Lawrence W. Sherman

The Rise of Evidence-Based Policing: Targeting, Testing, and Tracking

ABSTRACT

Evidence-based policing is a method of making decisions about "what works" in policing: which practices and strategies accomplish police missions most cost-effectively. In contrast to basing decisions on theory, assumptions, tradition, or convention, an evidence-based approach continuously tests hypotheses with empirical research findings. While research on all aspects of policing grew substantially in the late twentieth century, the application of research to police practice intensified in the early twenty-first century, especially for three tasks that make up the "triple-T" strategy of policing: targeting, testing, and tracking. Evidence-based targeting requires systematic ranking and comparison of levels of harm associated with various places, times, people, and situations that policing can lawfully address. Evidence-based testing helps assure that police neither increase crime nor waste money. Tracking whether police are doing what police leaders decide should be done may grow most rapidly in coming years by the use of GPS records of where police go and body-worn video records of what happens in encounters with citizens.

Can research help improve democratic policing? For over four decades, numerous colleagues and I have struggled to answer that question. When we began, policing was done very differently from how it is

Lawrence W. Sherman is Wolfson Professor of Criminology at Cambridge University, a director of the College of Policing (UK), Distinguished University Professor at the University of Maryland, and director of the Cambridge Police Executive Programme. He is grateful for comments on earlier drafts by Heather Strang, Peter Neyroud, Daniel Nagin, Stephen Mastrofski, John Eck, Alex Piquero, David Weisburd, Lorraine Mazerolle, Cynthia Lum, and Michael Tonry.

0192-3234/2013/4201-0007$10.00

done now. In the years to come, it will be done even more differently. Yet the trajectory of change is likely to remain consistent: from less research evidence to more and from lesser to greater use of the evidence available. The 50 years covered in this volume bear witness to the rise of evidence-based policing in both the quantity of evidence and its influence on police practices. This essay reports on my ongoing participant-observation study of these changes, for which I have tried to be a primary instigator.

Policing in 1975 was largely delivered in a one-size-fits-all strategy, sometimes described as the "three Rs": random patrol, rapid response, and reactive investigations (Berkow 2011). After World War II, random patrols in police cars were promoted on the theory that police "omnipresence" would deter crime (Wilson 1950). In the 1960s, the advent of three-digit emergency phone numbers turned random patrol into an airport-style "holding pattern" for rapid response, also based on a theory of deterrence, but producing a new theory of organizational action: Albert J. Reiss's (1971) distinction between reactive and proactive actions. What police did when they reacted to a citizen call was not subject to much police agency direction or analysis. The main organizational requirement was to arrive, do something, and leave as quickly as possible. As Reiss reported, his 1966 Chicago study found patrol officers spending 14 percent of their time "reactively" answering such calls and investigating crime, about 1 percent of their time "proactively" stopping people at their own initiative, and 85 percent of their time in unstructured random patrolling. Reactive investigations, when required, were taken over by detectives, who in theory investigated all reported crimes but in practice closed most cases as unsolvable. Cases that were solved relied primarily on the evidence presented to them by crime victims and the patrol officers who arrived first on the scene (Greenwood and Petersilia 1975).

By 1975, the three Rs had become the standard model of urban policing (Skogan and Frdyl 2004, p. 223) across the United States as well as in the United Kingdom and other predominantly Anglo-Celtic cultures (e.g., Canada, Australia, New Zealand). There was almost no targeting of patterns or predictions of crime or disorder, no testing of what worked best to prevent or solve crimes and problems, or much tracking and managing of what police were doing, where, when, and how, in relation to any specific objectives. Most police agencies lacked computers. The few computers in use in 1975 were mainframes de-

signed for dispatching police cars, not analyzing crime data. Police leaders rarely discussed crime at all unless a particularly gruesome crime received substantial publicity.

By 2012, the three Rs were changing into what I describe in this essay as the "triple-T" of targeting, testing, and tracking. While the standard model is far from gone, its resources are increasingly guided by statistical evidence. In the emerging triple-T strategy, both patrol and detective managers had moved toward far greater proactive management of police resources. Although these changes are far more evident in some police agencies than in others, Anglo-Celtic policing is increasingly targeting scarce resources on evidence of large, predictable, and harmful statistical patterns rather than on isolated cases (Spelman and Eck 1989). Compared to 1975, these focused police strategies are far more elaborate and differentiated, choosing from a wider range of priorities and objectives on the basis of extensive data analysis. Police methods have also become far more subject to testing, with evaluation and debate over "what works," the core idea of evidence-based policing (Sherman 1984, 1998). Yet the most evident use of new statistical evidence is the growth of tracking and managing what police were or were not doing in relation to the dynamic patterns of crime and public safety problems.

Examples of the growth of evidence-based policing abound. They include the US Supreme Court's citation of research on tests of restrictions on police shootings (Fyfe 1980, 1982; Sherman 1983), *Tennessee v. Garner* (471 U.S. 1 [1985]), a decision that fostered a major reduction in police killings of citizens in the United States (Tennenbaum 1994). Overall reductions in homicide have been associated with an increased targeting of US police work toward micro-level crime "hot spots" (Weisburd and Lum 2005; Police Executive Research Forum 2008; Braga and Weisburd 2010), where research showed that crime is heavily concentrated (Sherman, Gartin, and Buerger 1989). By 2012, evidence-based targeting for problem-oriented policing (POP) has, since Goldstein (1979) first proposed it, gradually become respected as real police work, not "social work," supported in some agencies with new case management systems. Crime analysts are more likely to identify repeat crime places, victims, suspects, situations, and other patterns to support targeting decisions for patrol, POP, and detectives. Even reactive detective work is more proactive, police-initiated, in gathering new evidence at crime scenes, using scientific meth-

ods to identify suspects and prove cases (Roman et al. 2009). What patrol officers do when responding to crime scenes has become more likely to be a method that has been tested. Where officers are at any moment is more likely to be tracked by global positioning satellite (GPS) systems generating weekly management analysis reports.

The best test of evidence-based policing is whether it has improved public safety and police legitimacy. There is certainly a correlation over time between the rise of evidence and a decline in serious crime, in both the United States and the United Kingdom. Proof that serious crime dropped because of evidence-based policing, however, is more elusive. There is some micro-level analysis of the New York City crime drop that offers a greater basis for causal inference than national trends (Zimring 2012), but it is impossible to rule out other plausible causes that also match the timing of the crime drop. There is also a danger that a sharper focus on crime can undermine police legitimacy, especially since proactive policing is structurally less legitimate than reactive policing with a personal victim's endorsement (Reiss 1971, p. 11). Yet there is no evidence that the rise of evidence in policing has caused reductions in police legitimacy. Indeed, more data about legitimacy have been incorporated into police management as part of evidence-based tracking than ever before. Overall, the vast scale of the rise of evidence in policing leaves it without a fair comparison group, just as it does in medicine (Sagan 1987), leaving causation unknowable. Still, the macro trends in reduced mortality suggest that we should celebrate the rise of evidence in both medicine and policing.

The changes from 1975 to 2012 raise the further question of what Anglo-Celtic policing will look like by 2025. The answers cannot be predicted just by extrapolating the trends of four decades, or even—ideally—by a causal model for prediction (Silver 2012). But by explaining why policing changed as it did from 1975 to 2012, a forecast using the same causal model can identify a range of possible scenarios for policing in 2025. Chief among these will be the possible global influence of the 2012 creation of the first professional body ever charged by a national government with recommending police practices on the basis of continuous review of new research evidence on what works: the United Kingdom's College of Policing (http://www.college.police.uk). This body has tremendous potential to follow the pathway to innovation Johnson (2010) associates with such major advances as the printing press, which was inspired by the wine press: a lateral-thinking

style of adaptation of an idea used in one setting (such as evidence-based medicine) to another (such as evidence-based policing).

The Rise of Evidence. The causal model this essay suggests for these changes is a dynamic system that links external demands on police (to cut crime, cut costs, or build legitimacy) to research evidence on how to meet those demands, supported by the increasing availability of innovative technologies. While external factors drive the availability of both research evidence and technology, how policing responds depends on a fourth factor: the governmentally generated human capital of education and police professionalism. The role of government policies is crucial in setting salaries and recruitment standards that shape the quality of police skills and creativity in general and of police leaders in particular. Those human resources are so influential that the entire causal model can reasonably be said to call for evidence-based professionalism.

Inside this abstract causal model are many human beings struggling hard for change: influential leaders of policing, research, and efforts to combine the two. These include UK Home Secretary R. A. Butler creating a research unit in 1958; future FBI Director Clarence Kelly supporting the Kansas City Patrol Experiment in 1971 (Kelling et al. 1974); Minneapolis Police Chief Anthony V. Bouza supporting the first random assignment of arrest in 1981 (Sherman and Berk 1984); criminologist David Weisburd and his colleagues (Weisburd et al. 2004, 2006; Weisburd, Groff, and Yang 2012) analyzing criminal careers of places in Seattle and Jersey City; New York Police Department (NYPD) Commissioner William Bratton creating COMPSTAT in the 1990s; Assistant US Attorney General Laurie Robinson making evidence from police testing readily accessible at http://www.crimesolutions.gov in 2010; and UK Chief Constable Peter Neyroud's proposal for the evidence-based College of Policing in 2011. What these leaders generally shared—with, most notably, NYPD Commissioner and US Police Foundation President Patrick V. Murphy—was a loosely articulated but deeply held belief that greater use of research could help transform policing into a more legitimate and respected profession.

A profession is widely defined as a public-interest occupation that restricts entry to those who have mastered knowledge and skills needed to provide a particular set of complex services. In the past century, the level of knowledge deemed necessary for a profession has become

linked to requirements for university-based education. It is only a century, for example, since the radical proposal to force doctors to attend university was presented to the United Kingdom's Haldane Commission (Sherman 2011*b*). Similarly, proposals in repeated US commission reports in the 1960s and 1970s recommended that a university degree be a requirement for appointment as a police officer (Sherman and National Advisory Commission on Higher Education for Police Officers 1978). That idea remains controversial and widely unacceptable on both sides of the Atlantic. By conventional definition, then, policing has yet to become a profession.

Yet by the same definition, policing has made great strides toward professionalism. The past 40 years have seen a huge increase in the educational levels of police leaders. Almost half of English chief constables in 2010 had been educated at the Cambridge University Police Executive Program, and three decades of major city chiefs in the United States attended the nondegree Harvard Police Executive Sessions. The role of universities in disseminating police research has been a remarkably successful top-down change process.

