

THE IRON LAW OF EVALUATION AND OTHER METALLIC RULES

Peter H. Rossi

INTRODUCTION

Evaluations of social programs have a long history, as history goes in the social sciences, but it has been only in the last two decades that evaluation has come close to becoming a routine activity that is a functioning part of the policy formation process. Evaluation research has become an activity that no agency administering social programs can do without and still retain a reputation as modern and up to date. In academia, evaluation research has infiltrated into most social science departments as an integral constituent of curricula. In short, evaluation has become institutionalized.

There are many benefits to social programs and to the social sciences from the institutionalization of evaluation research. Among the more important benefits has been a considerable increase in knowledge concerning social problems and about how social programs work (and do not

Research in Social Problems and Public Policy, Volume 4, pages 3-20.
Copyright © 1987 by JAI Press Inc.
All rights of reproduction in any form reserved.
ISBN: 0-89232-560-7

work). Along with these benefits, however, there have also been attached some losses. For those concerned with the improvement of the lot of disadvantaged persons, families and social groups, the resulting knowledge has provided the bases for both pessimism and optimism. On the pessimistic side, we have learned that designing successful programs is a difficult task that is not easily or often accomplished. On the optimistic side, we have learned more and more about the kinds of programs that can be successfully designed and implemented. Knowledge derived from evaluations is beginning to guide our judgments concerning what is feasible and how to reach those feasible goals.

To draw some important implications from this knowledge about the workings of social programs is the objective of this paper. The first step is to formulate a set of "laws" that summarize the major trends in evaluation findings. Next, a set of explanations are provided for those overall findings. Finally, we explore the consequences for applied social science activities that flow from our new knowledge of social programs.

SOME "LAWS" OF EVALUATION

A dramatic but slightly overdrawn view of two decades of evaluation efforts can be stated as a set of "laws," each summarizing some strong tendency that can be discerned in that body of materials. Following a 19th Century practice that has fallen into disuse in social science, these laws are named after substances of varying durability, roughly indexing each law's robustness.

The Iron Law of Evaluation: The expected value of any net impact assessment of any large scale social program is zero.

The Iron Law arises from the experience that few impact assessments of large scale¹ social programs have found that the programs in question had any net impact. The law also means that, based on the evaluation efforts of the last twenty years, the best a priori estimate of the net impact assessment of any program is zero, i.e., that the the program will have no effect.

The Stainless Steel Law of Evaluation: The better designed the impact assessment of a social program, the more likely is the resulting estimate of net impact to be zero.

This law means that the more technically rigorous the net impact assessment, the more likely are its results to be zero—or no effect. Specifically, this law implies that estimating net impacts through randomized controlled experiments, the avowedly best approach to estimating net impacts, is more likely to show zero effects than other less rigorous approaches.

The Brass Law of Evaluation: The more social programs are designed to change individuals, the more likely the net impact of the program will be zero.

This law means that social programs designed to rehabilitate individuals by changing them in some way or another are more likely to fail. The Brass Law may appear to be redundant since all programs, including those designed to deal with individuals, are covered by the Iron Law. This redundancy is intended to emphasize the especially difficult task faced in designing and implementing effective programs that are designed to rehabilitate individuals.

The Zinc Law of Evaluation: Only those programs that are likely to fail are evaluated.

Of the several metallic laws of evaluation, the zinc law has the most optimistic slant since it implies that there are effective programs but that such effective programs are never evaluated. It also implies that if a social program is effective, that characteristic is obvious enough and hence policy makers and others who sponsor and fund evaluations decide against evaluation.

It is possible to formulate a number of additional laws of evaluation, each attached to one or another of a variety of substances varying in strength ranging from strong, robust metals to flimsy materials. The substances involved are only limited by one's imagination. But, if such laws are to mirror the major findings of the last two decades of evaluation research they would all carry the same message: The laws would claim that a review of the history of the last two decades of efforts to evaluate major social programs in the United States sustain the proposition that over this period the American establishment of policy makers, agency officials, professionals and social scientists did not know how to design and implement social programs that were minimally effective, let alone spectacularly so.

HOW FIRM ARE THE METALLIC LAWS OF EVALUATION?

