

Police Innovation Contrasting Perspectives

Edited by

David Weisburd and Anthony A. Braga

 **CAMBRIDGE**
UNIVERSITY PRESS

17 Critic

Improving police through expertise, experience, and experiments

Mark H. Moore

The words “expertise,” “experience,” and “experiment” all come from the same Latin root: *ex per*, and it means *from danger*. The common root suggests that at one time in human history, these ideas were closely connected with one another. Interpreting unforgivably, we might conclude that human beings have long been aware of the danger of acting with uncertain knowledge of consequences; further, that an important way to reduce this danger was to rely on the *expertise* that was rooted in *experience* of which one important type was an *experiment*.

The commitment to “evidence-based policing” is today’s effort to reduce the danger associated with relying on police practices and methods whose consequences are, in an important sense, unknown, or at least uncertain. The proposed solution lies in the careful accumulation of actual concrete experience with particular police methods to discover which methods seem to work to produce desired results. But the focus on *evidence-based*, rather than *experience-based* knowledge suggests that it is not just any old experience that can be used in developing a more solid base for action. It is not, for example, the kind of experience we recognize as commonsense. Nor is it the kind that accumulates as police lore. It is not even the kind of experience captured in detailed case studies. It is instead the kind of experience that is captured in observational studies that reduce experience to numbers that can be systematically analyzed to discover the generality and reliability with which a particular intervention produces desired results. It may be even more particularly the kind of experience generated by carefully designed, randomized experiments.

As a social scientist, and as a person who longs to put police work on a more solid empirical basis, it is impossible to be against a movement that supports “evidence-based policing.” I think it is important to find out what works in policing. I think social science methods provide the most powerful methods available to us to determine what works in policing as well as in other fields. My only concern is that by focusing too much on the experience that can be captured in quantitative observational studies and controlled experiments – by assuming that these methods can stand

alone, and that they are the only ones that can provide a relatively firm basis for action – we will end up, paradoxically, both reducing the amount of experience that is available to us, and slowing the rate at which the field as a whole can learn about what works in policing.

I set out this argument below in the form of a skeptical commentary on evidence-based policing. For clarity and concreteness, I use Lawrence Sherman’s (1998) paper on evidence-based policing as the point of departure. I understand that the idea of evidence-based policing is broader than the approach suggested by Sherman, and that there are many others who propose different ideas. But Sherman’s paper provides such a convenient foil for my arguments, and is so clearly an exemplar of the commitments that lie behind a larger movement for evidence-based medicine, evidence-based education and evidence-based policing that it is irresistible to use it as a target. I hope both Larry and others working on evidence-based policing will forgive the distortion that this focus gives to my discussion. I only hope that my arguments will have merit even when considered against a wider and more nuanced idea of evidence-based policing.

“Scientific evidence”

Let’s begin where Sherman does: with the apparently incontrovertible claim that “police practices should be based on scientific evidence about what works best.” Anyone who believes in science as an important driver of human progress has to accept this idea. Since that includes virtually all police researchers, and no small number of police practitioners, Sherman starts off with a large and sympathetic audience. The difficulties arise when we begin to think hard about what we mean by “scientific evidence,” by “police practices,” and by “works best.”

Take, first, the issue of what constitutes “scientific evidence.” In Sherman’s view, too much current police practice is based on nothing more than “local custom, opinions, theories, and subjective impressions” (1998: 6). Current practices of policing are viewed as “unsystematic ‘experience’” that provides the raw material for, but not the true essence of the kind of “scientific evidence” that would put policing on a stronger intellectual footing (Sherman 1998: 4). What lies between this “unsystematic experience” guided by “local custom, opinions, etc.” is the rigorous examination of those practices through scientific methods – ideally, randomized trials of particular methods used in dealing with “repetitive circumstances.” Thus, Sherman sets up a simple dichotomy: mere “custom, opinion and theory” versus “scientific facts.”

The problem is that this is a false dichotomy. The options that Sherman describes anchor the ends of a continuum of scientific methods that can

be used to test the efficacy of police methods; not a stark choice between superstitious tradition on one hand, and scientific certainty on the other. Science, as a human disposition and as a method, has never been strictly limited to randomized experiments; it has always included many more different types of investigations to acquire and use knowledge.