Rising levels of police chief education have also fostered more support for research. First led by external scholars, and now increasingly led by police leaders in partnership with universities, the production of new research evidence has helped police respond to external demands for improvement. At minimum, research is perceived to be helpful with such demands. That is especially true when there is internal or external controversy about which police methods are best and whether new technologies are cost-effective.

Even a radical change in structures of police accountability can enhance the value of police research evidence. The 2012 devolution of control over chief constables in England and Wales to locally elected "police and crime commissioners," for example, created a potential for disputes between elected officials and professional police leaders. One possible solution is the professional College of Policing setting independent, evidence-based standards of police practice. Whether this institution, or its cognates in other countries, can create a global demand for more production and use of evidence is a crucial question for the next decade.

Targeting, Testing, and Tracking. Both the demand for, and uses of, research evidence have become clustered around three strategic principles:

1. Police should conduct and apply good research to target scarce resources on predictable concentrations of harm from crime and disorder.
2. Once police choose their high-priority targets, they should review or conduct tests of police methods to help choose what works best to reduce harm.
3. Once police agencies use research to target their tested practices, they should generate and use internal evidence to track the daily delivery and effects of those practices, including public perceptions of police legitimacy.

The growing adoption of those three principles has given shape to what is increasingly called evidence-based policing (EBP). Broadly defined as the use of best research evidence on "what works" as a guide to police decisions (Sherman 1998), the EBP framework was meant to be only a method of making decisions rather than a substantive strategy for police operations. In the 15 years from 1997 to 2012, however, police have shaped their own emergent definition of EBP as a substantive strategy for managing large police agencies on the basis of these three principles.

The rise of the triple-T strategy did not emerge from any theoretical plan to use evidence in such a coherent strategy. The clustering of evidence around three key strategic tasks was driven as much by innovative police leadership as by police scholarship. Its success resulted from the surprisingly rational convergence of police reform with a flood of new research in criminology.

This essay has four sections. Section I discusses the key social trends, institutions, and people who helped shape research aiming to improve policing. The triple-T strategy emerged from the users of that evidence. Section II offers major examples of how much the evidence has grown since 1975 and how widely police have applied that knowledge. These examples address all three triple-T principles, including instances in which evidence and practice have sharply diverged. They are selected to highlight the wide range of what has already been accomplished, even as so much still remains to be done.

Section III analyzes conceptual issues and confusion that hold back progress in targeting, testing, and tracking. The issues are arguably little different from those arising in the development of evidence-based practices in professional baseball, election campaigns, or marketing. They reflect the current struggle between "system I" and "system II"

thinking (Kahneman 2011) in a wide range of endeavors, all of which face the same challenge policing faces—to get much better at what they do in the next decade. In the conclusion I suggest 10 things institutions can do to institutionalize research evidence in ways that foster more fairness and effectiveness in democratic policing.

These are the most important points I make in this essay:

1. The evidence base for police decisions has grown enormously since 1975.
2. Use of that evidence lags behind the knowledge, but use has also grown.
3. Most police practices, despite their enormous cost, are still untested.
4. Targeting and testing require highly reliable measures of crime and harm.
5. Crime rates and counts are by themselves misleading; a crime harm index offers far better evidence to guide police decisions.
6. Police in 2012 used evidence on targeting much more widely than evidence from testing.
7. Research on tracking police outputs remains largely descriptive and incomplete, with great room for using new technologies to improve the quality of evidence.
8. More use of evidence can increase police legitimacy, both internal and external.
9. The State Boards of Police Officer Standards and Training in the United States and the College of Policing in the United Kingdom will be key institutions in making policing more effective, along with the practitioner-led Society for Evidence-Based Policing.

I. The Rise of Evidence for Police Decision Making

The rise of evidence in policing is mostly an Anglo-American story, with events on each side of the Atlantic influencing the other. Both sides began with parallel social trends in the 1960s that challenged police legitimacy. Both the United Kingdom and the United States faced rising crime rates and increasing racial tensions. Authoritative commissions in both set out ambitious agendas for "upgrading" the police, as one think tank put it (Saunders 1970). Both saw policing

become a matter of national politics in unprecedented ways. The different responses of police in each country helped build a special intellectual relationship that was conceptually far deeper than the military collaboration of World War II. One continuing theme on both sides of the Atlantic has been that "they do policing better" on the other side.

A. Setting the Stage

By the early 1970s, Anglo-American policing faced a kind of Christmas tree of ideas for police reform, each idea an ornament unconnected to the others. One idea was that police needed more research, including experiments and demonstration projects, with criminologists and other social scientists becoming part of the police agency workforce focusing on crime prevention (President's Commission 1967, pp. 25–27). There was no consensus about the questions research should answer. There was no particular vision of a new policing strategy to which research could contribute. Yet belief in innovation and evaluation as a strategy in itself became a major force in police reform (Reiss 1992).

That belief led to public and charitable investments in new institutions for police research. The growing role of the British Home Office in police research helped inspire the creation of a similar function in the US Department of Justice. Yet neither government agency's efforts had much impact until after the Ford Foundation in 1970 funded the Police Foundation in Washington to "foster innovation and improvement in American policing." The bold experiments the foundation conducted led government funders to test policing more deeply. In less than a decade, the Police Foundation's and government-supported research discredited the intellectual basis for the three-Rs strategy of random patrol, rapid response, and reactive investigations.

What replaced that strategy was not a coherent new theory. Only in retrospect can we use the triple-T framework of targeting, testing, and tracking to make sense of what emerged. The multicentered work of examining current practices, designing innovations, and evaluating new programs produced something that scholars now call "emergence," the confluence of properties arising from a combination of elements not found in any one of the elements (Johnson 2001). The essential new property is the capacity to lead police organizations with dynamic evidence rather than static doctrine. This property arose initially from the destruction of the prevailing three-Rs model. That, in turn, opened

the door to considering the implications of basic research provided by an earlier generation of thinkers.

The most important basic research had documented the existence of police discretion to choose different strategies rather than being handcuffed by the three Rs. Most notably, Reiss's (1971) conception of police work as divided between proactive and reactive discretion stimulated much creative thinking. At the same time, Banton's (1964) research in Britain and Wilson's (1968) research in the United States showed how the discretionary choices of police organizations differ on the basis of their political and cultural environments, despite their sharing a common legal system. These insights expanded the concept of police discretion from the level of case-by-case to agency-by-agency decision making and helped to set the stage for the emergence of triple-T. But that new, evidence-based strategy took almost four decades to emerge from the accumulation of new research evidence. What happened was a process of "prosumption" (Tapscott and Williams 2006) in which the producers of elements of the triple-T strategy were simultaneously its consumers.

B. Testing

The three-Rs strategy was intellectually discredited after three major efforts were made to assess the effects of key elements. These studies did not adequately support the conclusions attributed to them. They did not show that random patrol had no effect on crime, that rapid police response had no deterrent effect, or that detectives made no contribution to detection, prosecution, and general deterrence of crime. Regardless of what the research evidence really did show, those conclusions became widely accepted by police leaders and police scholars, then and now.

1. *Random Patrol.* The Police Foundation's Kansas City Preventive Patrol Experiment (Kelling et al. 1974) launched the rise of evidence-based policing. The first attempt to undertake a scientifically controlled test of the effects of patrol staffing levels, it proved that bold experiments were possible in policing. The experimental design of withdrawing patrols from five patrol beats (and doubling it in five others) was stunning and unprecedented. That the "sky did not fall" made the world safe for further bold experiments. The reported conclusion—that frequency of police patrols did not affect crime—opened many minds to think more critically about police strategy. The experiment's

leadership by a police chief who was subsequently appointed FBI director also suggested that research, even with negative results, could be good for police career advancement.

It is unfortunate that the Kansas City experiment did not actually support its conclusion. The limitations of the experiment's research design have been extensively reviewed elsewhere (Sherman 1986, 1992*b*; Sherman and Weisburd 1995). The main problem is that the actual frequency of patrols was unmeasured and may well have been identical in all beats because of reactive responses to calls (Larson 1976). Moreover, there were large differences across treatment groups in certain crime rates but too few beats to call the differences "statistically significant." But the most important critique has never been published before: that random patrol was not compared to any other pattern of patrol, so there was no evidence showing that random patrol did not "work" relative to other patterns of patrol, such as concentration on crime hot spots. Policing still needs a strong randomized trial comparing random patrol to hot spots patrol.

Nonetheless, the conclusion that random patrol did not work was widely accepted. Tens of thousands of police officers lost their jobs in the aftermath of the study, which came coincidentally just before a financial crisis in many US cities. None of that stopped the widespread use of random patrol. But it did help drive a research agenda seeking alternative police strategies.

2. *Rapid Response.* The theory that marginally faster response times would catch and deter more criminals was effectively falsified by a National Institute of Justice–funded research project led by staff of the Kansas City Police Department (1977). The study reported that it was necessary to divide crimes into victim-offender "involvement" crimes (e.g., robbery, assault, rape) and after-the-crime "discovery" crimes (e.g., burglary, car theft). It then focused response time analysis on involvement crimes, with "response time" including three time periods: crime occurrence to calling the police ("reporting time"), police receipt of call to dispatch ("dispatch time"), and "travel time" of police from receipt of dispatch to arrival at the scene. Using systematic observation methods and interviews of victims, the Kansas City study found that there was no correlation between response-related arrest probability and reporting time once the reporting time exceeded 9 minutes. The average reporting time for involvement crimes was 41 minutes (Kansas City Police Department 1977, vol. 2, pp. 23, 39). Replications of the

reporting time segment in other US cities found similar results (Spelman and Brown 1981).

What the research did not show is that a capacity for rapid response had no effect on crime. The Kansas City study was not an experiment comparing police agencies or areas with very rapid responses to similar agencies or areas with slower responses, or even no response. Much remains to be learned about the effect of any rapid response capacity compared to none, the latter being the state of police services in much of the world, from India to parts of rural America.

What mattered historically was the growing recognition that there were better things police could do with their time than just wait for calls. While the research did not prove that point, it helped open the minds of police leaders and scholars about what those other things might be.

3. *Reactive Investigations.* The view that detectives "solve" crimes that are reported to them was strongly rejected by another National Institute of Justice (NIJ) project, this one conducted by the RAND Corporation (Greenwood and Petersilia 1975). This report examined the value detectives add to the information that was in the record at the end of a preliminary investigation by the first responders to a crime, usually uniformed patrol officers. The conclusion was that detectives rarely uncovered new evidence that made a difference in solving the crime—contrary to a century of detective fiction.

