

Reviewing For Results

A Comment on "Policing For Results" by Lawrence W. Sherman

Journal of Criminal Law and Criminology

Mark H. Moore

October 31, 1991

In policing, Herman Goldstein and Larry Sherman are contemporary giants. They are particularly important to me because they have long sought to improve policing not just study it. As a result, their disagreements are apt to be of far more than academic significance. They are probably debating an issue that will profoundly affect the future development of policing. For the rest of us, then, it is crucial that we come to understand as best we can the nature of their apparent disagreement.

The Argument of Sherman's Review

In this journal, Sherman has written a sharply critical review of Goldstein's recent book entitled Problem-Oriented Policing. Of course, Sherman would be quick to point out that his review is far from wholly negative. Indeed, he praises the book somewhat fulsomely as "the blueprint for a new police institution and the most influential treatise on policing since Colquhoun's." (p.2) He also sets it apart from other contemporary works that merely "popularize" Goldstein's ideas, or try to set out the "flabby content of "community policing"",(p.4., p.3)

Presumably, it is precisely because Goldstein's book is so important that Sherman feels duty bound to attack it aggressively on one vital point -- a point that Sherman believes threatens to completely undermine the otherwise admirable foundation that Herman has laid. As Sherman observes:

"The paradox is that the book's emphasis on the importance of results is unsupported by any serious concern with the measurement of results. (author's emphasis) For despite the book's success in focusing police attention on the effects of what they do, the book is almost cavalier in its disregard for the complexity of the causal inference necessary for the assessment of those effects.

This paradox leaves the blueprint unfinished...He has omitted a design element that is at least as important as a roof, without which the building cannot last.

The roof we need is some protection against self-delusion about success....Unless the rules of determining success or failure are clearly laid out in advance, problem oriented policing may be destroyed by fraudulent claims of success.

Goldstein does not supply the rules."

Sherman then goes on to make his point about the importance of establishing the rules for evaluation by reviewing two examples of problem-solving policing taken from Goldstein's book. One case involves a Philadelphia Police Officer who negotiated the re-location of a barroom juke box to eliminate repeat calls for service. The other describes an apparently successful effort by the Gainesville, Florida Police Department to reduce convenience store robberies, through, among other things, the passage of a law requiring two clerks to be in the stores after dark. Sherman claims that Goldstein represents these activities (implicitly, perhaps, but nonetheless "strongly") as "successful results of policing." (p. 6) He then asks

whether these programs can justifiably be claimed as successes of "problem-solving policing".

The case involving the re-located jukebox, Sherman finds unexceptionable, but not very important. The case involving convenience store robberies, on the other hand, Sherman finds important, but quite problematic. Indeed, from Sherman's perspective, this case reveals the dangers of failing to take rigorous evaluation methodology seriously.

Sherman argues that if the evaluation of this program were done properly (rather than "as an advocate's brief" -- p. 11), the success of the program would be undermined. More particularly, confidence in the value of a law requiring all convenience stores to have two clerks on duty would be undermined both by doubts both about whether the crime rate decreased significantly following the passage of this ordinance, and whether any observed reduction could be properly attributed to the particular law.

What makes the situation even worse from Sherman's perspective is not only that the Gainesville police may have deluded themselves about their own efficacy, but also that they may be deluding others about the potential benefits of this particular approach to reducing convenience store robberies. Sherman notes that the New York Times recognition of this apparently successful example of problem-solving policing

"helped the Gainesville Police Department's advocacy of a Florida state law requiring all convenience stores to have two clerks on duty after dark. If passed, it would be the first such state law in the country, and a major innovation public policy apparently endorsed by the authority of Goldstein's writing.

Yet nowhere does Goldstein employ the principle of rival hypotheses. The only hint of qualification in judging the Gainesville case a success comes in one vague sentence.... The book clearly does not use the principle for any serious consideration of the possibility that the decline in robberies might have been the result of factors other than the two clerk law." (p.12)

Thus, Goldstein's praise of this second case is condemned as insufficiently rigorous, irresponsible, and revelatory of the hazards of being too eager to find problem-solving policing efforts succesful.