Science begins with curiosity about the size and character of some phenomenon in the world (for example, different kinds of crime). It continues with the development of some (as yet untested) ideas about the causal factors that are shaping the size and character of the phenomenon of interest (such as levels of poverty, or the development of a culture of violence). It proceeds with measurements made through particular instruments specially designed to make the link between a theoretical concept, and an empirical reality. It continues with the use of statistical techniques designed to discover the relationship, if any, between changes in the phenomenon being observed and changes in the variables that are thought to be causing the change.

It is worth noting that this basic method of science (interest in a phenomenon, speculation about causes, observation and measurement, analysis of differences) can be used not only to discover the "natural causes" of a phenomenon as it exists in nature, but also to investigate the consequences of a specifically imagined human intervention designed to alter the phenomenon that is observed. Thus, scientific methods can, in principle, be used to discover not only the causes of crime, but also the crime control effectiveness of particular interventions made by police departments. The search for causes is often described as basic research, while the search for remedies and cures is described as applied research.

But the distinction between basic and applied research should not blind one to the fact that both basic and applied sciences depend on exploring relationships between so-called "dependent variables" on one hand, and "independent variables" on the other.¹ The major difference is that when we are trying to explain the causes of crime, the independent variables are social structural and population characteristics that we hypothesize generate a certain amount of crime. When we are trying to explore the impact of a crime control intervention, the independent variable of particular interest is the intervention to be evaluated.²

It is also worth noting that the general form of "observational studies" described above (in which a phenomenon is observed in nature being affected both by natural cases and certain kinds of policy interventions) are generally only capable of showing a *correlation* between changes in the independent variables on one hand, and the dependent variables on the other (Winship and Morgan 1999). They cannot demonstrate *causation*

with any high degree of reliability. To show *causation*, nature has to be manipulated a bit more carefully (Heckman 1997).

Specifically, to demonstrate that a particular hypothesized factor actually caused an effect to occur; or to show convincingly that a particular effect was the result of a particular policy intervention, we have to be able to compare two situations that are identical in all respects except for the fact that one situation was affected by a cause, or received a particular intervention, and the other did not so that any observed difference in results can be attributed to that causal variable, or that particular intervention. Repeated enough times, if we find that the treated situation produces a different result than the untreated situation, we can conclude with a high degree of confidence that the intervention produced the result we observed. It is this simple fact that makes randomized trials the "gold standard" for reliably inferring causation whether looking for natural causes or estimating the impact of an intervention, since randomization of conditions is the best method of ensuring that the two situations that are being compared are similar (Campbell 1969).

While everyone agrees that randomized trials are the best way of creating confident knowledge about the existence and size of an effect of police practices on levels of crime, an important question remains about how much we should rely on randomized trials, and how much on other, lesser means. This would not be an important question if we thought we could produce randomized trials to test every police practice. But there are many things that stand in the way realizing this particular dream – including the cost of such experiments, the technical difficulty of executing them, and various ethical cautions that arise when we experiment with the fates of human beings (Cook and Campbell 1979).

If, realistically, we cannot expect to test every police practice through randomized trials, then the issue of whether it makes sense to use some of the lesser methods of science (simple inquiry, logical thought, direct observation of actions and apparent results, program evaluations carried out through statistical analyses rather than randomized trials) becomes important. Put somewhat differently, both the research and the practice field in policing face the important question of how far down the path of scientific sophistication they should go in their combined efforts to establish a firm experiential and empirical basis for policing. More provocatively put, they have to decide what to do with the knowledge that lies between mere opinion on one hand, and results established through randomized trials on the other.³

In my preferred vision of "evidence-based policing," the standards of what constituted "research," or "evidence," or "scientific knowledge" would be more open and flexible than in Sherman's conception.

Obviously, one would want to get to the “gold standard” with as many important police practices as we could. But the central aim of a movement for “evidence-based policing” wouldn’t be simply to push some arbitrarily selected police practices to the promised land of randomized trials, it would be to subject as many police practices as possible to increasingly stringent tests of their efficacy through methods that seemed appropriate and feasible.