Here again, the evidence is thin in relation to the conclusion. An experiment comparing cases prosecuted without detective work to cases prosecuted after detective work would be a strong test of the "no-effect" hypothesis. But what mattered was that yet another sacred cow was wounded. For many police leaders, this completed the well-justified execution of the three-Rs strategy, which legitimated trying alternatives.

These three studies created a strong appetite for more experiments and support for funding them from influential scholars. Franklin E. Zimring, James Q. Wilson, Albert J. Reiss Jr., and others shaped several National Academy of Science reports recommending more NIJ funding for randomized experiments (Zimring 1976; White and Krislov 1977). These reports led to NIJ funding streams for most of the more than 100 tests of police practices listed online by Lum, Koper, and Telep (2010), many of which were funded in 1983–89 by James K.

Stewart, the only NIJ director to have served as an operational police officer and leader.

C. Targeting

Important as these tests were, they could not answer the more fundamental question: how can police resources be allocated more effectively? Once the idea of proactive policing became widely understood, it sparked a revolutionary insight in targeting that Goldstein (1979) called problem-oriented policing (POP). For reactive investigations, the RAND report also led to a revolutionary question: can detectives target which cases they can solve by predicting the likely outcome before they begin working on each case?

1. *Problem-Oriented Policing.* After decades of observing police operations for a major study of police discretion (LaFave 1965) and assisting one of the first university graduate police executives in history (O. W. Wilson) as police superintendent of Chicago, Herman Goldstein offered a revolutionary new targeting strategy for policing. Describing a case-by-case response to events as a myopic failure to see larger patterns, Goldstein (1979, 1990) recommended that police invest more time in treating the causes of those patterns rather than their symptoms. POP is more than a targeting strategy, however, since it also develops ideas for the content of police actions to deal with targeted problems. Yet the tactical content of POP is appropriately broad and highly dependent on good diagnosis and good research on what action would work best. The major breakthrough of the strategy is the injunction to first look for patterns—proactively.

2. *Crime and Harm Concentrations.* It was therefore no accident that Reiss, the person who first used the term "proactive" in a policing context, suggested that I study the distribution of all reactive police responses to reported crimes and disorder by each and every address in a major city. The result was our discovery of tiny, micro-level crime hot spots (Sherman 1987; Sherman, Gartin, and Buerger 1989), in contrast to conventional concepts of much larger high-crime areas. The city of Minneapolis provided the first, but by no means the last, evidence of far greater concentrations of repeat problems at individual street addresses than among individual offenders. Only 3 percent of all street addresses in Minneapolis generated over half of all police rapid responses. This addition to the existing research on crime concentrations showed how police scholars could identify important new targets

for proactive policing on both sides of the Atlantic. While Goldstein's examples were richly qualitative, the research on crime and harm concentrations was intensely quantitative. American scholars found hot spots of crime. British scholars found patterns of repeat victimization (Farrell 1995). In both cases, the numbers of potential patterns were unlimited. Only newly available computers could say how many addresses or people had more than 10, or 20, or 200 crimes over any given time period. Only the computer could rank every known offender, victim, location, neighborhood, weapon, modus operandi, or crime recruiter by the number of repeat events in a year (Reiss 1988). What these studies showed were highly concentrated distributions of crime in a small proportion of any of the units at risk, from people to places to times and situations.

3. *Investigative Solvability Factors.* News reports intermittently "reveal" that only a small proportion of crimes lead to arrests or convictions, as if they had discovered something new. There is, however, no evidence to suggest that it was ever otherwise, at any time in the last two centuries (see, e.g., Monkkonen 1992). At the same time the RAND report on detectives was under way, scholars at Stanford University asked whether they could target solvable crimes for priority investment, allocating resources that might otherwise be wasted on pursuing unsolvable crimes.

This idea of "triage" was at least as old as the US Civil War, when doctors worked first on people who would die without immediate assistance. They postponed work on two other groups (hence the term "tri-age"): those who would die even with treatment and those who would live without immediate treatment. The strategy was intended to optimize the number of lives. For policing, it offered a way to make policing far more effective.

As Greenberg and Lang (1973) suggested, the likely "solvability" of each case can be reliably, but not perfectly, identified at the time an initial investigation is completed. By using a checklist of available forensic evidence in each case when it arrives on a detective's desk, the detective can decide whether to close the case or start working on it. The checklist that Stanford developed included such items as fingerprints, witnesses, footprints, names of likely suspects, and other facts. A test of one such solvability model on over 12,000 burglary cases found that it correctly predicted in 85 percent of the cases whether the case would be solved or not (Eck 1979), suggesting that nonsolvable

cases could be filed immediately with little difference in result. The next question, however, is still unanswered, as with the untested conclusions about response time and random patrols: how would a "checklist of solvability factors" regime compare to an "investigate every case" regime in terms of overall detection rates—or even to a "discretionary, nonchecklist decision about solvability" regime?

D. Tracking

No matter what targets are selected for police resources, no matter how well the police methods are tested, the central management question will always be, "what are police doing to accomplish our objectives, when, where, and with what apparent result?" In my first year at the NYPD in 1971, few police leaders ever discussed that question. By 1991, the question was asked all over the New York subways by then–Transit Police Chief William Bratton, and by 1995, he was grilling NYPD commanders on the question in front of 200 people at Police Headquarters. Yet long before New York reached that point, the British police were being driven hard by the numbers. Only the numbers, and the conversations, were very different.

1. *Performance Management, UK Style.* From the dawn of the Thatcher government, UK policing was dominated by two themes: huge pay increases and an explosion of statistical management. Key Performance Indicators were tabulated annually on a national level and monthly on a local level (Parkinson 2013). Eventually, police agencies were not only assessed on whether they had met specific goals set by central government but also compared to all other police agencies in a kind of "class rank," or football league table. One police agency would be ranked first, one ranked last.

These rankings created enormous pressure on police leadership to track what their officers were doing. Yet the pressure was not matched by any new tools, either hardware or managerial, for monitoring police activity. Only the numbers of police car dispatches, crime reports, or arrests were easily measurable. It was no surprise that these numbers were sometimes manipulated, including a "flexible" definition of crime types and other crimes "taken into consideration" when confessed by offenders caught red-handed. Even police response time was used in some years, despite its discredited status in the United States.

By 2012, these interagency competitions were mostly rejected in UK police culture. Yet the precedent they set for tracking in general is still

widely accepted. What police now understand is that the number of actions tracked should be relatively small, and each one should have a direct connection to something that is well researched and found useful. What US police learned from the British experience is that timely data could help drive a police agency.

2. *COMPSTAT.* No one put that lesson to better use than William Bratton (1998), who served as the chief executive of the Boston, New York, and Los Angeles Police Departments. With his colleague Jack Maple (1999), he developed a more focused process of tracking police activity. Using area commanders as the conduit for strategic initiatives, Bratton and Maple developed the COMPSTAT (computerized statistics) meetings as a means of pushing for better policing and less crime. Local commanders were arrested from time to time and charged with manipulating crime statistics, but for the most part, the system has been seen as operating with integrity. By 2012 the NYPD system was 18 years old and was still in operation, despite turnover in mayors and police commissioners. There is enormous potential to expand it to better indicators and performance.

3. *Auditing for Evidence.* In late 2012 the Metropolitan Police at New Scotland Yard launched a top-to-bottom effort to push evidence into every aspect of police operations, training, and promotions (Stanko 2013). This decision included a plan to appoint a Commissioner's Professor of Policing in a major university, a position that would entail 50 percent time doing research for a 50,000-employee police agency. This research could include comprehensive reviews of police practices against what the research evidence—especially testing—says about those practices. It would also likely entail new randomized controlled trials to be conducted in London. This plan for evidence-based policing was premised on the recognition of a vast body of research that had not yet been fully applied to police operations. This decision of Police Commissioner Sir Bernard Hogan-Howe provides perhaps the clearest indication of the Anglo-American rise of evidence for police decisions.

E. Explaining the Rise

How policing learned so much, and put so much of it to work, cannot be easily explained by any current theory of knowledge. Even the history of medicine offers thin explanations of why doctors do research, and why, if ever, they decide to use it (Millenson 1997) or

not, as in their continued failure to wash their hands after touching each patient (Pittet and Boyce 2001). What follows is my insider's hunch about a theory of police knowledge. This hunch is by no means a general theory of developing and using professional knowledge. My explanation is deeply, if not widely, "grounded" (Glaser and Straus 1967) in my continuous participant-observation study of Anglo-Celtic policing since 1968.

The theory has a major premise: that democratic policing is an extremely open system. Far more than any publicly held stockholder corporation, and probably more than any other government agency, policing is always vulnerable to the demands of any vocal constituency. No matter how small the group of demanders or how justified their complaints, police leaders must listen to them and give due consideration to them in providing better (or different) police practices. This pattern of organizational behavior has been observed in general since at least Thompson (1967, pp. 30, 37). Wilson (1968, p. 78) described a subset of this pattern in his discussion of "critical events" in policing, such as scandals, riots, police killings of citizens, and police strikes. The key question for evidence is what happens next: how critical events may shape evidence, and how evidence may then shape policy.

Figure 1 presents a causal model in which external demands for different policing are the major driving force. The theory claims that these demands are transformed by the human capital of the police institution, at both its core (police leaders) and its periphery (police scholars and research funders). In turn, human capital is shaped by the broader government policies that are also a target of public demands. The presence or absence of talented police leaders and scholars often reflects a 20- or 30-year lag in recruitment conditions and is highly dependent on national policies affecting police salaries, research funding, and other aspects of the capacity to generate and use research evidence. The people in a position to influence policy, if they are skilled enough, can then identify critical research questions and complete research projects, which are then communicated to police practitioners. Police managers and officers will then decide—immediately, somewhat later, or many years later—to apply the conclusions of the research evidence in their police practices. Sometimes this happens without their even knowing that the practice was shaped by research evidence (see, e.g., Police Executive Research Forum 2008).