Larry closes his review article with a more conciliatory and constructive discussion about how much evaluation is enough. He acknowledges (following Lindblom and Cohen) that "social science is only one of many possible

approaches to social problem solving" (p.21), but insists that "social science principles" have great value in evaluating (as against solving) social problems. (p.21.) He acknowledges that "police themselves can be trained in these principles to perform more reliable peer assessments of results" (p.21), and that "there may be far too many problems to be addressed for an independent review of every police result in any depth" (p.21), but then cautions that without some rigorous, independent evaluations, the police may once again revert to an organization that concentrates too much on following rules, and not enough on achieving important substantive results.(p.22)

He suggests that the important "middle ground" between no evaluations on one hand and rigorous independent reviews on the other might be occupied by increasing the skills of police managers and special evaluation units within the police departments, and by developing some "sliding scales" that would identify how much evaluation a particular initiative was worth according to such considerations as: 1) the scope of the problem or its solution; 2) the level of effort the police put into solving the problem; or 3) the cost of the solution -- either to the police or to others. The bigger the problem, the greater the effort, the more expensive the solution, the more important it would be to evaluate the effort rigorously.

Reviewing Sherman's Review

There is much of value in Sherman's review. I think it is both correct and important to focus attention on how the problem solving efforts of the police should be evaluated. I also think the ideas at the end of the piece about how to distribute responsibilities for evaluation across police organizations, and between police organizations and outside evaluators are quite interesting and important. And I particularly like the fact that Sherman's review will initiate (starting with this essay) a long, useful discussion about: 1) the role that rigorous, independent evaluations of particular police programs can and should play in the future development of the field of policing; and 2) how the general strategy of problem-solving policing (as distinct from particular program initiatives taken under its aegis) might be evaluated.

Nonetheless, I think there is also much to criticize in Sherman's review. And the criticisms may be important to help us all learn something from the disagreements between these giants.

The Justice of the Ad Hominem Attacks

First is the matter of tone, and the ad hominem quality of some of Sherman's criticisms. Sherman might well be surprised to have his review characterized as an attack on Goldstein whom he clearly admires. Yet, when a scholar is accused of being "cavalier in [his] disregard for the complexity of causal inference" (p.4); or when he is faulted with omitting "a design element" that provides "some protection against self-delusion" (p.5); or when it is predicted that police will be left "as ignorant of the results of their work as they were before the book was written" (p.6); or when he is told that by "failing to put the weight of one's enormous prestige and authority behind the importance of the principles of evaluation, one provides a de facto endorsement of overblown claims of success" (p.9); it is hard not to take the accusations at least somewhat personally.

Perhaps the tone was nothing more than the result of a misguided effort to write vividly and powerfully about what seems to Sherman to be a crucial weakness in Goldstein's book. Or, perhaps the tone could be accounted for by the disappointment that an apprentice commonly feels when the master or compromises a point that has always been central to the thinking of the apprentice.

But there seems to be more in the attack than just this. One feels, in the heat of the language, not only the presence of a disillusioned acolyte, but also a determined and ambitious priest eager to keep the temple of program evaluation pure. Even worse, one comes to wonder whether self-interest and professional pride aren't doing some mischief in the article as Sherman seeks to maintain the grip of social science evaluative methods on the right to say authoritatively what does and does not work in policing.

Indeed, one of the most startling things about Sherman's review is the amount of time he devotes to the detailed methodological discussion of the Gainesville, Fla. case. This occupies only a long, single paragraph in Goldstein's book, and is used primarily to illustrate the overall method of problem solving including diagnosis and design as well as evaluation. It is not offered as a definitive defense of a particular program that has been proved to deal effectively with convenience store robberies. Yet, this case, and the methodological weaknesses of concluding from the available evidence that a two clerk law will reliably reduce convenience store robberies, is given great prominence in Sherman's review.

To be fair to Sherman, Goldstein is a little casual in his use of this story. It is clear that the diagnosis and search for possible approaches to the problem are presented

as a "good" example of "problem-solving methods." Sherman agrees with this.

What is less clear is whether Goldstein also means to claim that, in this particular case, the particular police response generated by problem-solving methods were successful in reducing convenience store robberies, and if so, whether that implies that similar methods would be successful elsewhere.