How seriously should we take the metallic laws? Are they simply the social science analogue of poetic license, intended to provide dramatic emphasis? Or, do the laws accurately summarize the last two decades' evaluation experiences?

First of all, viewed against the evidence, the iron law is not entirely rigid. True, most impact assessments conform to the iron law's dictates in showing at best marginal effects and all too often no effects at all. There are even a few evaluations that have shown effects in the wrong directions,

opposite to the desired effects. Some of the failures of large scale programs have been particularly disappointing because of the large investments of time and resources involved: Manpower retraining programs have not been shown to improve earnings or employment prospects of participants (Westat, 1976–1980). Most of the attempts to rehabilitate prisoners have failed to reduce recidivism (Lipton, Martinson, and Wilks, 1975). Most educational innovations have not been shown to improve student learning appreciably over traditional methods (Raizen and Rossi, 1981).

But, there are also many exceptions to the iron rule! The “iron” in the Iron Law has shown itself to be somewhat spongy and therefore easily, although not frequently, broken. Some social programs have shown positive effects in the desired directions, and there are even some quite spectacular successes: the American old age pension system plus Medicare has dramatically improved the lives of our older citizens. Medicaid has managed to deliver medical services to the poor to the extent that the negative correlation between income and consumption of medical services has declined dramatically since enactment. The family planning clinics subsidized by the federal government were effective in reducing the number of births in areas where they were implemented (Cutright and Jaffe, 1977). There are also human services programs that have been shown to be effective, although mainly on small scale, pilot runs: for example, the Minneapolis Police Foundation experiment on the police handling of family violence showed that if the police placed the offending abuser in custody over night that the offender was less likely to show up as an accused offender over the succeeding six months (Sherman and Berk, 1984). A meta-evaluation of psychotherapy showed that on the average, persons in psychotherapy—no matter what brand—were a third of a standard deviation improved over control groups that did not have any therapy (Smith, Glass, and Miller, 1980). In most of the evaluations of manpower training programs, women returning to the labor force benefitted positively compared to women who did not take the courses, even though in general such programs have not been successful. Even Head Start is now beginning to show some positive benefits after many years of equivocal findings. And so it goes on, through a relatively long list of successful programs.

But even in the case of successful social programs, the sizes of the net effects have not been spectacular. In the social program field, nothing has yet been invented which is as effective in its way as the small pox vaccine was for the field of public health. In short, as is well known (and widely deplored) we are not on the verge of wiping out the social scourges of our time: ignorance, poverty, crime, dependency, or mental illness show great promise to be with us for some time to come.

The Stainless Steel Law appears to be more likely to hold up over a

large series of cases than the more general Iron Law. This is because the fiercest competition as an explanation for the seeming success of any program—especially human services programs—ordinarily is either self- or administrator-selection of clients. In other words, if one finds that a program appears to be effective, the most likely alternative explanation to judging the program as the cause of that success is that the persons attracted to that program were likely to get better on their own or that the administrators of that program chose those who were already on the road to recovery as clients. As the better research designs—particularly randomized experiments—eliminate that competition, the less likely is a program to show any positive net effect. So the better the research design, the more likely the net impact assessment is likely to be zero.

How about the Zinc Law of Evaluation? First, it should be pointed out that this law is impossible to verify in any literal sense. The only way that one can be relatively certain that a program is effective is to evaluate it, and hence the proposition that only ineffective programs are evaluated can never be proven.

However, there is a sense in which the Zinc law is correct. If the a priori, beyond-any-doubt expectations of decision makers and agency heads is that a program will be effective, there is little chance that the program will be evaluated at all. Our most successful social program, social security payments to the aged has never been evaluated in a rigorous sense. It is “well known” that the program manages to raise the incomes of retired persons and their families, and “it stands to reason” that this increase in income is greater than what would have happened, absent the social security system.

Evaluation research is the legitimate child of skepticism, and where there is faith, research is not called upon to make a judgment. Indeed, the history of the income maintenance experiments bears this point out. Those experiments were not undertaken to find out whether the main purpose of the proposed program could be achieved: that is, no one doubted that payments would provide income to poor people—indeed, payments by definition are income, and even social scientists are not inclined to waste resources investigating tautologies. Furthermore, no one doubted that payments could be calculated and checks could be delivered to households. The main purpose of the experiment was to estimate the sizes of certain anticipated side effects of the payments, about which economists and policy makers were uncertain—how much of a work disincentive effect would be generated by the payments and whether the payments would affect other aspects of the households in undesirable ways—for instance, increasing the divorce rate among participants.