These practices, and the ongoing critical events associated with

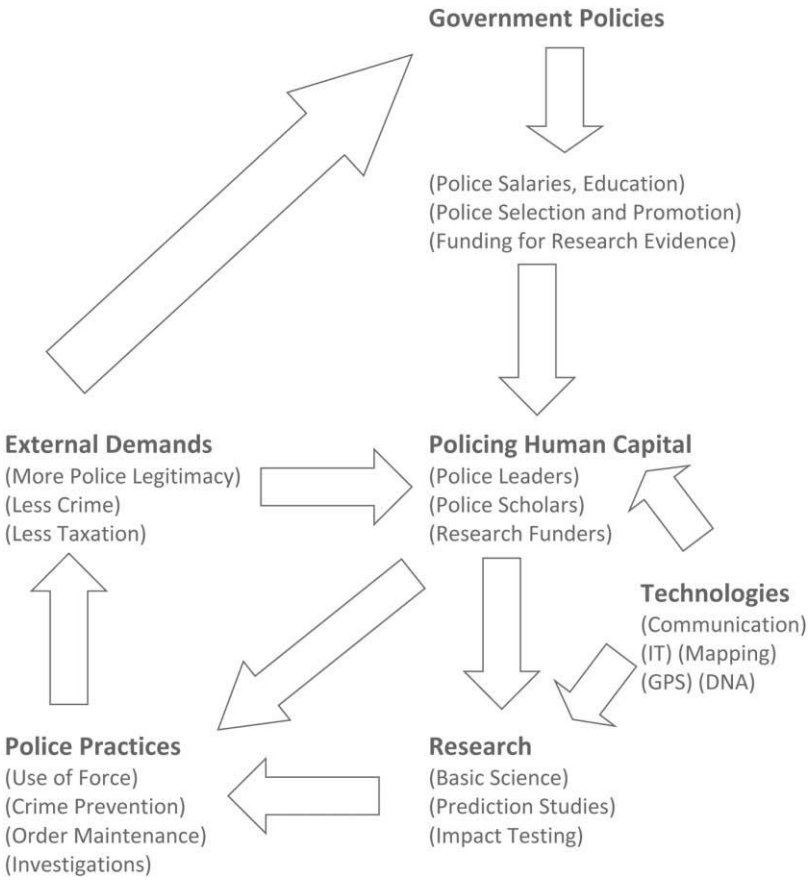

FIG. 1.—A theoretical model of the rise of evidence-based policing

them, continue to shape external demands on police. This happens only in part on the basis of police success or failure in meeting those demands. Many other things shape those demands, of course, and may be far more important than what police actually do. These other factors include economic ups and downs, emigration and immigration, ethnic and racial conflict, baby booms and busts, and possibly even a lag time from policies on leaded gasoline (Nevin 2000). Thus external demands on police are largely exogenous to the model, even though police can try to shape them to some degree. This makes the external demands

fairly insensitive to how well police are performing in any objective sense. Yet the model leaves room for improvements in police practice at least to mitigate criticism and possibly even to increase legitimacy of the police and trust and confidence in them.

The role of technology is also exogenous to the model and may be even more important than citizen demands (Manning 1992). Both police and societies adopt new technologies rapidly when their advantages are obvious and their cost declines. Many technologies, such as the automobile and the computer, are sold initially at high cost and then become much cheaper. By the time most US households owned an auto, for example, most police patrols were also performed in cars. The adoption of personal computers also lagged but was much faster. The desktop computer that analyzed police mainframe data to identify hot spots of crime (Sherman 1987) was unavailable in the Minneapolis Police Department in the 1980s, but later, faster, cheaper laptop models were widely available in policing by the 1990s. The rise of DNA testing still hinges on reducing the cost of speedy results (Roman et al. 2009). The spread and impact of body-worn video may also hinge on reliability and the cost of data storage.

The model does not suggest that there is a direct pathway between demands for specific improvements in policing, such as more legitimate rape investigations, and research on how this demand can be met. To the contrary: demands for immediate solutions send governments and police in a frantic search for research that has already been done. If none exists, that merely provides evidence that policing lacks an evidence base. For example, the death of a bystander hit by a constable in a 2009 London protest was associated with a controversial police tactic called "kettling," or restraining people for hours in a cordoned-off area. The UK government launched a rapid evidence assessment on kettling but found that little research had been done. Rather than investing in such research over a longer time frame, the government chose to report the knowledge gap and then ignore it.

The direct pathway from specific demands for police innovations usually travels to police leaders through theory and precedent from other jurisdictions, but not through new research. The way research can affect policy is to anticipate the chronically recurring issues and have research evidence ready for the next critical event. As a member of Parliament for Cambridge once said, research findings are "money in the bank" (Howarth 2010). They do not have to be used immedi-

ately and can go unused for years until there is a demand for them. That is why the line from research goes directly to police practices, since research can be kept in readiness to shape the decisions on police practices when "critical events" press police leaders to meet external demands to change them.

This theory explains the past but need not predict the future. A more rational model of using evidence to improve policing would be proactively led by the police profession, not reactively driven by the political environment. That model might even reduce the number of critical events that threaten police legitimacy. It is one that, as applied to health care, has helped to keep down medical costs in the United Kingdom to less than half the percentage of gross domestic product as in the United States, where there is no system to target resources on the basis of professional research evidence (Sherman 2011*b*). It is also the plan for what the College of Policing might do as a champion and arbiter of evidence-based practice.

II. What We Know and What Police Do

The clearest test of my claims about evidence is whether police decisions are more consistent with the evidence now than in 1975. As of 2012 the record on that test was mixed but encouraging. There are many areas of triple-T in which the evidence is useful and police are using it. There are others in which the evidence is useful but is not used. The adoption of evidence-based decision making has been stronger for targeting and tracking than for testing, but all three areas have seen a substantial impact of research on practice. This section supports those conclusions with a series of short summaries of good evidence on many aspects of policing, followed by even shorter summaries of whether that evidence is used.

There are many other strategies that can use EBP besides triple-T. Evidence can be useful, for example, in the recruitment and promotion of more women police leaders, in developing better ways to train police, or in finding lighter-weight body armor with greater protection. This essay focuses on triple-T because it can shape the core of policing and holds the greatest potential for the direct reduction of harm in society.

The most accessible evidence of the large quantity of the testing police have done can be found by Googling the Evidence-Based Po-

licing Matrix at George Mason University's Center for Evidence-Based Crime Policy (Lum, Koper, and Telep 2010). This website lists over 100 controlled tests of police practices, in both randomized experiments and quasi experiments. The matrix is designed by a former Baltimore detective to increase the use of research evidence in daily police operations. Other easily accessible sources, such as http://www.crimesolutions.gov (at the US Department of Justice) and http://www.campbellcollaboration.org (operated by the Norwegian government), provide additional evidence on tests of police effectiveness.

Evidence on targeting and tracking resources is more scattered and inaccessible than it is for testing. The likely solution to that problem is for George Mason, the College of Policing, and other knowledge centers to broaden their knowledge repositories to include evidence on all areas of triple-T.

The evidence on the use of research in policing is much harder to come by and is badly in need of more evidence of its own. This section draws heavily on qualitative evidence gathered in over 54 weeks of discussions at Cambridge University with police executives from around the world in 2007–12 (primarily reporting on the United Kingdom, United States, and Australia). That evidence may be the least reliable of any reviewed here, but it is a consistent source applied equally to all the different topics addressed below.

A. Targeting Police Resources

In this subsection, I consider a range of evidence about ways police agencies can set priorities for investing their personnel with the greatest return on investment. It includes studies of how police use their time and of how they could focus their time on concentrations of high-priority events. It includes tests of targeting strategies as well as tests of predictions about high-priority targets, which must be distinguished from the tests of police impact on public safety discussed below.

1. *Differential Police Response.*

The Evidence. There is good evidence that police can greatly limit their use of rapid response by car without offending citizens or putting them at risk. Callers will accept police decisions not to dispatch a police car for nonemergencies if the policy is explained to them by the person who answers the telephone. In randomized experiments in three US cities (McEwen, Connors, and Cohen 1986), citizens calling about

matters that were not emergencies were given one of two responses: that police will come "as soon as they can" (standard policy at the time) or that no response is necessary since the matter can be handled by telephone or other methods. Citizens were equally satisfied with both responses. The benefit was that call takers performed the same work as patrol officers in about one-third of the work time, with no transportation costs or air pollution. In some agencies, rapid police responses could be reduced by 50 percent or more with substantial savings—or more resources preventing serious crime.

What Police Do. The use of "differential police response" (DPR) remains very mixed. Even when DPR is used, it is rarely evidence-based. The 1986 evaluation is generally unknown to contemporary police professionals, yet it is freely available online and could save millions annually in any large police agency. Every police agency could easily conduct its own randomized trial of different kinds of incidents for which callers would receive a nondispatch response. At a time when police resources are declining in the United States and the United Kingdom (Sherman 2011*b*), an evidence-based approach to targeting resources where they are needed most is now needed more than ever.

2. *Hot Spots of Crime.*

The Evidence. There is good descriptive evidence that police can target resources on most crimes by identifying the small fraction of small places in any city where crime happens repeatedly (Sherman, Gartin, and Buerger 1989; Weisburd, Groff, and Yang 2012). These hot spot concentrations are clearest at a very micro level, such as a single address, a cluster of addresses, or a "block-face" street segment from one corner to the next on a single street. While some hot spots may wax and wane from year to year or month to month, many are stable for 15 years or more (Weisburd et al. 2004). Hot spots can be mapped, ranked, classified by offense type, and analyzed in many ways relevant to police operations.

What Police Do. Crime mapping is widespread, with many police agencies using it to identify hot spots. By the 2007 US Law Enforcement Management and Administrative Statistics Survey, over half of all cities of 50,000 or more residents used computers to generate crime hot spots, including almost all cities of over 500,000 people (Reaves 2010, p. 22). But the definition of hot spots varies widely. In many UK agencies, these locations are usually much larger than what the original research evidence analyzed. Entire patrol beats, rather than street cor-

ners or face blocks, are the United Kingdom's more common unit of analysis, while the opposite tendency is found in the United States. A sample of 192 big US cities showed slightly higher use of intersections (63 percent) than of neighborhoods (57 percent) as the definition of hot spots (Police Executive Research Forum 2008, p. 3). UK police in 2012 were at a special disadvantage because of their dispatch systems, which often failed to pinpoint the locations where crimes occur; some audits have shown the mapped locations of crimes to be great distances from the actual locations. Any police agency that is serious about reducing crime must, arguably, invest in correcting crime location errors, ideally by using GPS devices to include the exact latitude and longitude of every crime reported. In Andhra Pradesh, India, some police stations had incorporated GPS data into all crime reports by early 2011, and Trinidad and Tobago adopted it nationally in 2012, when no UK agency had yet done so. GPS measures would also improve the accuracy of targeting analysis for directed patrol and problem solving.