Goldstein would clearly like to be able to claim that the application of problem-solving methods, and the particular response they suggested to police, did produce a good result. Otherwise, the claim that the use of problem-solving principles is valuable to police departments is undermined.

On the other hand, Goldstein is far less invested in being able to claim that the particular methods used by the Gainesville Police to control convenience store robberies represent a proven, nationally effective approach to this particular police problem. Indeed, one of the most important ways that the idea of "problem-solving policing" departs from what might be thought of as the "research and development" approach to policing is that "problem-solving policing" is much less confident that there are general answers to how particular police problems might best be addressed.

Central to the idea of "problem solving" is the idea that problems might differ significantly from one place to another, and that even if the problems were the same, the solutions might be different since the materials available for a solution might well differ from place to place. What is important to problem-solving, then, is the method of analysis, not the general validity of the particular solution arrived at in one place for all other places. Even if general answers existed, the philosophy of problem-solving suggests that they will be discovered slowly, over time, as we learn how to describe problems, and work our way through several failed evaluation efforts.

In contrast, Sherman is far more wedded to the "research and development" approach to policing. In this conception, it is possible to test the efficacy of particular approaches to particular problems definitively, and to know through one experiment the answer once and for all whether a particular approach works. From this vantage point, any place can be a site for a nationally significant experiment. And any particular program can be made to serve this purpose if only it is evaluated properly. The reason is simply that the underlying behavioral laws that produced the effects in one place can be counted on to produce the same behavioral effects elsewhere.

Of course, if one is a scientist, one has to believe these things. But one can remain a scientist and still observe that the limits of our ability to observe, measure and make inferences, even applying the most sophisticated methods, may prevent us from seeing clearly the workings of these behavioral laws in any individual experiment. And it is for this reason that replications of even the most skillfully designed experiments are important.

Ironically, Sherman has been at the center of one of the most ambitious efforts "research and development efforts" in policing: an NIJ sponsored program to rigorously test the value of arrests as a response to domestic violence. What that experience has shown is that rigorously produced results will not necessarily be replicated in subsequent experiments. Thus, it is not at all clear that the "research and development" approach that seeks to leap to a conclusion about the effectiveness of a particular approach through one brilliantly conducted experiment will succeed. It may take a whole host of different kinds of inquiries to finally reveal what kinds of responses to domestic violence are most effective in what particular situations. That is more consistent with the "problem-solving" view of how technical progress occurs than the "research and development" view.

Given Sherman's background and interests, then, it is quite natural for him to respond to the Gainesville story not simply as an illustration of problem-solving methods at work, but as a claim that the particular police responses made were effective in controlling convenience store robberies, and if that was true, that these approaches might become a national model for how to deal with this problem. Indeed, this becomes an even more natural perspective for him to adopt given that Sherman's Crime Control Institute was hired by the National Association of Convenience Stores to analyze the Gainesville experience to determine what lessons it held for the country as a whole as it responds to this particular crime problem.

Indeed, one might say that Larry has really written two articles in his review. One, consisting of the sections entitled "The Paradox of Results" and "How Much Evaluation is Enough", is an interesting but somewhat disconnected discussion of what role evaluation of police programs should play in the future of policing. (I say somewhat disconnected because the first section makes an impassioned argument for rigorous standards all the time, while the second presents a much more moderate position that is closer to what Goldstein recommends.) The second, consisting of the sections entitled "Two Case Studies" and "Principles of Evaluation", is an excellent primer on the techniques of rigorous program evaluation, and an application of those principles to a

substantive discussion of whether convenience store robberies actually declined in Gainesville following the police problem-solving initiative, and if so, what particular features of the intervention caused that effect to occur.

Each of these articles, in themselves, is quite interesting. What is upsetting is the way that they are joined in the review of Goldstein's book to suggest that Goldstein has behaved irresponsibly by "provid[ing] a de facto endorsement of overblown claims of success and leav[ing] more rigorous independent evaluators vulnerable to charges of sour grapes." (p.9) The logic that links the two together, of course, is Sherman's belief that it is irresponsible to make claims for the general success of a particular problem solving method in dealing with a particular problem until it has been validated by rigorous independent review, and his natural assumption that Goldstein has done precisely this in using, without sufficient caution, the example of Gainesville's efforts to deal with convenience store robberies. I think Sherman has attacked Goldstein personally and unfairly on these points.

