 1998. Evaluation with Power: A New Approach to Organizational Effectiveness, Empowerment, and Excellence. San Francisco, CA: Jossey-Bass Publishers.

Grønbjerg, K.A., Martell, L., & Paarlberg, L. 2000. "Philanthropic Funding of Human Services: Solving Ambiguity through the Two-Stage Competitive Process." Nonprofit & Voluntary Sector Quaterly 29 (Supplement 2000), 9-40.

Guba, E.G., & Lincoln, Y.S. 1989. Fourth Generation Evaluation, Newbury Park, CA: Sage Publications.

Hall, M. 1977. "Private Foundations and Evaluation." Presented to the Working Group on Evaluation, Blacksburg, VA.

Hall, M. 1979a. "Evaluation and Private Philanthropy: Recent Events and Emerging Trends." Presented at the College of Education, Virginia Polytechnic Institute and State University, Blacksburg, VA.

Hall, M. 1979b. "Evaluation in Corporate Philanthropy." Presented at the Conference Board Seminar, San Francisco, CA.

Hall, M. 1982. "Self-Studies in Corporate Philanthropy." In Corporate Philanthropy: Philosophy, Management, Trends, Future, and Background, pp. 49-51. Washington, DC: Council on Foundations.

Hall, P. D. 1992. Inventing the Nonprofit Sector and Other Essays on Philanthropy, Voluntarism, and Nonprofit Organizations. Baltimore, MD: Johns Hopkins University Press.

Hammack, D. C., and Wheeler, S. 1994. Social science in the making : essays on the Russell Sage Foundation, 1907-1972New York : Russell Sage Foundation, c1994.

Independent Sector. 1999. Obedience to the Unenforceable: Ethics and the nation's Voluntary and Philanthropic Community. Washington, DC: Independent Sector.

Kanter, R.M., & Summers, D.V. 1987. "Doing Well While Doing Good: Dilemmas of Performance Measurement in Nonprofit Organizations and the Need for a Multiple-Constituency Approach." In W.W. Powell (ed.), The Nonprofit Sector: A Research Handbook, 154-166. New Haven, CT: Yale University Press.

Karl, B. D. 1987. "Nonprofit Institutions." Science 236 (22 May), 984-985.

Lagemann, E.C. 19 1983. Private Power for the Public Good: A History of the Carnegie Foundation for the Advancement of Teaching. Middletown, CT: Wesleyan University Press.

Lagemann, E.C. 1989. The Politics of Knowledge: The Carnegie Corporation, Philanthropy, and Public Policy. Middletown, CT: Wesleyan University Press.

Lagemann, E.C. 2000. An Elusive Science: The Troubling History of Education Research. Chicago, IL: University of Chicago Press.

Lemann, N. 1999. The Big Test: The Secret History of the American Meritocracy. New York, NY: Farrar, Straus and Giroux.

Levin, H.M. 1983. Cost-Effectiveness - A Primer. Beverly Hills, CA: Sage Publications.

Marais, L. 1998. "The Relationship between Organizational Culture and the Practice of Program Evaluation in Human Service Organizations." Doctoral Dissertation. Kalamazoo, MI: Western Michigan University.

McKenna, C.D. 2000. "The World's Newest Profession: Management Consulting in the Twentieth Century." Doctoral Dissertation. Baltimore, MD: Johns Hopkins University.

Moore, M.. 1997. Transcript of Speech to Evaluation Task Force (December 3). Author's personal collection.

Moore, M. 2002. "The 'Public Value Scorecard': A Rejoinder and an Alternative to 'Strategic Performance Measurement and Management in Non-Profit Organizations,' by Robert Kaplan." Manuscript in author's personal collection.

Morrell, J.A. and Flaherty, E.W. 1978. "The Development of Evaluation as a Profession: Current Status and Some Predictions." Journal of Evaluation and Program Planning I, 11-17.

Morris, A. 2003. "Charity, Therapy, and Poverty: Private Social
Service in the Era of Public Welfare." Doctoral dissertation. Department of History, University of Virginia.

Murray, V. 2001. "The State of Evaluation Tools and Systems for Nonprofit Organizations." New Directions for Philanthropic Fundraising 31 (Spring), 39-49.