3. *Hot Spots of Distrust.*

The Evidence. There is good evidence that distrust of police is concentrated in a small proportion of urban neighborhoods. Police legitimacy is defined as a public (or police) perception of the moral right to exercise legal powers. This means that policing can be legal but illegitimate in the eyes of some (or most) beholders (Tankebe 2009; Bottoms and Tankebe 2012). The legitimacy of policing is highly correlated with minority or majority status in many populations (Sherman 2002), with majority populations attributing much greater legitimacy to police than minorities. Riots in US cities were concentrated in such hot spots (Rossi, Berk, and Eidson 1974). One recent study of public trust in a UK police jurisdiction showed very high levels of distrust in certain minority areas (Murray 2009). Yet not all majority-minority areas have highly negative views of police; some (as in Chicago) have more positive views (Skogan 1990; Skogan and Hartnett 1997; Sampson and Bartusch 1998).

What Police Do. Very few police agencies take annual surveys of public attitudes toward police. The US agencies that pay for such evidence tend to focus on citywide ratings rather than neighborhood-level differences. Some UK agencies have paid for neighborhood-level breakdowns with samples as large as 30,000 responses agencywide. Targeting strategies for increasing police legitimacy cannot be accom-

plished on a geographic basis unless such evidence is generated annually, so that deterioration or progress can be identified.

4. *Hot Times and Days.*

The Evidence. Crime risks vary enormously by time of day and day of week (Sherman 1992*a*), as well as season by season and holiday by holiday. Weather also has a strong effect on police workload (Cohn 1993), and in some regions it can be predicted days in advance with high reliability.

What Police Do. This issue remains the major failure of evidence-based targeting (Sherman 1992*a*). More money is wasted by ignoring "hot times" targeting than by any other police management practice. As one UK inspection report observed (Her Majesty's Inspectorate of Constabulary 2010, p. 3), more police officers were on patrol on Monday mornings than on Friday nights. The failure to schedule a workforce based on high-crime times and days is largely due to police labor contracts. These agreements serve the reasonable interests of police officers in a predictable work schedule. What is unreasonable to taxpayers is chronic and predictable understaffing during the peak crime times and overstaffing during "cool" times.

5. *Repeat Offenders.*

The Evidence. A small number of all offenders commit a high proportion of all crimes (Blumstein et al. 1986). Few specialize in any particular crime, and most of the crimes they commit are offenses against property. Controlling crime rates, and even the aggregated crime harm, may depend on police identifying, managing, or arresting prolific offenders.

What Police Do. Some police agencies target high-volume offenders who are not in jail (Spelman 1990). Yet almost no police agency has an instant query capacity to determine whether a particular known offender is in prison or at large. No agency appears to develop objectives for reducing recidivism of high-rate offenders, or even tracking the proportion of known repeat offenders they have put behind bars or under correctional supervision at any given time. Since the population of such offenders at large changes daily, a substantial investment in evidence gathering would be needed to target repeat offenders day by day. Nonetheless, most agencies would report that "we do that already," without actually investing in what is required to target repeat offenders in an evidence-based way.

6. *Dangerous versus "Supersafe" Offenders.*

The Evidence. More important than repeat offenders, and far fewer in number, are the most dangerous offenders. People who are most likely to kill, rape, rob, or commit sex crimes against children can be identified and forecasted from large databases (Berk et al. 2009). While the error rates in forecasting dangerousness can be substantial, the errors in forecasting "supersafe" conduct are minimal. A UK analysis of 100,000 convicted offenders found 96 percent accuracy in classifying offenders as not dangerous if somewhat lower accuracy (84 percent) in forecasting zero recidivism over a 2-year period (Sherman, Cosma, and Neyroud 2012).

What Police Do. The techniques used to classify known offenders as very dangerous, supersafe, or in between are new to both criminology and policing. From around 2000, UK police chiefs tried hard to identify the most dangerous offenders, but with highly subjective "clinical" methods that have failed to pass basic statistical tests (Thornton 2011). UK chiefs have recently expressed great interest in applying statistical forecasting methods to their local offender populations, requesting the statistical models as soon as they could be made available (Peter Neyroud, personal communication). Meanwhile, the few police agencies employing their own statisticians are free to replicate the published models.

7. *Crime Recruiters.*

The Evidence. A small proportion of repeat offenders have a high propensity for being arrested with, and apparently "recruiting," many different co-offenders (Reiss 1988; Reiss and Farrington 1991; Sherman 1992*a*). Where these co-offenders are almost always younger than the more experienced offender, the pattern suggests that the frequent co-offender may be recruiting novices into crime or certain types of crime. The hypothesis is even more compelling if the younger offenders have no prior record. That is exactly what an analysis of all arrestees in Sacramento, California, found over a multiyear period (Englefield 2013).

What Police Do. Only one police agency so far is known to generate lists of high-volume crime recruiters (Sacramento), but others may follow suit. The fact that the strongest evidence was generated not by an academic but by a police officer writing a master's thesis (as a self-funded student at Cambridge) is a good indication of the rise of professionally driven research in policing (Weisburd and Neyroud 2011).

8. *Traumatized Victims.*

The Evidence. Crime victims suffer great trauma from rape and rape attempts (Foa 1997). They also suffer lesser but high levels of trauma from noninjury crimes such as burglary and robbery, especially women (Angel 2005). While they are rarely diagnosed with post-traumatic stress disorder, they suffer elevated levels of post-traumatic stress symptoms.

What Police Do. Anglo-Celtic police have become far more sensitive to rape-related trauma since 1975 and have developed elaborate protocols for rape investigations and related medical care. Police do not, however, generally target the victims of lesser trauma, largely because there has been little external pressure to do so. Yet well-tested methods exist that police could use to massively reduce the symptoms of post-traumatic stress.

9. *Repeat Victims.*

The Evidence. High proportions of residential and personal crimes are committed against repeat victims (Pease 1998).

What Police Do. Police programs on repeat victimization tend to target all victims rather than to identify repeat victims. Other than the Minneapolis RECAP program in the 1980s (Buerger 1993), few police agencies have even developed lists of repeat victims. This may be a missed opportunity for reducing both crime and harm.

10. *Near-Miss Victims.*

The Evidence. One of the most predictable risks of a residential burglary is in the zone of a few buildings on either side, behind, and in front of a burglarized residence for the first 10–14 days after the burglary (Bowers, Johnson, and Pease 2004; Johnson and Bowers 2004).

What Police Do. For a decade after these findings were published, almost no police agencies identified these "near-miss" areas for prevention programs. Then a few UK police agencies began doing so in 2011–12. The word-of-mouth interest in this approach came from presentations by University of London academics to practitioner audiences at Cambridge and Cardiff conferences of the Society of Evidence-Based Policing and was still spreading in 2012.

11. "*Solvability Factors.*"

The Evidence. There is good evidence, summarized above, that police investigations can be targeted on the cases most likely to be solved and that investigations are wasted on cases in which a detection is predicted to be highly unlikely.

What Police Do. This is another major failure of getting research into practice, at least by 2012. It is hard to find a police agency that employs a statistical model of solvability factors to allocate the very scarce resource of investigators' time. Many police agencies still say they investigate all crimes. Others admit that they will close cases that are not solved after a certain period, but they do not use a statistical algorithm to decide which ones to abandon at the outset. Many confuse this research with closing cases on the basis of seriousness rather than solvability, such as when the value of property stolen is too low. It remains a prime opportunity for policing to save money and do more to help crime victims.

B. Testing Police Practices

The following subsections discuss a selection of major police practices that have been tested with at least a comparison group. Many have been tested with randomized controlled trials, and some have been tested repeatedly with up to 20 or more randomized trials. It is the most robust area of evidence supporting the triple-T strategy but also the area of biggest gaps between evidence and practice.

1. *Random Marked Car Patrol.*

The Evidence. There is little direct evidence that random patrol of areas that have little crime can cause any less crime than having no patrol. The major exception is police strikes, which consistently cause crime to skyrocket (Sherman 1992*a*, pp. 192–93).

What Police Do. Random patrol by "rapid response" cars is still widespread across Anglo-Celtic policing. Even when police are told to "concentrate" patrols in, or "give special attention to," certain areas, this can amount to only a few minutes per work shift. Random patrol predominates in practice if not in rhetoric; hot spot patrol is rising rapidly, at least in rhetoric if not yet in practice.

2. *Hot Spot Car Patrol.*

The Evidence. Doubling the dosage of marked car patrols in high-crime hot spots at higher-crime times can measurably reduce crime and disorder in those hot spots (Sherman and Weisburd 1995). In the Minneapolis hot spots experiment, an average of 15 percent of at-risk time with police present produced moderately less crime, and half the disorder, than in a comparison group with police present for only 7 percent of at-risk time. Koper (1995) reported that the greatest residual deterrence in each hot spot (Sherman 1990) was associated with 15-

minute periods of patrol presence, compared to both shorter and longer periods of patrol. Various police strategies have been tested in other hot spots experiments (Braga, Papachristos, and Hureau 2012), but none has apparently replicated either the unit of analysis or the measured dosage levels of the initial Minneapolis experiment (Sherman and Weisburd 1995). Displacement was not measured in the initial experiment, but it has been falsified in a range of other hot spots patrol tests (Weisburd et al. 2006). No experiment, however, has compared hot spots to random patrol at the level of police districts. No test has yet made a direct comparison between, for example, 50 districts using random patrols and 50 districts concentrating patrols in crime hot spots. Until such a test is done, a huge gap remains in the evidence on hot spots policing.

What Police Do. Directed patrols at crime hot spots in marked cars is a growing practice in the United States and the United Kingdom. Hot spots patrol has been identified by the NYPD as a major strategy to which it attributes a decline in homicide rates (Zimring 2012). Weisburd and Lum (2005) found that among the 92 US police departments responding to a survey of a random sample of 125 with 100 or more police officers, 62 percent had adopted computerized crime mapping by 2001. The most common reason these early adopters gave for their decision—43 percent—was to "facilitate hot spots policing." Thus by 2001, 27 percent of the responding agencies said that they had adopted computerized crime mapping to facilitate hot spots policing. By 2007, the Police Executive Research Forum survey of 192 large police agencies found that hot spots had become the most common police strategy for fighting violent crime since "nearly nine out of 10 agencies use hot spots enforcement efforts directed either at larger hot spot areas like neighborhoods (57 percent), smaller hot spot places like intersections (63 percent), or both" (2008, p. 3). In the United Kingdom, hot spots patrol is growing in tests as well as in practice, with three different police agencies testing such patrol in 2012–13.

3. *Random Foot Patrol.*

The Evidence. Two quasi-experimental tests of random foot patrol found no effect on crime (Kelling et al. 1981; Trojanowicz and Baldwin 1982). The Newark Foot Patrol Experiment, however, found reduced public perceptions of crime and disorder in areas with foot patrols compared with control areas (Kelling et al. 1981).