Sherman's claim that Goldstein ignores the issue of "rigorous evaluation" is simply wrong. Sherman says this:

"He [Goldstein] does provide some general discussion of the issue, consuming less than three pages. [Pages 145-147] But the discussion lacks clear examples, and seems half-hearted."

Sherman has apparently missed another passage in Herman's book on p.49, entitled "Evaluating Responses of Newly Implemented Responses." Indeed, there is no reference to this section in Larry's footnotes.

It is worth quoting from this section at some length not only to show that Goldstein's commitment to evaluation is as great as Sherman's but also to indicate that Goldstein comes to many of the same conclusions that Sherman presents later in his essay when he stops attacking Goldstein for the Gainesville case and begins focusing on the problem that Goldstein has been focused on all along: namely, how to divide the tasks of evaluation within a police department, and between the department and outside evaluators.

"Concern about effectiveness naturally includes a commitment to evaluation. It is obviously important to guard against replacing one ineffective response with another and to ensure that the claims made for new responses are substantiated. ...

The problem-oriented approach calls for developing -- preferably within a police agency -- the skills, procedures, and research techniques to analyze problems and evaluate police effectiveness as a continuing part of management. Given the wide range in the type, level and size of problems, there will be equal variety in the type, level, and size of evaluation that is required.....A substantial effort has been made to provide guidance to police personnel in research methodology so that, in modest but sound ways, they can do a better job than they have ever done in the past of measuring the impact of their actions on the problems they handle...High quality evaluations of the effectiveness of major changes are difficult because of the large number of variables that can affect outcomes and because of the enormous effort and cost involved in setting up controlled experiments. For these, the police must develop new and more productive relationships with academics for the skills, time and independence required. And they must turn to foundations and governmental agencies for the needed financial support." (Goldstein, p.49)

Of course, Sherman could still claim that this was not enough emphasis given to the subject of evaluation, but this passage is a far stronger statement than the one he cites.

Similarly, Sherman's claim that Goldstein has provided a "de facto endorsement of overblown claims of success" of which the convenience store clerks is cited as "a prime example" (p.9) is also false. Goldstein presents the Gainesville case not as an argument that a "two clerk ordinance" will inevitably, in all circumstances, reduce convenience store robberies. Instead, he also offers it as an illustration of how the method of problem-solving proceeds, and to show that the methods can sometimes seem to produce satisfactory results.

Indeed, even in Sherman's discussion, it was the New York Times editorial page that "heralded the Gainesville case as dramatic evidence of the successful results of his new approach to policing" and that "helped the Gainesville Police Department's advocacy of a Florida State law" -- not Goldstein. And it is the Gainesville police view that "anyone who questions their claim of success must be representing an "interest group", and that alternative explanations should be fought off as an attack on the police department." Goldstein is held responsible in these events because his "account [of the Gainesville case] is so impressive" that the New York Times was stimulated to recommend problem solving as a new approach to policing; " and for writing a sentence that "supports" the Gainesville police view.

What we have here is Sherman complaining about the difficulty he has had in persuading people that it is not certain that the two clerk rule has been successful in controlling convenience store robberies, and using Goldstein's account of the Gainesville initiative as the whipping boy. What we do not have is any evidence that Goldstein has been sloppy in making inferences about the effectiveness of crime control methods, or irresponsible in representing evidence about these matters.

At any rate, it would be best if these personal attacks could be set aside. There is plenty of work to do in figuring out the substantive issues we face in improving policing without getting distracted by the question of whether one has behaved responsibly or not. Such discussions generate far more heat than light, and I am sorry to have given as much space as I already have to these personal recriminations.

How Much Independent, Scientific Program Evaluation?

The most obvious and important substantive issue dividing Goldstein and Sherman is the question of what role "evaluation" will play in the future of problem oriented policing. Sherman is primarily concerned that Goldstein has

not given serious "evaluation" a sufficiently prominent role in problem oriented policing, and that without this key ingredient, problem oriented policing will fail to help the field as a whole develop. Sherman finds this particularly disappointing since, in his view, it was Goldstein who first powerfully directed the field's attention to the substantive results of their organizational efforts, and away from their pre-occupation with internal organizational arrangements. From Sherman's perspective, it is only continued attention to "evaluation" that can keep the police focused on achieving substantive results, and produce reliable information about whether they have done so.