In short, when we look at the evidence for the metallic laws, the evidence appears not to sustain their seemingly rigid character, but the

evidence does sustain the "laws" as statistical regularities. Why this should be the case, is the topic to be explored in the remainder of this paper.

IS THERE SOMETHING WRONG WITH EVALUATION RESEARCH?

A possibility that deserves very serious consideration is that there is something radically wrong with the ways in which we go about conducting evaluations. Indeed, this argument is the foundation of a revisionist school of evaluation, composed of evaluators who are intent on calling into question the main body of methodological procedures used in evaluation research, especially those that emphasize quantitative and particularly experimental approaches to the estimation of net impacts. The revisionists include such persons as Michael Patton (1980) and Egon Guba (1981). Some of the revisionists are reformed number crunchers who have seen the errors of their ways and have been reborn as qualitative researchers. Others have come from social science disciplines in which qualitative ethnographic field methods have been dominant.

Although the issue of the appropriateness of social science methodology is an important one, so far the revisionist arguments fall far short of being fully convincing. At the root of the revisionist argument appears to be that the revisionists find it difficult to accept the findings that most social programs, when evaluated for impact assessment by rigorous quantitative evaluation procedures, fail to register main effects: hence the defects must be in the method of making the estimates. This argument *per se* is an interesting one, and deserves attention: all procedures need to be continually re-evaluated. There are some obvious deficiencies in most evaluations, some of which are inherent in the procedures employed. For example, a program that is constantly changing and evolving cannot ordinarily be rigorously evaluated since the treatment to be evaluated cannot be clearly defined. Such programs either require new evaluation procedures or should not be evaluated at all.

The weakness of the revisionist approaches lies in their proposed solutions to these deficiencies. Criticizing quantitative approaches for their woodenness and inflexibility, they propose to replace current methods with procedures that have even greater and more obvious deficiencies. The qualitative approaches they propose are not exempt from issues of internal and external validity and ordinarily do not attempt to address these thorny problems. Indeed, the procedures which they advance as substitutes for the mainstream methodology are usually vaguely de-

scribed, constituting an almost mystical advocacy of the virtues of qualitative approaches, without clear discussion of the specific ways in which such procedures meet validity criteria. In addition, many appear to adopt program operator perspectives on effectiveness, reasoning that any effort to improve social conditions must have some effect, with the burden of proof placed on the evaluation researcher to find out what those effects might be.

Although many of their arguments concerning the woodenness of many quantitative researches are cogent and well taken, the main revisionist arguments for an alternative methodology are unconvincing: hence one must look elsewhere than to evaluation methodology for the reasons for the failure of social programs to pass muster before the bar of impact assessments.

SOURCES OF PROGRAM FAILURES

Starting with the conviction that the many findings of zero impact are real, we are led inexorably to the conclusion that the faults must lie in the programs. Three kinds of failure can be identified, each a major source of the observed lack of impact:

The first two types of faults that lead a program to fail stem from problems in social science theory and the third is a problem in the organization of social programs:

1. *Faults in Problem theory:* The program is built upon a faulty understanding of the social processes that give rise to the problem to which the social program is ostensibly addressed;
2. *Faults in Program theory:* The program is built upon a faulty understanding of how to translate problem theory into specific programs.
3. *Faults in Program Implementation:* There are faults in the organizations, resources levels and/or activities that are used to deliver the program to its intended beneficiaries.

Note that the term *theory* is used above in a fairly loose way to cover all sorts of empirically grounded generalized knowledge about a topic, and is not limited to formal propositions.

Every social program, implicitly or explicitly is based on some understanding of the social problem involved and some understanding of the program. If one fails to arrive at an appropriate understanding of either, the program in question will undoubtedly fail. In addition, every program

is given to some organization to implement. Failures to provide enough resources, or to insure that the program is delivered with sufficient fidelity can also lead to findings of ineffectiveness.