What Police Do. Random foot patrol has almost disappeared in

Anglo-Celtic policing. Its use is limited to the highest-density areas of the largest cities, such as New York and London. Hot spot foot patrol is far more common.

4. *Hot Spot Foot Patrol.*

The Evidence. A brief but randomized controlled test of foot patrol in high-crime areas of Philadelphia found that foot patrol caused significant reductions in violent crime with only modest local displacement effects (Ratcliffe et al. 2011). The areas were much larger than the Minneapolis hot spot street corners but still smaller than a typical police district.

What Police Do. In the United States, foot patrol in high-crime hot spots is most likely to be used in densely populated cities. In the United Kingdom, Police Community Support Officers, whose uniforms look much like those of police constables with full arrest powers, are often used for foot patrols in town centers and suburban shopping strips.

5. *Rapid Response Policing.*

The Evidence. There is no direct evidence that rapid response can make any difference in detection or crime rates and some indirect evidence that it cannot. It is very rare that rapid response can catch an offender. Rarer still does rapid response save a crime victim from serious injury. The capacity to do so may be a core feature of police legitimacy, an aspect of this issue on which there is no evidence.

What Police Do. Police agencies still spend substantial funds ensuring the capacity to respond very quickly. Some of them track average response times very closely and advertise the results by public announcements.

6. *Problem-Oriented Policing.*

The Evidence. Both randomized experiments and quasi experiments have tested Goldstein's (1979, 1990) proposed strategy of attacking underlying causes, rather than just symptoms, of patterns of crime. Most reported studies have been before-after studies with no control group, most of which report substantial success. The largest randomized trial (Sherman, Gartin, and Buerger 1989) reported that POP caused a 15 percent reduction in calls over 6 months in a sample of some 250 residential locations, but not in a sample of 250 commercial locations. A systematic review of randomized trials of POP found, on average, that POP produces modest but clear reductions in the problems it targets (Weisburd et al. 2010).

What Police Do. Thousands of US and UK police are enthusiastic

about POP and often do it despite peer pressure not to. For many it is an unacknowledged addition to their random patrol work, which many police agencies have encouraged. What some, but still few, agencies have done is to create either specialized units of experts in POP (such as the Minneapolis RECAP Unit) or case management systems for identifying problems and assigning specific time periods for specific officers to undertake POP work and be held accountable for the results.

7. *Stop-and-Question.*

The Evidence. There is highly consistent evidence that stop-and-question or stop-and-search causes reductions in weapons violence and homicide (Koper and Mayo-Wilson 2006), based on seven out of seven quasi experiments reducing either homicide or gunshot wounds. Repeated cross-sectional analyses across US cities show lower levels of robberies and other crimes where police proactively issue more traffic violations or make arrests for disorderly conduct (e.g., Kubrin et al. 2010). No evidence exists on the life course effects of police stops on those who are stopped, especially whether a contact with police that is seen as hostile can engender defiance that increases future offending (Sherman 1993). One survey of a residential area with high levels of gun violence found substantial support for this practice among residents (Shaw 1995), but no surveys were done of those stopped.

What Police Do. For many police agencies, stop-and-question is an embedded part of patrol practice in high-crime areas. Its use in crime hot spots in New York and London is highly visible and highly controversial. Few debates distinguish between the questions of how stops should be done and whether they should be done at all. Many police agencies track the occurrence of stops, but not their qualities of procedural fairness. As Reiss (1971) suggested, all citizen encounters could end with police giving citizens a receipt with the officer's identifying details and the date and time. Such a system could lead to measures of quality.

8. *Covert Surveillance.*

The Evidence. The use of covert surveillance on repeat offenders alleged to be committing serious crimes was tested in a Washington, DC, experiment (Martin and Sherman 1986). Random assignment of surveillance, compared to a control group, substantially increased the likelihood of a suspect being arrested as well as convicted and sentenced to prison.

What Police Do. Police rarely use covert surveillance against on-

going street crime. UK police use it sparingly for very serious sex offenders who have been released from prison.

9. *Arrests for Domestic Assault.*

The Evidence. There is consistent evidence that arresting unemployed domestic abusers causes more violence but no consistent evidence that arrest reduces domestic violence. Three randomized controlled trials of arrests for domestic assault have examined the interaction between employment and arrest on repeat violence (Pate and Hamilton 1992; Sherman 1992*b*); a fourth examined the same interaction on repeat domestic calls including arguments as well as violent crime (Berk et al. 1992). All four tests showed that arrest had different effects on employed suspects than on unemployed suspects. The three studies of assault found that arrest reduced repeat crimes for employed suspects but substantially increased repeat crimes for unemployed suspects. Two of the five randomized trials of the main effect of arrest for assaults did not report tests for an interaction effect with suspect employment. But the Milwaukee experiment found that arrests for common assault caused significantly more domestic violence repeat offending than an on-site warning even after 24 years (Sherman and Harris 2012).

What Police Do. Elected legislators have prevented police from using evidence-based practice on domestic violence in over half of US states and all across England and Wales. Either a legal or a policy mandate requires police to make arrests whenever there is sufficient legal evidence to do so. Police in some US agencies have widely ignored this mandate for decades (Ferraro 1989). Others in the United States and most of the UK agencies have been more compliant, but with frequent sensitivity to victims' preferences. New experiments in policing domestic violence were launched in Hampshire (UK) in 2012, when evidence-based changes in policy were widely discussed across the United Kingdom.

10. *Warrants for Domestic Assault.*

The Evidence. There is good evidence that issuing an arrest warrant for an absent domestic violence suspect has a "sword of Damocles" effect in suppressing repeat assaults. A randomized controlled trial in Omaha, Nebraska, enrolled cases in which suspects had already left the scene before police arrived. The police then randomly assigned either the standard police policy of advising the victim how to pay for an arrest warrant to be issued in court or went to the court to arrange

issuance of a warrant free of any charge to the victim. In the cases assigned to police-initiated warrants, there was significantly less repeat violence reported by the victims of abuse to the university interviewers. There were also substantially fewer police reports of repeat offending (Dunford 1990).

What Police Do. Police rarely issue warrants for simple assault. Some English police go looking for an absent suspect when there is evidence of an assault and "hand over" the case to the next shift if the absent suspect has not been found. No evidence exists on the rate at which absent offenders are found or the rates of those not found being subjected to a police-initiated warrant, but these rates could be part of tracking.

11. *"Second Responder" Follow-ups after Domestic Incidents.*

The Evidence. Several randomized controlled trials have tested follow-up visits by "second responder" police or social workers, or both, some days after police have responded to a domestic violence call or made an arrest. A systematic review of these experiments included both repeat calls to police and interviews with the victims (Davis, Wiesburd, and Taylor 2008). The review found no difference between the self-reported victimizations of those victims who had been visited and those who had not, but there were more calls to the police about new violence by those who had been visited than among those left alone. While the evidence is ambiguous, second response cannot be considered beneficial. Increased reporting might be of some benefit, but it may also just indicate more violence.

What Police Do. Few police agencies deploy second responders, yet some UK agencies "invented" the idea in 2010–12 without knowledge of the results of the Davis, Weisburd, and Taylor (2008) systematic review.

12. *Police-Led Restorative Justice Conferences.*

The Evidence. Eight randomized controlled trials have tested police-led restorative justice conferences (RJCs) with crime victims, their offenders, and their respective friends or family in Australia (two), the United Kingdom (five), and the United States (one). Seven of these eight tests found a reduction in repeat offending among offenders assigned to restorative justice (Sherman and Strang 2012). The one failure had a high proportion of Aboriginal offenders, for whom restorative justice badly backfired (Sherman et al. 2006). A systematic review and meta-analysis of these experiments, combined with two others in

which nonpolice led the conference, found an average effect that was small but statistically significant among robbers, burglars, violent offenders, and property offenders (Strang et al. 2013).

What Police Do. Few Anglo-Celtic police officers were leading RJCs in 2012, although other kinds of people led them frequently in New Zealand and Northern Ireland. Police in England and Wales have been trained to use a "mini" version of these conferences in street settings, immediately after responding, which are often called "restorative disposals" or "community resolutions." This version of RJCs has not been tested.

13. *Prosecution of Juveniles.*

The Evidence. There is good evidence that when police prosecute most juvenile offenders, they cause more crime than they prevent—at least among those offenders. In 29 randomized controlled trials over four decades, in studies enrolling a total of 7,304 juvenile offenders, those who were prosecuted after arrest had more repeat offending than those who were diverted or cautioned (Petrosino, Turpin-Petrosino, and Guckenburg 2010). All tests were conducted in the United States except two of the restorative justice experiments in Australia. The offenders had committed a range of property and less serious violent crimes. Those who were diverted from prosecution had lower offending rates when the disposition required them to do less follow-up work, and the best success was found with no conditions at all.

What Police Do. Police practices with juvenile offenders remain mixed and largely unmeasured. In the United Kingdom, there has been a substantial increase in formal prosecution since 1975, but this trend may have reversed since 2010.

14. *Neighborhood Watch.*

The Evidence. Neighborhood watch appears to cause a small but consistent reduction in rates of burglary and other crimes. A systematic review examined 18 quasi-experimental controlled tests of neighborhood watch programs, either alone or in combination with other situational prevention strategies, such as property marking and home security surveys (Bennett, Holloway, and Farrington 2006). Crime was measured by both police-recorded crime and victimization surveys. The review found, on average, statistically significant crime reductions of 16–26 percent. Given the low levels of crime in many of the middle-class neighborhoods in which the tests were done, it remains unclear whether the crime reductions were cost-effective. Yet that issue can be

raised about any evidence of a positive effect of a police strategy relative to a control group.

What Police Do. There is no good evidence about what police do in neighborhood watch programs after a local group has held its first meeting. There appears to be no tracking evidence on how much time police invest in starting or maintaining neighborhood volunteer interest in preventing crime. Signs warning would-be offenders that "this area is protected by neighborhood watch" are ubiquitous in many US suburbs, but whether there have been any neighborhood meetings to discuss crime in months or years is impossible to say.

15. *Drug Abuse Resistance Education* (*DARE*).

The Evidence. Repeated quasi experiments (Sherman et al. 1998) and a major randomized controlled trial (Perry et al. 2003) have failed to show that the standard DARE program reduces drug use or delinquency. The Perry et al. trial of 24 schools in Minneapolis did find that a greatly enhanced program called DARE Plus reduced multidrug use (but not marijuana use) among boys, with no effect on girls. But Dare Plus involved much greater use of parents than the standard DARE program, which relies heavily on police visits to classrooms for its delivery.