Part of the difficulty here is that Goldstein and Sherman seem to mean somewhat different things by the idea of evaluation. To Sherman, "evaluation" means a quite particular thing. It means measuring the results or outcomes of policing, not just the activities of police organizations -- including, in particular, effects on crime and fear. It also means relying on research designs that can reliably attribute any observed effect on the outcomes of interest to the measures taken by the police rather than to other factors. And it often seems to mean using outsiders to perform these evaluations to ensure their objectivity and validity.

In principle, these attributes of evaluation efforts need not go together. For example, one could have evaluations with elaborate measures of outcomes and rigorous research designs performed by insiders as well as outsiders. Or, one could have outsiders look closely at measured outcomes, but without the benefit of a rigorous research design.

In practice, however, these three different attributes are often bundled together in what has become the common paradigm of outside, scientific, program evaluation. They are bundled together because they are all necessary to establish whether a program did or did not succeed in achieving its purposes with a high degree of confidence.

Goldstein seems to have a much wider and more eclectic view of what constitutes "evaluation." He is certainly interested in measuring the results of police efforts. But he is more catholic in his view about what sort of evaluations would be worth doing. Evaluations based on research designs that could not attribute measured effects to police efforts would be rejected by Sherman, but would probably still be of interest to Goldstein. Moreover, unlike Sherman, Goldstein seems to prefer inside evaluations to outside evaluations (Note the passage quoted above where Goldstein inserts the clause "preferably by insiders"). So there is a real difference in their views about what sorts of evaluations should be done.

Now, we all know that if we want to get close to the truth of whether a program or activity has succeeded in its purposes that Sherman's methods are the correct ones. The question, then, is why someone with Goldstein's sophistication is not insisting on sole reliance on these methods? Indeed, if Sherman wanted to take Goldstein to task more effectively, it would be to challenge him to give reasons for tolerating less than the best kinds of evaluations. I don't know exactly what Goldstein would say if challenged on this point, but I can imagine four justifications for his more tolerant views.

First, Goldstein observes that the high cost and practical difficulties of accurate measurement and research design make it impossible to evaluate every problem-solving activity with the most elaborate, high-tech methods of evaluation. There are some things that are simply not worth evaluating this way. For these activities, then, one faces a choice between doing the evaluation imperfectly, or not at all. In these cases, Goldstein would opt for imperfect evaluations, simply on the grounds that something is better than nothing. It seems, from the tail end of Sherman's review that Sherman agrees with this position, though they might still disagree about exactly where in the distribution of different kinds of police activities one ought to draw the line, and might also differ about the value of the efforts that are not full fledged scientific evaluations.

Second, financial constraints are such that there will always be less high tech "research and development" type evaluations than could technically be accomplished. There is simply not enough money in research institutes and foundations interested in policing to support all the important, technically feasible evaluations that could be done. The implication, then, is that if one wants to do evaluations, one must do it with resources from within the field of policing itself. That means that many of the evaluations will be less than ideal. To the extent that one thinks high quality evaluations require outside evaluations, the quality is immediately degraded by the fact that insiders are doing them. In addition, however, it is certainly true that, in the short run at least, inside personnel are less skilled than many outsiders in performing the technical parts of the evaluation. From Goldstein's perspective, however, given the mass of problems to be understood and managed, increasing the overall volume of evaluative activity is better than suppressing evaluations because they cannot meet standards for scientifically validating the accomplishment of the police.

Third, in trying to decide which of the relatively small number of police activities should be the focus of expensive (but convincing) high tech evaluations, having a large number of less sophisticated evaluations on varied

police activities may be extremely helpful. In the past, it was relatively clear what sorts of police activities should be the focus of research attention. The reason was that the field thought of itself as using only a limited number of crime control techniques, including random patrol, directed patrol, rapid response to calls for service, and retrospective investigation of crimes. Moreover, these were thought to be valuable in addressing all crimes.