Problem Theory

Problem theory consists of the body of empirically tested understanding of the social problem that underlies the design of the program in question. For example, the problem theory that was the underpinning for the many attempts at prisoner rehabilitation tried in the last two decades was that criminality was a personality disorder. Even though there was a lot of evidence for this viewpoint, it also turned out that the theory is not relevant either to understanding crime rates or to the design of crime policy. The changes in crime rates do not reflect massive shifts in personality characteristics of the American population, nor does the personality disorder theory of crime lead to clear implications for crime reduction policies. Indeed, it is likely that large scale personality changes are beyond the reach of social policy institutions in a democratic society.

The adoption of this theory is quite understandable. For example, how else do we account for the fact that persons seemingly exposed to the same influences do not show the same criminal (or noncriminal) tendencies? But the theory is not useful for understanding the social distribution of crime rates by gender, socio-economic level, or by age.

Program Theory

Program theory links together the activities that constitute a social program and desired program outcomes. Obviously, program theory is also linked to problem theory, but is partially independent. For example, given the problem theory that diagnosed criminality is a personality disorder, a matching program theory would have as its aims personality change oriented therapy. But there are many specific ways in which therapy can be defined and at many different points in the life history of individuals. At the one extreme of the lifeline, one might attempt preventive mental health work directed toward young children: at the other extreme, one might provide psychiatric treatment for prisoners or set up therapeutic groups in prison for convicted offenders.

Implementation

The third major source of failure is organizational in character and has to do with the failure to implement properly programs. Human services

programs are notoriously difficult to deliver appropriately to the appropriate clients. A well designed program that is based on correct problem and program theories may simply be implemented improperly, including not implementing any program at all. Indeed, in the early days of the War on Poverty, many examples were found of non-programs—the failure to implement anything at all.

Note that these three sources of failure are nested to some degree:

1. An incorrect understanding of the social problem being addressed is clearly a major failure that invalidates a correct program theory and an excellent implementation.
2. No matter how good the problem theory may be, an inappropriate program theory will lead to failure.
3. And, no matter how good the problem and program theories, a poor implementation will also lead to failure.

Sources of Theory Failure

A major reason for failures produced through incorrect problem and program theories lies in the serious under-development of policy related social science theories in many of the basic disciplines. The major problem with much basic social science is that social scientists have tended to ignore policy related variables in building theories because policy related variables account for so little of the variance in the behavior in question. It does not help the construction of social policy any to know that a major determinant of criminality is age, because there is little, if anything, that policy can do about the age distribution of a population, given a commitment to our current democratic, liberal values. There are notable exceptions to this generalization about social science: economics and political science have always been closely attentive to policy considerations; this indictment concerns mainly such fields as sociology, anthropology and psychology.

Incidentally, this generalization about social science and social scientists should warn us not to expect too much from changes in social policy. This implication is quite important and will be taken up later on in this paper.

But the major reason why programs fail through failures in problem and program theories is that the designers of programs are ordinarily amateurs who know even less than the social scientists! There are numerous examples of social programs that were concocted by well meaning amateurs (but amateurs nevertheless). A prime example are Community Mental Health Centers, an invention of the Kennedy administration, apparently

undertaken without any input from the National Institute of Mental Health, the agency that was given the mandate to administer the program. Similarly with Comprehensive Employment and Training Act (CETA) and its successor, the current Job Partnership Training Act (JPTA) program, both of which were designed by rank amateurs and then given over to the Department of Labor to run and administer. Of course, some of the amateurs were advised by social scientists about the programs in question, so the social scientists are not completely blameless.

The amateurs in question are the legislators, judicial officials, and other policy makers who initiate policy and program changes. The main problem with amateurs lies not so much in their amateur status but in the fact that they may know little or nothing about the problem in question or about the programs they design. Social science may not be an extraordinarily well developed set of disciplines, but social scientists do know something about our society and how it works, knowledge that can prove useful in the design of policy and programs that may have a chance to be successfully effective.