In this conception of evidence-based policing, it would be counted as a gain if one simply laid out the common sense logic that lay behind a particular police practice, and subjected it to what I would describe as “the giggle test” that measured the common-sense plausibility of a claim of efficacy.⁴ It would also count if one gathered facts about particular kinds of crimes and problems that police face that could suggest innovative ideas about how that sort of crime, or that kind of problem might best be addressed. It would be counted as a bigger gain if one actually went ahead and captured the experience a police department produced by recording and measuring what the department did, and what happened later, possibly but not necessarily as a consequence of the department’s actions. These ideas are key elements in the problem-oriented policing model proposed by Herman Goldstein (1990). It would be counted as an even bigger gain if the experience produced in one domain was checked with naturally occurring experiences in other domains that could be used as “quasi-control groups” (i.e., rough equivalents to the kind of planned variation that one would introduce in the design of randomized experiments). And so on.

The point is to recognize that different pieces of evidence about efficacy come with different degrees of weight and credibility. The task is not only to produce evidence that has the particularly heavy weight associated with randomized trials; it is also to produce and responsibly use evidence that is less strong than that produced by randomized trials.

Widening the range of acceptable methods, and learning how to calibrate their differential weight in policymaking is important for at least three different reasons. First, sometimes the lesser methods produce important knowledge. This occurs when the effect of an intervention is so large, and so obvious that one doesn’t really need a fine-grained statistical analysis to “tease out” the effect. Second, sometimes the lesser methods are the only practicable methods to use. There is not enough time to wait for the result, or not enough money to pay for the experiments, or not enough capacity to structure the world into the form demanded by randomized experiments. Third, many more people can get involved in the effort to learn what works if we work hard on the cruder parts of science such as developing “thick descriptions” of the phenomenon of interest,

reasoning about interventions that might make sense, and remembering to look as closely as one can at what happened when one tried a particular intervention. Developing these basic scientific orientations among a mass audience can provide not only a valuable base for learning, but also help to create a strong base of support for the use of more sophisticated methods. The alternative approach – allowing only a few elite folks to carry out these investigations – raises doubts and generates resistance among those for whom the research is ostensibly being conducted (Hartley, Bennington, and Binns 1997; Hartley and Bennington 2000).

Police practices

Take next the question of what constitutes a “police practice” that could be evaluated through scientific means. Again, the scientific method works best (most reliably, most simply, most inexpensively) when the phenomenon being examined is causally simple, easy to observe, and occurs often. Such conditions often occur in manufacturing processes where physical inputs are converted to physical outputs through a well-defined technological process engineered to produce a specific result. That is the reason that statistical control systems work as well as they do in manufacturing operations.

In principle, one can always characterize some specific police practice as a kind of manufacturing technology that takes inputs (human labor, professional skill, assistance from victims and witnesses, forensic support, etc.) and produces a potentially valuable output (e.g., arrest of a murder suspect with enough evidence to support an effective prosecution). All one has to do is draw a “black box” around the inputs and the outputs, and carry out a statistical analysis of the relationship between the inputs and the outputs to identify the relevant “production function” (see e.g., Pindyck and Rubinfeld 2001). Look at enough homicide investigations and burglaries in this way, and one can identify the factors that contribute to the solution of the case (Greenwood, Chaiken, and Petersilia 1977). Similarly, one can imagine that we have particular approaches to particular crimes – such as arresting offenders for domestic violence – and experiment with the use of this technique to see whether it reduces the recurrence of these events, and thereby results in a reduction in these kinds of crimes.

In principal, at one higher level of abstraction, one can also think of a police department as a whole as the cumulative sum of these particular practices, brought out by the personnel of the department at the right time when triggered by an external demand for performance. Thus, one can conceptualize a police department as a kind of manufacturing

or service facility designed to prevent crime in a particular jurisdiction, draw a box around the police department as a whole, and ask about the degree to which it is effective in reducing crime throughout its particular jurisdiction.

While this is fine in principle, when one begins to try to scientifically design and evaluate police practices, some important difficulties emerge. The first comes in defining a particular "police practice" that is worth scientifically evaluating.