In the future, if the field takes Goldstein's advice and begins both disaggregating types of crime and adding new methods of dealing with the new types, the number of police activities that will need to be evaluated will increase. The question then will be which of the many different types of problems then faced by the police will be sufficiently important, sufficiently common and sufficiently well addressed to merit the cost of a high tech evaluation. Will it be the proper response to domestic assaults, or the handling of the mentally ill on the street, or identifying and apprehending serial killers? At a minimum, the existence of a large number of unsophisticated internal evaluations may provide some indication of how the field as a whole wants to aggregate problems, which kinds of problems seem to be the focus of the field's attention, and what have been more or less successful methods of measuring the results.

Fourth, and perhaps most importantly, encouraging low tech evaluations may be important not because it answers definitively whether a police activity did or did not have the desired effect, and not because it produces certain knowledge of how a particular problem can be attacked in other departments as well as the one that first initiated the effort, but simply because the act of conducting evaluations will have important effects on the attitudes and capabilities of the organization itself. Sherman himself recognizes that it would be value to introduce the "principles of evaluation...slowly to police culture" and to present them as "integral to the very idea of problem oriented policing." (p.9) But he fails to take from this observation the point that this goal might be advanced by letting the police perform evaluations they way that they want to do it for a while until they become more familiar with the risks and value of the methods, and that that would be aided by giving encouragement to evaluations that depart in many ways from the pure academic style.

What should be discouraged, of course, is the kind of evaluation that does nothing other than glorify the organization, and does so with bad data and worse inferences. But the point is that in order to make problem solving work, and even increase the future utility of high tech evaluations, one might have to prime the pump of doing evaluation work inside police organizations, and that that

goal would be aided by having the police rather than outsiders do the work. If the price of such encouragement is looser standards in the newly expanded sector of evaluation, then that is a price one should be willing to pay.

Actually performing evaluations inside police organizations could be expected to produce important changes in police culture. It might encourage the sense of purposefulness that Sherman thinks is the essence of problem-solving policing since it encourages the police to keep thinking in terms of their objectives and how their particular accomplishments might be measured. It might also encourage a greater sense of accountability, since it puts in their hands the tools to help themselves become accountable. It might also allow the police to find within themselves reserves of analytic talent and imagination that they did not know they had, and to widen the range of people doing analytic work in police departments beyond the planning staffs and crime analysts to the managers and the officers themselves.

Such hopes seem reasonable given the increased professionalism of the police, their high professional aspirations, developments in computer technology, and the emergence of a broad philosophy of policing that re-distributes the right and the responsibility for thinking about police objectives and methods from the academics or

the chief to ordinary officers. From Goldstein's perspective, the worst thing that could now happen would be for the police managers and officers to once again be told that they should not and could not think for themselves -- this time because it is only outside, scientifically trained evaluators who can accurately and reliably determine whether police activities have been successful.

Of course, to some extent the claim that it would be valuable to support low tech internal evaluations as well as high tech external evaluations because the low tech evaluations will produce desirable organizational changes sounds suspiciously like a return to what Sherman takes to be the "bad old days" when the police were primarily concerned about organizational arrangements rather than substantive outcomes; when they were like "bus drivers who cannot stop to pick up passengers because it would make them fall behind schedule." (p.2) Once again, it seems that we are pleading for the virtue of something less than certain knowledge about performance by claiming some potentially desirable organizational effect.

But that claim, it seems to me, misunderstands the relevance and value of both thinking about and trying to produce some important organizational effects. Indeed, it lies at the heart of a third problem with Sherman's critique of Goldstein's work -- the failure to distinguish between:

1) the evaluation of particular programmatic or problem solving activities initiated by the police to deal with different problems faced by police organizations; and 2) the evaluation of problem oriented policing as a general idea about how police departments ought to be organized and operated.

Evaluating Problem Oriented Policing as an Overall Philosophy and Managerial Approach to Policing

Throughout Sherman's review, he consistently confuses the evaluation of particular programs developed by police departments to respond to particular problems facing them, with the evaluation of problem-oriented policing as an overall style or philosophy of policing with significant implications for both the management and the evaluation of police departments. Thus, for example, Sherman declares: "Unless the rules of determining success or failure [of particular projects that are to be evaluated] are clearly laid out in advance, problem oriented policing may be destroyed by fraudulent claims of success."(p.5) He also treats the Gainesville case as though it were, in itself, a test of the value of "problem oriented policing."