What Police Do. According to the DARE America home page, 220 US communities started DARE programs over the preceding 3 years—more than one per week. These new startups are in addition to the website's claim that 75 percent of the more than 14,000 US school districts offer DARE programs. Despite the example of former Mayor Rocky Anderson of Salt Lake City, Utah (population 190,000), ordering his police department in 2000 to stop wasting money on DARE (which he called a "fraud"; Eyle 2001–2), most US police agencies still appear to be sending police officers into classrooms to talk about drugs—with no demonstrated return on the taxpayers' cost.

16. *"Pulling Levers": Focused Deterrence.*

The Evidence. The use of directed sanction threats to specific individuals and groups gained worldwide attention when it preceded a complete cessation of youth homicide in Boston in the 1990s. A systematic review of 10 quasi-experimental evaluations of this approach, seven of which used nonequivalent (unmatched) control groups, examined the effect of this strategy on such crimes as homicide and gang violence, mostly at the citywide level (Braga and Weisburd 2012). The review found that nine out of 10 evaluations showed statistically sig-

nificant reductions in crime. Given the many threats to the internal validity of these evaluations, the review authors described the evidence as "promising."

What Police Do. Police in communities with high homicide rates, from Trinidad to Scotland, often try to replicate the "Boston miracle," which has received further attention from David Kennedy's (2008, 2011) books on his experiences in Boston and elsewhere.

C. Tracking Outputs and Outcomes

1. *Police Killings of Citizens.*

The Evidence. Fifty years ago, criminologist Gerald Robin (1963) published the first cross-city comparison of the rates at which police committed justifiable homicide against the citizens of their communities. This study helped to spark two decades of research on the issue (e.g., Sherman and Langworthy 1979; Fyfe 1980, 1982; Sherman 1980, 1983) that supported an appeal of a police killing case to the US Supreme Court (*Tennessee v. Garner*, 471 U.S. 1 [1985]). The later research showed that limiting police killings to defense-of-life situations had reduced deaths, without any increases in crime or violence against police. The *Garner* decision, which cited the research evidence, ruled that common law powers to kill unarmed fleeing felons were an unconstitutional seizure without due process. The decision restricted police powers to kill in over half of the 50 US states.

What Police Do. Tennenbaum (1994) reports that killings by US police, which had already been dropping prior to the 1985 *Garner* decision, dropped by 16 percent in its immediate aftermath. He also showed that police shootings for all reasons, not just to catch fleeing felons, declined. Most US police shootings are now tracked very closely, and police gun use is regulated by agency managers far more intensely than in 1975. Police killings in the United States have risen and fallen since then, but shootings of unarmed fleeing felons have virtually disappeared.

2. *Police Use of Force.*

The Evidence. In the mid-1970s, city manager–run police agencies began to require police to file a use-of-force report each time they used weapons, handcuffs, or physical restraint against a citizen (Croft 1985). This paper file tracking system enabled police managers to identify statistical outliers in the frequency of use-of-force reports, as well as to justify in court or arbitration hearings any dismissal of overly

frequent users of force from police service. An independent commission led by Warren Christopher (1991) in the aftermath of the Los Angeles Police Department's Rodney King case documented extreme concentrations of use of force among a small percentage of police officers.

What Police Do. Use-of-force reports are now widely used in the United States, usually with stand-alone digital databases that allow frequent in-house tracking analysis of the evidence. Litigation over police use of force is defended with this evidence and possibly even prevented by managerial application of these analyses to personnel and assignment decisions.

3. *Complaints against Police.*

The Evidence. In 1975, systems for receiving, recording, investigating, and resolving citizen complaints against the police were very scarce. Cities as big as Louisville, Kentucky, made it almost impossible for citizens to complain against police (Sherman 1976). By the mid-1980s, most US cities over 100,000 recorded and investigated complaints against police. By 2012 most US and UK forces had computerized systems for tracking complaints by officer characteristics, complainant characteristics, area demographics, type of encounters, and other factors of managerial interest.

What Police Do. Most large police agencies in 2012 had "professional standards" units reporting to the police executive on trends and patterns in complaints against police. Few of these, however, appear to be adjusting the analyses for such crucial risk factors as how many encounters each officer had with citizens each year, in each unit, or in each area. As in tracking evidence generally, police managers often look at raw data that are not meaningful unless they are standardized by relevant denominators (Kahneman 2011). High levels of complaints may indicate high levels of misconduct. But unless complaints are divided by contacts, they could also indicate a high level of professional commitment to making as many appropriate contacts (and arrests) as possible. Fairness alone requires the use of a standardizing denominator, as well as good tracking measurement.

4. *Trust and Confidence in Police.*

The Evidence. There is now abundant research tracking trust and confidence in policing. A recent government in the United Kingdom made public confidence the single performance measure it said would be used to judge police performance. Yet no governmental data sup-

plement raw data with risk-adjusted data, the latter taking into account factors affecting public confidence that are irrelevant to police performance. Risk factors for public confidence may include population demographics, structural inequality, lack of collective efficacy (Sampson and Bartusch 1998), cultural backgrounds of populations, and other factors (e.g., Jackson and Bradford 2009). There is a great need for better risk-adjusted models of "expected" levels of trust and confidence in police.

What Police Do. Many police agencies fail to track trust and confidence at all. Some track raw levels citywide, and a few track it by neighborhoods. There appears to be no use of risk-adjusted confidence measures in any of the police agencies from 10 countries whose representatives were asked about this in recent classes of the Cambridge Police Executive Programme.

5. *COMPSTAT.*

The Evidence. Since 1975, nothing has done more than the COMPSTAT idea to increase the availability of evidence for tracking police performance at micro levels of activity. At highly structured and periodic meetings of senior police officials, operational commanders report the tracking data on their crime patterns and (less often) on their anticrime strategies. Pioneered by Bratton (1998) and Maple (1999), the COMPSTAT program was probably the first police "interaction ritual" (Collins 2004) to make a "sacred" meaning. By reporting the crime trends within days rather than months at very local levels, the institutional meaning and police legitimacy associated with crime control became far more intense than they had ever been before (Willis, Mastrofski, and Weisburd 2007). My own observations in the NYPD in 1972 and in 1995 revealed an enormous increase under COMPSTAT in the extent to which specific evidence about crime and police practices was portrayed in statistical and graphical visualizations. My own attempt to introduce something similar for the mayor of Indianapolis in 1994–95 had failed to mobilize the same kind of intensity, largely because the New York version engaged almost 200 people in the room listening very closely (compared to six or seven in Indianapolis). New York also used COMPSTAT with its unusually great legal authority to press unambitious area commanders to retire and to reward successful commanders with promotions—something few other agencies are able to do. NYPD officials who had used COMPSTAT moved to other large cities, where they also used COMPSTAT for citywide reviews of specific issues, from domestic violence to police

corruption (Sherman 1978). No other system of tracking evidence on outcomes or outputs has been so influential or comprehensive.

What Police Do. In 1999, some 5 years after New York's COMPSTAT had become highly publicized, the Police Foundation conducted a national survey of US police agencies in cities over 100,000 residents (Weisburd et al. 2004). The survey found that over half of those agencies (with an 85 percent response rate) either had adopted a COMPSTAT-like program or were planning to do so. In the early twenty-first century, COMPSTAT-like meetings were held regularly in many UK police agencies. Few such meetings, however, track the key question under discussion: where are the police doing what, and for how much of their time on duty? That is a question that two major advances in technology will soon make it easier to answer.

6. *GPS Tracking of Officer Locations.*

The Evidence. Police agencies have been able to track the locations of police cars since at least the 1980s but have rarely done so. Police unions once resisted the idea, especially in the 1970s, when some US officers regularly slept in their cars on the midnight shift. The rise of GPS devices in police radios and "smart" phones now makes it possible to track the whereabouts of each officer every few minutes or seconds. Police officers also seem more willing to allow this, if only as a safety device that could enable a rapid response of other police to help them. An experiment in providing police supervisors with data on how much time each patrol officer spent at directed patrols in crime hot spots has been completed in Dallas, with encouraging results (Weisburd 2012). Other experiments testing the use of tracking evidence for effects on police performance are needed to determine whether tracking at this level of detail is cost-effective or not.

What Police Do. From the limited evidence available, it appears that as of 2012, no major police agency has made statistical use of GPS data about locations of police outputs as a standard part of tracking or COMPSTAT. For reasons discussed in the first section of this essay, however, it seems likely that such tracking will be widespread by 2025.

7. *Body-Worn Video.*

The Evidence. Inexpensive body-worn video (BWV) cameras with long-running batteries came on the market around 2010. As a student in the Cambridge Police Executive Program, Police Chief Tony Farrar of Rialto, California, decided to test the hypothesis that police officers wearing BWV would generate fewer complaints from encounters with

citizens. One theory is that officers would be more polite and compliant with rules knowing that every word and action was on digital record. The other theory states exactly the same for citizens encountering police. In a randomized controlled trial of police shifts in which all patrol officers either did or did not wear the cameras, Farrar (2013) found initial evidence of fewer complaints on shifts when officers wore their BWV devices. Given spillover diffusion effects to shifts when officers did not wear cameras, it was no surprise that total complaints against police dropped sharply. Related hypotheses can be tested in this design for the rates at which arrestees plead guilty as charged or the rates at which either citizens or police are injured in their encounters. Note that this evidence is a prime example of testing a means of tracking police performance as distinct from testing the effects of performance itself. It asks whether tracking has benefits regardless of what police do and whether tracking improves police quality.

What Police Do. BWV is at a very early stage of development in policing. One UK agency tried to incorporate it into stop-and-question procedures, but citizens who were stopped refused to answer questions while police had their cameras on. A large Australian police agency has had problems with equipment failures. What no agency appears to have tried is the use of BWV to replace typing, with video-recorded crime reports, victim and investigative interviews, and other work. Such use of videos could keep police out of stations typing up text for many hours when they could be on patrol taking video statements as well as increasing patrol visibility. In the United Kingdom's ongoing effort to "cut red tape" and reduce reports, no solution could be easier or more elegant than downloading police video and audio as a digital logbook. If text reports need to be created for any reason, including prosecution, then civilian video specialists can do that rather than police trained to patrol, prevent, and investigate.