Our social programs seemingly are designed by procedures that lie somewhere in between setting monkeys to typing mindlessly on typewriters in the hope that additional Shakespearean plays will eventually be produced, and Edisonian trial-and-error procedures in which one tactic after another is tried in the hope of finding out some method that works. Although the Edisonian paradigm is not highly regarded as a scientific strategy by the philosophers of science, there is much to recommend it in a historical period in which good theory is yet to develop. It is also a strategy that allows one to learn from errors. Indeed, evaluation is very much a part of an Edisonian strategy of starting new programs, and attempting to learn from each trial.²

PROBLEM THEORY FAILURES

One of the more persistent failures in problem theory is to under-estimate the complexity of the social world. Most of the social problems with which we deal are generated by very complex causal processes involving interactions of a very complex sort among societal level, community level, and individual level processes. In all likelihood there are biological level processes involved as well, however much our liberal ideology is repelled by the idea. The consequence of under-estimating the complexity of the problem is often to over-estimate our abilities to affect the amount and course of the problem. This means that we are overly optimistic about how much of an effect even the best of social programs can expect to achieve. It

also means that we under-design our evaluations, running the risk of committing Type II errors: that is, not having enough statistical power in our evaluation research designs to be able to detect reliably those small effects that we are likely to encounter.

It is instructive to consider the example of the problem of crime in our society. In the last two decades, we have learned a great deal about the crime problem through our attempts by initiating one social program after another to halt the rising crime rate in our society. The end result of this series of trials has largely failed to have significant impacts on the crime rates. The research effort has yielded a great deal of empirical knowledge about crime and criminals. For example, we now know a great deal about the demographic characteristics of criminals and their victims. But, we still have only the vaguest ideas about why the crime rates rose so steeply in the period between 1970 and 1980 and, in the last few years, have started what appears to be a gradual decline. We have also learned that the criminal justice system has been given an impossible task to perform and, indeed, practices a wholesale form of deception in which everyone acquiesces. It has been found that most perpetrators of most criminal acts go undetected, when detected go unprosecuted, and when prosecuted go unpunished. Furthermore, most prosecuted and sentenced criminals are dealt with by plea bargaining procedures that are just in the last decade getting formal recognition as occurring at all. After decades of sub-rosa existence, plea bargaining is beginning to get official recognition in the criminal code and judicial interpretations of that code.

But most of what we have learned in the past two decades amounts to a better description of the crime problem and the criminal justice system as it presently functions. There is simply no doubt about the importance of this detailed information: it is going to be the foundation of our understanding of crime; but, it is not yet the basis upon which to build policies and programs that can lessen the burden of crime in our society.

Perhaps the most important lesson learned from the descriptive and evaluative researches of the past two decades is that crime and criminals appear to be relatively insensitive to the range of policy and program changes that have been evaluated in this period. This means that the prospects for substantial improvements in the crime problem appear to be slight, unless we gain better theoretical understanding of crime and criminals. That is why the Iron Law of Evaluation appears to be an excellent generalization for the field of social programs aimed at reducing crime and leading criminals to the straight and narrow way of life. The knowledge base for developing effective crime policies and programs simply does not exist; and hence in this field, we are condemned—hopefully temporarily—to Edisonian trial and error.

PROGRAM THEORY AND IMPLEMENTATION FAILURES

As defined earlier, program theory failures are translations of a proper understanding of a problem into inappropriate programs, and program implementation failures arise out of defects in the delivery system used. Although in principle it is possible to distinguish program theory failures from program implementation failures, in practice it is difficult to do so. For example, a correct program may be incorrectly delivered, and hence would constitute a "pure" example of implementation failure, but it would be difficult to identify this case as such, unless there were some instances of correct delivery. Hence both program theory and program implementation failures will be discussed together in this section.

These kinds of failure are likely the most common causes of ineffective programs in many fields. There are many ways in which program theory and program implementation failures can occur. Some of the more common ways are listed below.