Nielsen, W.A. 1972. The Big Foundations, New York, NY: Columbia University Press.

Nielsen, W.A. 1985. The Big Donors: A New Anatomy of the Great Foundations. New York, NY: E.P. Dutton.

Olasky, M. N. 1992. The Tragedy of American Compassion. Washington, D.C.: Regnery Gateway.

Ostrander, S. A. 1995. Money for Change: Social Movement Philanthropy at Haymarket People's Fund. Philadelphia, PA: Temple University Press.

Patrizi, P., & McMullan, B.J. 1999. "Realizing the Potential of Program Evaluation: A New Study Looks at How 21 Foundations 'do' Evaluation." Foundation News and Commentary (May/June), 30-35.

Poole, D.L., Davis, J.K., Reisman, J., Nelson, J.E. 2001. "Improving the Quality of Outcome Evaluation Plans." Nonprofit Management & Leadership 11 (Summer), 405-421.

Rodriguez-Spagnolo, R. 1992. "A Study of Program Evaluation Capacity in Nonprofit Organizations. Doctoral Dissertation. Ann Arbor, MI: University of Michigan.

Rogers, D.E. 1987. "On Building a Foundation. Foundation News (July/August),

Rossi, P.R., & Wright, S.R. 1977. "Evaluation Research: An Assessment of Theory, Practice, and Politics." Evaluation Quarterly I(1)(February), 5-

Rossi, P.R., Freeman, H.E., & Wright, S.R. 1979. Evaluation: A Systematic Approach. Beverly Hills, CA: Sage Publications.

Russell Sage Foundation. 1956. Report of the Princeton conference on the history of philanthropy in the United States. New York, NY: Russell Sage Foundation.

Saidel, J. 2002. "Interview: Sara L. Engelhardt of the Foundation Center." Nonprofit Management & Leadership12(4)(Summer), 485-336.

Salasin, S. 1972. "What is Evaluation Good For?" Evaluation I(1)(Fall), 2.

Sawhill, J.C., & Williamson, D. 2001. "Mission Impossible? Measuring Success in Nonprofit Organizations." Nonprofit Management & Leadership 11 (Spring), 371-386.

Schwartzman, H.B. 1993. Ethnography in Organizations. Newbury Park, CA: Sage Publications.

Sieber, J.E. 1981. "Purchasing and Selling Evaluation: A Contract Negotiation Process." Journal of Evaluation and Program Planning IV, 173-176.

Shadish, W.R., Cook, T.D., & Leviton, L.C. 1991. Foundations of Program Evaluation: Theories of Practice. Newbury Park, CA: Sage Publications.

Smith, N.L. 1981. "Classic 1960s Articles in Educational Evaluation." Journal of Evaluation and Program Planning IV, 177-183.

Suchman, E.A. 1967. Evaluative Research: Principles and Practice in Public Service & Social Action Programs. New York, NY: Russell Sage Foundation.

Tassie, B., Murray, V., Cutt, J., & Bragg, D. 1996. "Rationality and Politics: What Really Goes on When Funders Evaluate the Performance of Fundees." Nonprofit & Voluntary Sector Quarterly 25 (September), 347-363.

"Three Funders: Process or Outcome?" Foundation News and Commentary 37 (march/April), 46-48.

"To the Reader." 1978. Evaluation and Program Planning I(1)(January), iii.

Tuchman, B.S. 1981. "Accountability to Whom?" Journal of Evaluation and Program Planning IV, 169-172.

W.K. Kellogg Foundation. 1998. W.K. Kellogg Evaluation Handbook. Battle Creek, MI: W.K. Kellogg Foundation. URL: http://www.wkkf.org/ WKKFhandbook.pdf.

Weiss, C. H. 1973a. "Between the Cup and the Lip. . ." Evaluation I(2), 49-

Weiss, C. H. 1973b. "Where Politics and Evaluation Research Meet." Evaluation I(3), 37-45.

Wisely, D.S. 1990. Evaluation Handbook. Indianapolis, IN: Lilly Endowment, Inc.

Worthen, B.R., Sanders, J.R., & Fitzpatrick, J.L. 1997. Program Evaluation: Alternative Approaches and Practical Guidelines - Second Edition. New York, NY: Addison Wesley Longman.