At the outset of modern police research, this was not a particularly difficult problem. The "professional model" of policing relied on only a few large "practices" that were thought to be just and effective in dealing with all kinds of crime. The "logic model" of professional policing hypothesized (reasonably, I think) that all kinds of crime could be prevented and controlled through a few general methods of policing: namely, patrol, rapid response to crime calls, and retrospective investigations (Moore 2002). By threatening offenders with arrest, such methods could be expected to prevent and control crime through general deterrence. By producing arrests that led to convictions, such tactics could prevent and control crime through the additional mechanisms of specific deterrence, incapacitation, and (sometimes) rehabilitation.⁵

These police practices – patrol rapid response, retrospective investigation – represented large, significant targets for evaluative research (including both observational studies and randomized experiments) since they encompassed a great deal of what police departments actually did to control crime. What that research showed, however, was that these methods had much smaller effects on levels of crime and fear in the community than we initially thought (Kelling, Pate, Dieckman, and Brown 1974; Greenwood *et al.* 1977; Spelman and Brown 1984). Nor did they result in the apprehension of offenders as often as we thought desirable (Greenwood *et al.* 1977). And they were not able to cope very well with crimes that were invisible to the police because they did not necessarily happen in public, or produce a victim or witness who could alert the police (Moore 1983).

One of the most important effects of these studies was to encourage the police to search for more specialized practices designed to deal with particular kinds of crime. And it is this approach to dealing with crime that recently won the endorsement of the National Academy of Science's Panel on Policing (Committee to Review Research 2004). But an important consequence of this development for those who would like police methods to be carefully evaluated is that the more specialized focus on particular kinds of crime implied a dramatic increase in the number of specific police practices that had to be developed and evaluated. Instead of evaluating a

few police practices that used most police resources and were considered robust in dealing with all manner of crime, the research task now became the evaluation of a much larger number of more specialized practices designed to deal with more particular kinds of crimes. We can, and have, looked at how the police respond to domestic violence, to serial killers, to robberies, to street-level drug dealing, to vandalism and vagrancy, and to mentally ill individuals who are creating disturbances, or threatening to commit suicide, and so on (Sherman 1992; Plotkin and Narr 1993; Rossmo 1995).

Once we begin to think that it is wrong to look at a general method for dealing with a homogenous problem called crime, and to rely instead on a wide array of more particular methods for dealing with a highly heterogeneous set of crimes, the challenge of evaluating police practices begins to resemble the problem of investigating the huge variety of medical practices used to cope with different kinds of illnesses. In medicine, instead of imagining that there is a general cure for all disease, and that hospitals are the organizations that do only that general thing to protect us from disease, we have begun to distinguish many different kinds of diseases, each with their own proper response, specially tailored to the unique circumstances of an individual. We have also come to view the practice of medicine in hospitals as bundles of these particular practices that are hauled out by doctors and nurses when particular cases require them. Perhaps police departments should be seen as collections of professional police officers organized in a particular organization that brings out particular treatments for particular cases as needed. Ideally, then, the knowledge base of policing would consist of the knowledge of what particular interventions work with what particular diseases, and the adaptation of those particular interventions to the particular condition of a given circumstance – just what the idea of problem-oriented policing prescribed for the field.

Unfortunately, just as in medicine, only a portion of what is now thought to constitute the "best practices" in policing with respect to particular kinds of crime have actually been tested in randomized trials. That means that the police are still operating in the dangerous situation of relying on untested best practices. On the other hand, it also suggests that all we have to do is to increase the rate of experimentation so science can catch up to our current practices. That is what Sherman seems to believe. But I think there is a deeper problem here, and the difficulties that medicine faces in evaluating its own practices should provide a cautionary tale to policing as it seeks a firmer basis for its professional practice.

One part of the difficulty has to do with knowing how to divide up the world of crime problems into meaningful categories for which

particular best practices can be developed. So far, much of that work seems to be guided by common sense – not some deductive logic. For example, we started out distinguishing domestic violence from other kinds of aggravated assault because we thought it was a “different” kind of problem in at least the following ways: (1) we were less likely to hear about it from victims; (2) the victims would be less cooperative in supporting arrests and prosecutions; (3) there were often repeat offenses, so that each offense could be used as an occasion to do something that would make the next offense more or less likely to occur, and more or less serious; (4) we had to be worried about the effect of the intervention we made not only on the likelihood of future violence against the victim, but also on the impact that the intervention had on children living in close proximity to and dependent on the victims, and so on. All these features of domestic violence seemed to make it a special category of crime that deserved its own diagnosis and its own treatment (Chalk and King 1998).