Wrong Treatment

This occurs when the treatment is simply a seriously flawed translation of the problem theory into a program. One of the best examples is the housing allowance experiment in which the experimenters attempted to motivate poor households to move into higher quality housing by offering them a rent subsidy, contingent on their moving into housing that met certain quality standards (Struyk and Bendick, 1981). The experimenters found that only a small portion of the poor households to whom this offer was made actually moved to better housing and thereby qualified for and received housing subsidy payments. After much econometric calculation, this unexpected outcome was found to have been apparently generated by the fact that the experimenters unfortunately did not take into account that the costs of moving were far from zero. When the anticipated dollar benefits from the subsidy were compared to the net benefits, after taking into account the costs of moving, the net benefits were in a very large proportion of the cases uncomfortably close to zero and in some instances negative. Furthermore, the housing standards applied almost totally missed the point. They were technical standards that often characterized housing as sub-standard that was quite acceptable to the households involved. In other words, these were standards that were regarded as irrelevant by the clients. It was unreasonable to assume that households would undertake to move when there was no push of dissatisfaction from the housing occupied and no substantial net positive benefit in dollar

terms for doing so. Incidentally, the fact that poor families with little formal education were able to make decisions that were consistent with the outcomes of highly technical econometric calculations improves one's appreciation of the innate intellectual abilities of that population.

Right Treatment But Insufficient Dosage

A very recent set of trial policing programs in Houston, Texas and Newark, New Jersey exemplifies how programs may fail not so much because they were administering the wrong treatment but because the treatment was frail and puny (Police Foundation, 1985). Part of the goals of the program was to produce a more positive evaluation of local police departments in the views of local residents. Several different treatments were attempted. In Houston, the police attempted to meet the presumed needs of victims of crime by having a police officer call them up a week or so after a crime complaint was received to ask "how they were doing" and to offer help in "any way." Over a period of a year, the police managed to contact about 230 victims, but the help they could offer consisted mainly of referrals to other agencies. Furthermore, the crimes in question were mainly property thefts without personal contact between victims and offenders, with the main request for aid being requests to speed up the return of their stolen property. Anyone who knows even a little bit about property crime in the United States would know that the police do little or nothing to recover stolen property mainly because there is no way they can do so. Since the callers from the police department could not offer any substantial aid to remedy the problems caused by the crimes in question, the treatment delivered by the program was essentially zero. It goes without saying that those contacted by the police officers did not differ from randomly selected controls—who had also been victimized but who had not been called by the police—in their evaluation of the Houston Police Department.

It seems likely that the treatment administered, namely expressions of concern for the victims of crime, administered in a personal face-to-face way, would have been effective if the police could have offered substantial help to the victims.

Counter-acting Delivery System

It is obvious that any program consists not only of the treatment intended to be delivered, but it also consists of the delivery system and whatever is done to clients in the delivery of services. Thus the income maintenance experiments' treatments consist not only of the payments, but the entire system of monthly income reports required of the clients,

the quarterly interviews and the annual income reviews, as well as the payment system and its rules. In that particular case, it is likely that the payments dominated the payment system, but in other cases that might not be so, with the delivery system profoundly altering the impact of the treatment.

Perhaps the most egregious example was the group counselling program run in California prisons during the 1960s (Kassebaum, Ward, and Wilner, 1972). Guards and other prison employees were used as counseling group leaders, in sessions in which all participants—prisoners and guards—were asked to be frank and candid with each other! There are many reasons for the abysmal failure³ of this program to affect either criminals' behavior within prison or during their subsequent period of parole, but among the leading contenders for the role of villain was the prison system's use of guards as therapists.

Another example is the failure of transitional aid payments to released prisoners when the payment system was run by the state employment security agency, in contrast to the strong positive effect found when run by researchers (Rossi, Berk, and Lenihan, 1980). In a randomized experiment run by social researchers in Baltimore, the provision of 3 months of minimal support payments lowered the re-arrest rate by 8 percent, a small decrement, but a significant one that was calculated to have very high cost to benefit ratios. When, the Department of Labor wisely decided that another randomized experiment should be run to see whether YOAA—"Your Ordinary American Agency"—could achieve the same results, large scale experiments in Texas and Georgia showed that putting the treatment in the hands of the employment security agencies in those two states cancelled the positive effects of the treatment. The procedure which produced the failure was a simple one: the payments were made contingent on being unemployed, as the employment security agencies usually administered unemployment benefits, creating a strong work disincentive effect with the unfortunate consequence of a longer period of unemployment for experimentals as compared to their randomized controls and hence a higher than expected re-arrest rate.