To a degree, treating the success of particular programs designed with the use of problem solving methods as

a test of the success of problem-oriented policing more generally is fair and appropriate. There are some distinctive approaches associated with problem-solving (such as the use of situational preventive approaches, or the reliance on other governmental agencies to assist the police in their work) that can be tested in particular applications, and give some rough indication of these general kinds of approaches. It is also true that if problem oriented policing as a whole is to be successful, it will be because it succeeds in a whole series of individual problem-solving efforts like the example of Gainesville, Fla.

Nonetheless, the success of any particular problem-solving attempt can be treated as only a partial test of the success of the overall strategy of problem-oriented policing. The reason is simply that problem-oriented policing is an idea about how police organizations should be directed, structured, managed and operated, as well as a bundle of particular ideas about how police work should be done.

Goldstein is clear about this. He is writing to feel the need for "a broad conceptual framework" that "helps the police build a strong, sensitive institution, with refined methods of operating, that can better transcend the crisis of the day, whether that crisis be labor-management strife, racial conflict, political protest, drugs, or yet-to-be-

identified social problem." (Goldstein, p.xiii.) Moreover, he locates problem-oriented policing in the broad context of historical changes in the overall philosophies of policing, and devotes an entire chapter to a discussion of the basic changes in police organization and management that would be necessary to support a consistent pattern of problem-solving police efforts. And Sherman himself points out that the "calm and balanced tone" of Goldstein's work "understates the book's radical attack on current police practice, with its revolutionary implications for all aspects of the police institution: who becomes police, how they are trained, how they are rewarded, what powers society will grant them."(p.4)

So, both scholars understand problem oriented policing both to be a statement of organizational approach, and to require significant organizational changes as well as a particular set of programs that the police use to deal with particular programs. They understand that the task is to create purposeful, proactive, analytically resourceful, and accountable police organizations as well as to accumulate knowledge about the best methods for dealing with individual problems.

If the idea of problem-oriented policing is an organizational idea, however, then its worth cannot be evaluated by the substantive success of any given program.

It could only be judged over the long term by comparing the success of a police department committed to problem-solving approaches methods with some other department that was committed to some other methods -- presumably the traditional ones.

Note that in measuring the overall value of a problem-solving police organization, some particular problem solving efforts will inevitably be recorded as failures, and others as successes that could not be attributed exclusively to police action. But these would not necessarily indicate failure of the organizational strategy. Indeed, overall success might well depend on the learning that comes from individual program failures. Evaluation would be focused the "bottom line" of the organization as a whole, not of any particular "product line" that had been introduced as a result of problem-solving methods. Thus, in evaluating problem solving policing in Gainesville, I would be as interested in knowing whether and how Gainesville applied their problem-solving methods to other problems than convenience store robberies, and what their overall "batting average" was in controlling crime and reducing fear than in the success of that one individual program.

Note also that if one thought an important task at hand was to help to build problem-solving organizations -- ones that could be analytically resourceful, purposeful, and

curious about their own performance -- then one might well think about the question of what kinds of evaluation should be done by whom with an eye to the effects of this decision not only on the accuracy of any particular evaluation, but also on the prospect of seeing problem-solving police organizations in this century.

This, it seems to me, is the principal difference between Goldstein and Sherman. Goldstein is actually interested in building organizations that can become purposeful and self-evaluating. He knows that unless the police organizations begin performing the evaluation function for themselves, that things will remain as they now are: with most police activities going unevaluated because there isn't enough money, or social scientists, or interest, to evaluate more than a handful of police activities. He also knows that by encouraging the police to do problem-solving and evaluation, however crudely, a base is laid for future organizational development and a continued tolerance of and capacity for learning.

For his part, Sherman knows what a powerful program evaluation looks like, and is skeptical of the police every being able to do the job. Moreover, he is suspicious of any effort to think about the problem of developing police organizations, for fear that such thoughts will drag us back to the time when we spent all our time thinking about police organization and not about results.

What he seems not to have considered is exactly how we get from where we now are in policing to a world in which police organizations are purposeful and well-evaluated without thinking about issues of organizational development. That remains a challenge to us all, as does the even more difficult question of how overall strategies of policing might be evaluated even as police departments are making strategic changes.