Yet, as we have gone deeper into the exploration of methods for dealing with domestic violence, two surprises have occurred that raise doubts about categorizing domestic violence as a unique kind of crime. On one hand, we have found that other kinds of violence share some important characteristics with domestic violence. For example, we have discovered that many homicides and aggravated assaults among young men occur in the context of continuing relationships that are as criminogenic as those between domestic partners (Kennedy, Piehl, and Braga 1996). They are rivals for turf, for the leadership of a gang, or for the affections of a young woman. Violence they commit on one another will create another round of violence as families and friends of the victims seek revenge. In short, gang murders are similar to domestic violence in that it is a continuing relationship among individuals that seems to stimulate and focus the violence.

On the other hand, we have learned that there are many different kinds of domestic violence cases that seem to respond differently to different kinds of interventions. Indeed, Sherman (1992) illustrates this phenomenon in his own account of what happened with the experiments in policing domestic violence. We began with the idea that there was a relatively homogenous problem called “domestic violence,” and the challenge was to determine whether a particular method – mandatory arrest – was successful in dealing with this general problem. What we learned (after more than \$10 million in experimental studies) was: (1) that arrest reduces domestic violence in some cities but increases it in others; (2) that arrest reduces domestic violence among employed people but increases it among unemployed people; (3) that arrest reduces domestic violence in the short run but can increase it in the long run; and (4) that the police

can predict which couples are most likely to suffer future violence, but our society seems to value privacy too highly to allow effective preventive action.

These observations suggest that the search for effective crime control methods may not go on as systematically or as scientifically as we might wish. While we could follow the approach of differentiating different kinds of crime, and looking for the best method of dealing with each particular kind of crime, it is not at all clear how such categories of crime should be constructed, nor how many will have to be investigated. The more particular our characterization of crimes, the more experiments we will have to run. On the other hand, we could start with a generalized idea of crime, and find a method that would be effective in dealing with all kinds of crime. Unfortunately, we tried that, and produced only mixed results. Perhaps our best chance is the method suggested by Goldstein (1990): namely, to learn as much as we can about the likely causes and potential points of intervention in dealing with particular crime problems through the development of a thick description of a certain class of crimes, and then use common sense to imagine and test interventions to see if they work. That clinical (as opposed to scientific) approach might be both necessary and sufficient to help us develop the knowledge we need to deal with a highly differentiated crime problem, just as it was necessary and sufficient to give us most of the means we now rely on in medicine.

If it is hard to evaluate any particular police operation focused on a particular problem, imagine how much harder it is to evaluate the performance of a police department as a whole. Of course, one could rule out the evaluation of a whole department’s performance as not the kind of “practice” Sherman had in mind. But Sherman seems inclined to think that the overall performance of an entire department could be viewed as a kind of “macro-practice” that could be evaluated scientifically as well as the kind of “micro-practices” we associate with the “best practices” that exist for dealing with particular kinds of crime.⁶ He suggests this inclination by explicitly introducing the analogy to manufacturing organizations on one hand, and by claiming that departments as a whole could be evaluated, motivated, and guided by scientific evidence of what works in policing, on the other. In both moves, he changes the definition of the police practice to be evaluated from a specific operational program designed to deal with a specific kind of crime to the operations of the police department as a whole.

When we focus on the “scientific evaluation” of the performance of a police department as a whole rather than the efficacy of a particular procedure in dealing with a particular kind of crime, many things change.

One key difference is that the kinds of police practices that need to be evaluated change. Police practices observed at the organizational level of the police would include all the specific *operational* practices used by the department to deal with particular crimes as defined above. But police practices observed at the organizational level would include the *administrative* practices that the police use to manage themselves; the methods they use to recruit and train personnel, the methods they use to motivate their line commanders and individual officers to achieve organizational goals, the methods they use to minimize corruption and abuses of force, and so on. These administrative practices might even include the arrangements the department made to evaluate their own operational practices (Moore, Sparrow, and Spelman 1997; Moore 2003).

A second key issue, however, is that while effective crime control is certainly one thing that citizens want from police organizations, it is also quite clear that citizens want other things from their police as well (Moore 2002). They want to have a subjective sense of security in their streets that may be somewhat independent of the objective risks of victimization. They want certain kinds of services from the police that make their individual and collective life better.