Pilot and Production Runs

The last example can be subsumed under a more general point—namely, given that a treatment is effective in a pilot test does not mean that when turned over to YOAA, effectiveness can be maintained. This is the lesson to be derived from the transitional aid experiments in Texas and Georgia and from programs such as The Planned Variation teaching demonstration. In the latter program leading teaching specialists were asked to develop versions of their teaching methods to be implemented in actual

school systems. Despite generous support and willing cooperation from their schools, the researchers were unable to get workable versions of their teaching strategies into place until at least a year into the running of the program. There is a big difference between running a program on a small scale with highly skilled and very devoted personnel and running a program with the lesser skilled and less devoted personnel that YOAA ordinarily has at its disposal. Programs that appear to be very promising when run by the persons who developed them, often turn out to be disappointments when turned over to line agencies.

Inadequate Reward System

The internally defined reward system of an organization has a strong effect on what activities are assiduously pursued and those that are characterized by "benign neglect." The fact that an agency is directed to engage in some activity does not mean that it will do so unless the reward system within that organization actively fosters compliance. Indeed, there are numerous examples of reward systems that do not foster compliance.

Perhaps one of the best examples was the experience of several police departments with the decriminalization of public intoxication. Both the District of Columbia and Minneapolis—among other jurisdictions—rescinded their ordinances that defined public drunkenness as misdemeanors, setting up detoxification centers to which police were asked to bring persons who were found to be drunk on the streets. Under the old system, police patrols would arrest drunks and bring them into the local jail for an overnight stay. The arrests so made would "count" towards the department measures of policing activity. Patrolmen were motivated thereby to pick up drunks and book them into the local jail, especially in periods when other arrest opportunities were slight. In contrast, under the new system, the handling of drunks did not count towards an officer's arrest record. The consequence: Police did not bring drunks into the new detoxification centers and the municipalities eventually had to set up separate service systems to rustle up clients for the dextoxification systems.

The illustrations given above should be sufficient to make the general point that the appropriate implementation of social programs is a problematic matter. This is especially the case for programs that rely on persons to deliver the service in question. There is no doubt that federal, state, and local agencies can calculate and deliver checks with precision and efficiency. There also can be little doubt that such agencies can maintain a physical infra-structure that delivers public services efficiently, even though there are a few examples of the failure of water and sewer systems on scales that threaten public health. But there is a lot of doubt that human

services that are tailored to differences among individual clients can be done well at all on a large scale basis.

We know that public education is not doing equally well in facilitating the learning of all children. We know that our mental health system does not often succeed in treating the chronically mentally ill in a consistent and effective fashion. This does not mean that some children cannot be educated or that the chronically mentally ill cannot be treated—it does mean that our ability to do these activities on a mass scale is somewhat in doubt.

CONCLUSIONS

This paper started out with a recital of the several metallic laws stating that evaluations of social programs have rarely found them to be effective in achieving their desired goals. The discussion modified the metallic laws to express them as statistical tendencies rather than rigid and inflexible laws to which all evaluations must strictly adhere. In this latter sense, the laws simply do not hold. However, when stripped of their rigidity, the laws can be seen to be valid as statistical generalizations, fairly accurately representing what have been the end results of evaluations "on-the-average." In short, few large-scale social programs have been found to be even minimally effective. There have been even fewer programs found to be spectacularly effective. There are no social science equivalents of the Salk vaccine.

Were this conclusion the only message of this paper, then it would tell a dismal tale indeed. But there is a more important message in the examination of the reasons why social programs fail so often. In this connection, the paper pointed out two deficiencies:

First, policy relevant social science theory that should be the intellectual underpinning of our social policies and programs is either deficient or simply missing. Effective social policies and programs cannot be designed consistently until it is thoroughly understood how changes in policies and programs can affect the social problems in question. The social policies and programs that we have tested have been designed, at best, on the basis of common sense and perhaps intelligent guesses, a weak foundation for the construction of effective policies and programs.

In order to make progress, we need to deepen our understanding of the long range and proximate causation of our social problems and our understanding about how active interventions might alleviate the burdens of those problems. This is not simply a call for more funds for social science research but also a call for a redirection of social science research toward understanding how public policy can affect those problems.