But the problem continues. Often specific police activities and operations produce results that register on many different objectives of policing – not just one. The DARE Program, for example, has long been evaluated as though its most important justification is the impact such a program would have on future levels of drug use by students exposed to the program (Esbensen, Friend, Taylor, Peterson *et al.* 2002). But the most important practical effect of the DARE Program might be to build relationships between the police on one hand, and parents and youth on the other, that will allow them to be more effective in dealing with all kinds of problems both encountered and created by these individuals in the future.

There may also be synergies (positive or negative) in the activities of police viewed across the department as a whole. For example, the strong work of an officer who was focused on building community relations in a particular neighborhood could be undermined in an instant by a drug raid carried out by a centralized unit that ended up in the wrong apartment arresting the wrong person in a way that terrorized rather than reassured the neighborhood about the competence and intentions of the police. In short, policing may simply be too complex an endeavor to ever be sorted out – just as it would be too complex to sort out the operations of a public health system that included not only emergency rooms, but also long-stay hospitals, neighborhood clinics, and public health immunization and well-baby programs.

“What works?”

Take, finally, the question of what we mean when we say that a particular police practice “works.” The obvious next question is: “Works to do what?” A question that follows closely after that is: “Works at what cost, and with what unintended and unexpected side effects?” In answering these questions, one is gradually constructing an analytic framework that one could use normatively to decide whether a given police practice was worth continuing.

This is importantly an empirical question, of course. To determine whether a practice is a good one or not, we have to be able to describe the effects it produces in the world. And that is a profoundly empirical issue.

But the construction of an analytic framework for *evaluating* a given police practice also requires a *normative* stance. To identify an effect of a program as something that would be worth noticing in any attempt to evaluate its value to the wider society is inherently a normative enterprise. The effects that we consider when evaluating a program are important precisely because they have normative significance to the world.

For example, we could observe that police crackdowns on gang members could reduce gun violence and fatalities in a city. That alone might seem to make a strong case for engaging in more police crackdowns on gang members. But suppose those police crackdowns produced other effects as well such as increased resentment and fear of the police in the neighborhoods in which the crackdowns occurred. There would be an empirical question to be answered about whether this was a real effect of the police crackdown. There would be a normative question about whether such an effect should be counted when we were considering the overall value of the crackdowns as a police response to a social problem. If we thought this effect was a plausible one, and that it was normatively significant, then the analytic framework for measuring the impact of the crackdown would have to look at the effect it had on local residents’ attitudes toward the police as well as on the level of gun violence. A new dependent variable would have to be introduced into the analysis. This is an example of an important conclusion drawn by the National Academy of Sciences Panel on Policing. They concluded that police practices had to be evaluated in terms of their impact on the legitimacy of the police as well as their crime control cost effectiveness (Committee to Review Research 2004).

Often, we do not feel obliged to say what we mean by a police practice that works, because we assume that we know what the point of police practices should be: namely, to control and reduce crime. We assume

that all police practices should be evaluated only in terms of this single dimension of performance. But I think it is clear that both micro police practices and the overall operations of macro police institutions need to be evaluated in more dimensions than their crime control efficacy. Some of these added dimensions focus on the output or value side of police operations. For example, we might be as interested in the capacity of a given police practice to reduce fear, and reduce the burden of self-defense as well as on their capacity to control a particular crime. We might also be interested in the capacity of a given police practice to produce the kind of justice we associate with both calling offenders to account, and protecting the rights of citizens as well as efficacy in dealing with crime. We might even be interested in the impact that the police practice has on perceptions of police responsiveness and fairness as well as effectiveness in dealing with particular crime problems (see e.g., Moore 2002).

Other dimensions of evaluation focus not on the value of the outputs, but on the costs of mounting the operation. That includes monetary costs for sure. But it also could include costs associated with abridging individual rights, or exacerbating a sense of unfairness in the way that the police do their business. These costs could show up as monetary costs (when the police were successfully sued for wrongful conduct); or they could show up in terms of reduced effectiveness as disillusioned citizens stopped cooperating with the police in the production of justice and security; or they could show up in nothing more than the continued disgust and disappointment that citizens feel when they think their tax dollars and liberty are being abused by an organization they would like to be able to trust.

To say that a police practice, or a police organization, works, then, is to make a set of claims about not only what effects a police department produces, but also what a political community that authorizes the police department does or should want. This means that we have to admit both values and politics into the construction of the analytic framework we use to evaluate police practices and police departments as a whole. What effects of a police practice should be considered important for evaluation purposes is not, strictly speaking, a value-free scientific enterprise. A fact taken by itself may be value neutral. And in trying to develop a fact it may be important to push one's values to the side as best one can. But to develop a fact that is useful in *evaluating* police practices, one has to construct a value framework that makes the fact interpretable as an evaluation of whether something is good or bad. And that is inevitably a value question.

How does policing improve?

In the end, both police researchers and police practitioners have to concern themselves with the way in which policing can be improved. An important part of that effort depends on developing more reliable knowledge about what works. And that, in turn, depends on deploying the methods of science to help us separate false claims of efficacy from true claims, or to gradually increase our confidence in the methods we are using to accomplish the goals of policing. Thus, science, and particularly the use of science in carrying out randomized trials of police practices, could be the "rate determining step" in determining the rate of police improvement.

If this is true, then the principal responsibility for improving policing seems to lie primarily in the research community. It is there that police practitioners must look to find out what is known to work. It is there that police practitioners must turn to develop the guidelines they use to instruct their officers to do their work. And it is from that source that any new doubts about the efficacy of old methods, or any new ideas about the proven efficacy of new methods must come.

That, at least, seems to be the core of the vision of how to improve policing that is contained in the idea of evidence-based policing: improvement through science, and the work of scientists.

I hope I can be forgiven if I suggest that this puts a bit too much of the load on science and scientists, and that it uses the commitment and knowledge of the police profession a bit too little. Science cannot become useful in the practical world of policing unless the practitioners embrace some aspects of science as an important part of the way they do their work. We have to create the equivalent in the profession not only of academic researchers and clinicians, but also of clinical professors, and reflective practitioners; and they have to learn to work together in a kind of intellectual and practical partnership to solve concrete problems as best they can – not compete with each other over whose knowledge is more authoritative.

Science alone cannot answer important questions that are central to the development of effective clinical practice. Science, by itself, cannot tell us how to divide up the world of practical problems we see in front of us in a way that is most amenable to their solution. Science, by itself, cannot necessarily suggest the particular intervention that should be tried to deal with a given problem. Science, by itself, cannot tell us in what terms we ought to evaluate the interventions we make. For all this work, we need the experience of practitioners as well as the experiments of science.

If our vision of evidence-based policing going forward is one that embraces these aspects of science, I am all for it. If our vision is one that seeks to privilege a certain kind of science and a particular kind of scientist, I am against it. The world of crime and policing is far too important, far too complex, and far too urgent to leave entirely in the hands of scientists. We need a great deal of practical wisdom as well as a rigorous and responsive science to move the field forward.

NOTES

1. Karl Popper describes what is here called the dependent variable, the *explanandum* – the thing to be explained, and the independent variables, the *explanans* – or the thing that explains the *explanandum* (Popper, 1902).
2. James Q. Wilson (1985) has made some important observations about the difference between criminology on one hand, and the effort to find important means for controlling crime on the other.
3. Howard Raiffa (1924) has developed some systematic ways of thinking about using imperfect information in practical decisions. He recommends a Bayesian approach in which the decisionmaker begins with more or less strong prior probability estimates of the likely effects of a given policy. He then updates those prior probability estimates as new information comes in through observation and experimentation. Since each piece of information has a different weight, each piece of information can be brought in to improve the estimate. We don't have to discard imperfect information; only discount its weight.
4. This is often accomplished as a necessary part of any program evaluation. Wes Skogan (1990) has demonstrated the importance of developing the "logic model" that connects a planned intervention to a desired effect as an important step in carrying on a program evaluation. Often, once one goes through this particular discipline, one can see errors in thought and planning even before one gets into the field. That is what I mean by the "giggle test" – the use of logic, common sense, and a sense of proportion to test the very plausibility of an idea before one goes to the trouble of trying it out and carefully evaluating it. In my experience, many police practices would be improved by taking this very first step in science.
5. These basic ideas remain strong in policing because they align both with common sense, and with certain ideas of justice; namely that it would be just, as well as practically useful, to call offenders to account for their crimes, and to put them in prison. They show up these days in police tactics that encourage aggressive preventive patrol focusing on disorder offenses.
6. Note that thinking about the idea of a police department as a macro practice could include two quite different ideas. One is that the department had a general approach to dealing with crime that worked. That could be located in a particular robust operational procedure such as aggressive preventive patrol. Or it could be located in a particular administrative approach such as Compstat. Or it could be located in a particular organizational strategy such as community-oriented or problem-oriented policing. Or, it could be in some combination of the police department having in its repertoire a wide variety

of responses that it uses, along with some principles that help it decide which particular procedures to use.