Second, in pointing to the frequent failures in the implementation of

social programs, especially those that involve labor intensive delivery of services, we may also note an important missing professional activity in those fields. The physical sciences have their engineering counterparts; the biological sciences have their health care professionals; but social science has neither an engineering nor a strong clinical component. To be sure, we have clinical psychology, education, social work, public administration, and law as our counterparts to engineering, but these are only weakly connected with basic social science. What is apparently needed is a new profession of social and organizational engineering devoted to the design of human services delivery systems that can deliver treatments with fidelity and effectiveness.

In short, the double message of this paper is an argument for further development of policy relevant basic social science and the establishment of the new profession of social engineer.

NOTES

1. Note that the law emphasizes that it applied primarily to "large scale" social programs, primarily those that are implemented by an established governmental agency covering a region or the nation as a whole. It does not apply to small scale demonstrations or to programs run by their designers.

2. Unfortunately, it has proven difficult to stop large scale programs even when evaluations prove them to be ineffective. The federal job training programs seem remarkably resistant to the almost consistent verdicts of ineffectiveness. This limitation on the Edisonian paradigm arises out of the tendency for large scale programs to accumulate staff and clients that have extensive stakes in the program's continuation.

3. This is a complex example in which there are many competing explanations for the failure of the program. In the first place, the program may be a good example of the failure of problem theory since the program was ultimately based on a theory of criminal behavior as psychopathology. In the second place, the program theory may have been at fault for employing counselling as a treatment. This example illustrates how difficult it is to separate out the three sources of program failures in specific instances.

REFERENCES

- Cutright, P. and F. S. Jaffe
1977 *Impact of Family Planning Programs on Fertility: The U.S. Experience*. New York: Praeger.
- Guba, E. G. and Y. S. Lincoln
1981 *Effective Evaluation: Improving the Usefulness of Evaluation Results Through Responsive and Naturalistic Approaches*. San Francisco: Jossey-Bass.
- Kassebaum, G., D. Ward, and D. Wilner
1971 *Prison Treatment and Parole Survival*. New York: John Wiley.
- Lipton, D., R. Martinson, and L. Wilks
1975 *The Effectiveness of Correctional Treatment*. New York: Praeger.

- Patton, M.
1980 *Qualitative Evaluation Methods*. Beverly Hills, CA: Sage Publications.
- Police Foundation
1985 *Evaluation of Newark and Houston Policing Experiments*. Washington, DC.
- Raizen, S. A. and P. H. Rossi (eds.)
1980 *Program Evaluation in Education: When? How? To What Ends?* Washington, DC: National Academy Press.
- Rossi, P. H., R. A. Berk and K. J. Lenihan
1980 *Money, Work and Crime*. New York: Academic.
- Sherman, L. W. and R. A. Berk.
1984. "Deterrent effects of arrest for domestic assault." *American Sociological Review* 49: 261-271.
- Smith, M. L., G. V. Glass, and T. I. Miller
1980 *The Benefits of Psychotherapy: An Evaluation*. Baltimore: The Johns Hopkins University Press.
- Struyk, R. J. and M. Bendick
1981 *Housing Vouchers for the Poor*. Washington, DC: The Urban Institute.
- Westat, Inc.
1976- *Continuous Longitudinal Manpower Survey, Reports 1-10*. Rockville, MD:
1980 Westat, Inc.

THE "LAW AND ECONOMICS" MOVEMENT

Daniel A. Farber

INTRODUCTION

During the last twenty years, an increasingly influential group of legal scholars has turned to economic analysis as the key to understanding law. Generally, their preferred version of economic analysis has been the Chicago School, of which Milton Friedman is probably the member best known to the public. The Chicago School is characterized by its high regard for free markets and its corresponding disdain for government. It is no coincidence that the "law and economics" movement has been centered at the University of Chicago's law school.

The "law and economics" movement—more familiarly called "L&E"—is important for reasons that go well beyond the law school world. Much of the significant scholarly research about legal problems is done today by members of this group. As a dominant faction at many law

Research in Social Problems and Public Policy, Volume 4, pages 21-37.

Copyright © 1987 by JAI Press Inc.

All rights of reproduction in any form reserved.

ISBN: 0-89232-560-7