REFERENCES

- Campbell, D. (1969). Reforms as experiments. *American Psychologist* 24, 409–429.
- Chalk, R. and King, P. (eds.). 1998. *Violence in families: assessing prevention and treatment programs*. Washington, DC: National Academy Press.
- Committee to Review Research on Police Policy and Practices. (2004). *Fairness and effectiveness in policing: The evidence*. Washington, DC: The National Academies Press.
- Cook, T. and Campbell, D. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston: Houghton Mifflin Company.
- Esbensen, F., Freng, A., Taylor, T., Peterson, D., and Osgood, D. (2002). National evaluation of the gang resistance education and training (G.R.E.A.T.) program. In W. Reed and S. Decker (eds.), *Responding to gangs: Evaluation and research*. Washington, DC: National Institute of Justice, US Department of Justice.
- Goldstein, H. (1990). *Problem-oriented policing*. Philadelphia, PA: Temple University Press.
- Greenwood, P., Chaiken, J., and Petersilia, J. (1977). *The investigative process*. Lexington, MA: Lexington Books.
- Hartley, J. and Bennington, J. (2000). Co-research: A new methodology for new times. *European Journal of Work and Organizational Psychology*, 7, 1–16.
- Hartley, J., Bennington, J., and Binns, P. (1997). Researching the role of internal change agents in the management of organizational change. *British Journal of Management*, 8, 61–73.
- Heckman, J. (1997). Instrumental variables: A study of implicit assumptions used in making program evaluations. *Journal of Human Resources*, 33, 441–462.
- Kelling, G., Pate, T., Dieckman, D., and Brown, C. (1974). *The Kansas City preventive patrol experiment: A technical report*. Washington, DC: Police Foundation.
- Kennedy, D., Piehl, A., and Braga, A. (1996). Youth violence in Boston: Gun markets, serious offenders, and a use-reduction strategy. *Law and Contemporary Problems*, 59, 147–196.
- Moore, M. H. (1983). Invisible offenses: A challenge to minimally intrusive law enforcement. In G. Kaplan (ed.), *ABSCAM ethics: Moral issues and deception in law enforcement*. Washington, DC: Police Foundation.
- (2002). *Recognizing value in policing*. Washington, DC: Police Executive Research Forum.
- (2003). Sizing up compstat: An important administrative innovation in policing. *Criminology and Public Policy*, 2, 469–494.
- Moore, M. H., Sparrow, M., and Spelman, W. (1997). Innovation in policing: From production lines to jobs shops. In A. Altshuler and R. Behn (eds.), *Innovation in American government*. Washington, DC: Brookings Institution Press.

- Pindyck, R. S. and Rubinfeld, D. (2001). *Microeconomics* (5th ed.). Upper Saddle River, NJ: Prentice-Hall.
- Plotkin, M. and Narr, T. (1993). *The police response to the homeless*. Washington, DC: Police Executive Research Forum.
- Popper, K. (1902). *The poverty of historicism*. Boston: Beacon Press [1957].
- Raiffa, H. (1924). *Decision analysis*. New York: McGraw-Hill [1997].
- Rossmo, D. K. (1995). Place, space, and police investigations: Hunting serial violent criminals. In J. Eck and D. Weisburd (eds.), *Crime and place*. New York: Criminal Justice Press.
- Sherman, L. W. (1992). *Policing domestic violence*. New York: The Free Press.
- (1998). *Evidence-based policing: Ideas in American policing series*. Washington, DC: Police Foundation.
- Skogan, W. (1990). *Disorder and decline*. New York: The Free Press.
- Spelman, W. and Brown, D. (1984). *Calling the police: Citizen reporting of serious crime*. Washington, DC: US Government Printing Office.
- Wilson, J. Q. (1985). *Thinking about crime*. Revised edition. New York: Vintage Books.
- Winship, C. and Morgan, S. (1999). The estimation of causal effects from observational data. *Annual Review of Sociology*, 25, 659-706.