D. Summary: Knowledge and Its Use

The examples offered above support a key conclusion: there has been a massive growth of policing knowledge over the levels in 1975. Equally important is a second conclusion: there has been less progress in using knowledge than in generating it. But the crude (and perhaps inaccurate) measures of evidence in practice fail to suggest what may be the most important point: that police interest in using evidence grew most rapidly in the last decade before 2012. Even though the 1980s

was the most intense period for producing police research, the 2000s and 2010s were the most intense periods for consuming it. If the demand for evidence continues to grow, so may the future production of evidence.

III. Improving Triple-T: Targeting, Testing, and Tracking

This section highlights the major conceptual issues that must be addressed for policing to use evidence in a fully professional way. These issues all relate to the way in which the concept of evidence is defined. The greatest threat to police professionalism is that the word "evidence" will be hijacked to mean what it is intended to replace: intuition, anecdote, and opinion. Just because a word is fashionable does not mean that people who use it will understand it, let alone accept it and agree with its message. Even with the best of intentions, the word "evidence" is often applied incorrectly by people who lack adequate training in how evidence should be used. This risk requires that police leaders and scholars pay close attention to the integrity of the concept and relentlessly challenge any claims that "the evidence shows" when no such evidence exists.

This threat has major implications for how police training and promotion processes teach and communicate about evidence-based practices. Simple lists of "best practices" may have great value in some contexts. But for people sitting at conference tables deciding on what best practices they should adopt, simplicity is not good enough. Policing is a complex enterprise, and so is drawing conclusions from evidence. People analyzing evidence need to be qualified to do so, one way or another. At minimum, they should understand the concepts presented in this section.

It is reassuring that these same conceptual issues bedevil other complex enterprises as well. There is nothing wrong with police thinking about evidence that is not also wrong with the thinking of highly seasoned politicians (Issenberg 2012), baseball managers (Lewis 2003), and nuclear power plant engineers (Silver 2012, chap. 5). As Kahneman (2011) demonstrates, it is very hard for humans to discipline their thinking away from intuitive snap judgments ("system I") to accept more systematic analyses that follow strict rules of evidence ("system II"). Fortunately, he concludes that it is at least possible for humans to achieve that discipline, but only if we work pretty hard at it.

This section considers four questions. Most important is the question of what good evidence is, and what it means to make policing "evidence-based." The other three questions cover related issues in targeting, testing, and tracking. The targeting subsection introduces the concept of a crime harm index as the best way to measure what matters in valuing police effectiveness. The testing subsection shows how testing is increasingly done by "pracademics," not just academics. The tracking subsection discusses how to use evidence for ensuring better implementation of evidence-based policies.

A. What Is Good Evidence?

What is "good evidence"? This question must be addressed almost daily if policing is to raise its level of professionalism. Even among those strongly committed to using good evidence, it is easy to lose the core principles of testing.

At an early meeting of the Society of Evidence-Based Policing, for example, a presentation led a session chair to endorse an innovation in taking witness statements that had just been described by two presenters. One speaker described laboratory tests of the new procedure, while the other speaker described its use in (just) three cases in the field. No controlled field test had been conducted, nor any test measuring the key outcome of conviction rates. Yet on the basis of the strong results of the lab experiments, the session chair enthusiastically described the research as "really strong" evidence and suggested that it was unethical for police at the conference not to adopt the innovation. I felt compelled to challenge that claim by pointing out the major difference between laboratory and field tests. I noted that "evidence-based medicine" limits its definition of evidence to field tests only, viewing lab tests as a necessary precursor to testing treatments on real people. In this police context, this means that the proper channel for the enthusiasm we all felt for the innovation would be to conduct a randomized field experiment, which had not yet been done. The session chair readily accepted this amendment to his initial conclusion. Yet few advocates of new ideas would be so open-minded as the session chair, who is widely seen as one of the most promising police leaders in Britain. It is people like him who will make a key difference in how well evidence is understood, used, and defined.

In science, as in law, "evidence" is an objective finding that can be confirmed by repeated observations of independent observers and that

can help to support a conclusion. Most people, as philosopher David Hume (1748) observed, find great difficulty in accepting that their own subjective impressions, hunches, prejudices, theories, or opinions are not objective evidence. As the late Senator Daniel P. Moynihan observed, "the plural of anecdote is not data." The rules of science and statistics require many complex procedures and conceptual tests in order to establish that evidence is reliable, including probability sampling, causal inference, selection bias, statistical power, confidence intervals, and many other challenges to deciding what is a "fact"—let alone whether the known facts can support a conclusion. The US Supreme Court in *Daubert v. Merrell Dow* (509 U.S. 579 [1993]) codified these principles into federal court procedures for admitting scientific evidence of all kinds.

At a more advanced level of understanding, the challenge is grasping the full breadth of decision-making processes for which good evidence is essential—including but not limited to targeting, testing, or tracking. To some critics, "evidence" is only about testing, or even just about randomized controlled trials. To baseball fans familiar with the *Moneyball* story of the Oakland Athletics team (Lewis 2003), it is all about targeting: hiring the most cost-effective baseball players by replacing experienced baseball scouts with computer geeks. To some observers of President Obama's "victory lab" (Issenberg 2012), better evidence means replacing intuitive political "gurus" with computerized "geeks" for tracking the delivery of more favorable voters to the polls.

These partial understandings of evidence-based practice converge on a single "straw man" caricature: that quantitative thinking beats—and should replace—qualitative judgment. Nothing could be further from the truth. As the doctors who helped develop evidence-based medicine argued, the best evidence is a blend of individual clinical experience with the best quantitative and qualitative "external" evidence: research findings and experience about similar cases or problems (Sackett et al. 1996). As baseball statistician Nate Silver (2012, p. 100) observed, two baseball players could have identical statistics, but one could spend his evenings volunteering at a homeless shelter while the other is snorting cocaine in nightclubs. He writes that "there is probably no way to *quantify* this distinction. But you'd sure as hell want to take it into account."

Similarly, if a police commander knows that a robbery hot spot has developed in a racially tense minority area (targeting) and that the

average effect of doubling police patrols in robbery hot spots is to cut robbery 50 percent (testing), that is not enough evidence for a good decision. The commander would also think of many other qualitative factors, such as the recent or long history of police-community relations in that area, the extent to which stop-and-frisk would be used, the likely weather or school vacations, the availability of minority group officers to work in the area, and the track record or personality of a sergeant who would directly supervise any effort to double patrols. The commander might also wish to consult with neighborhood leaders on how best to bring in more patrols, perhaps with a handout flyer for police to give everyone they encounter on the streets—a practice for which there is experimental evidence from Australia (Mazerolle et al. 2013).

Evidence-based practice does not and cannot replace judgment based on experience. It can only inform such judgment, and usually improve it. Attempts to portray EBP as trying to replace rather than supplement experience (Sparrow 2011) are merely "noise" in relation to the signal. Not even baseball teams, as Silver (2012, pp. 105–6) points out, replaced qualitative scouts with quantitative geeks after the events depicted in *Moneyball* (Lewis 2003). Instead, baseball teams increased their investments in both, integrating qualitative and quantitative information supported by a profession-led science (see Weisburd and Neyroud 2011).

The body of knowledge supporting EBP embraces and incorporates police experience and craft skills. It does not replace traditional thinking about legality, legitimacy, and common sense. Nor does it replace the wisdom and experience of police professionals. What it adds to these assets is the crucial step of an objectivity check about questions that have often been answered on the basis of subjective opinions (Sherman 1998): is there any systematic, scientifically generated evidence that bears on the decisions we are about to make? Moreover, is the evidence the right kind of evidence for the decision?

While the meaning of the phrase "what works in policing" may seem obvious, it is not. Whether a prediction model "works" simply means whether its predictions are accurate; no randomized controlled experiment is needed. Whether GPS works to track police locations accurately is simply a matter of auditing by physical tests: there is no cause and effect relationship, just measurement reliability. Whether arrest for domestic violence works, however, is a matter of testing for cause and

effect. This means that in a controlled experiment, cases in which police made arrests would have to have had less repeat offending than in cases in which police did not make arrests. That is why the Maryland Scale of Scientific Methods, levels 1–5 (Sherman et al. 1998), is relevant to testing police actions in triple-T but not to assessing the accuracy of predictions for police targeting decisions. The Maryland scale is also irrelevant for audits, measures, and descriptions intended for tracking police performance. The best evidence for answering any question of "what works" thus depends on "which T?"

Table 1 summarizes the different standards for good evidence of "what works" for the three different Ts. The left-hand column shows three different levels of strength of evidence, combining several key dimensions: replications of evidence collection (for measurement reliability), consistency of findings (for content reliability), and—only in the case of testing—the capacity of the research design to rule out other causes besides the police practice being tested. These standards are certainly subject to debate. But perhaps the bare minimum, rock-bottom standard for EBP is this: a comparison group is essential to every test to be included as "evidence" that a police decision causes a certain result. That was the standard used for reporting on evidence to the US Congress (Sherman 1998), and the global police profession should arguably have standards that are at least as high or higher.

These distinctions illustrate the need for police professionals to raise their game in mastering the (scientific) rules of evidence. What follows is an overview of some of the key issues and of my ideas for addressing them. Further readings of the sources cited would be useful for anyone who proposes to use or, especially, produce evidence for policing.

B. Issues in Targeting Resources

The targeting of scarce resources can be compared to an investment portfolio. Like police agencies, investors have a variety of objectives, such as growth, income, and security. Like police, investors make a variety of investments to accomplish different objectives. Like police, investors face an endless array of choices about how to invest scarce resources. But investors have one great advantage over police that makes the investors' job much easier: a common currency. Police can have a common currency as well, but only if they create it.

This subsection starts by describing a way in which police can guide targeting decisions with a common currency of return on investment.

TABLE 1

"Good Evidence" Scales for Different Kinds of Police Decisions: Criteria for Strength of Evidence

Strength of Evidence	Kinds of Decisions		
	Targeting	Testing	Tracking
Strong	85% or greater accuracy of predictions in over 5,000 cases	Multiple random control trials producing similar findings in field settings	Monthly audits of measurement systems; less than 5% magnitude of error
Medium	60% or greater accuracy of predictions in over 1,000 cases	One random control trial, or 5–10 controlled quasi experiments with similar findings, in field settings	Annual audits of measurement systems; less than 5% magnitudes of error
Suggestive	Correlations without prospective tests of prediction accuracy	One or two before-after field tests with large effect sizes	Audit within past 3–5 years; less than 5% magnitude of error

It then shows how different methods of predicting returns of investment in that currency will produce much higher or lower rates of error in making targeting decisions, including the new decision tool called "predictive policing," as well as the neglected tool called "solvability factors."

1. *A Crime Harm Index.* The biggest obstacle to reducing crime may be the misleading way in which it is counted: as a raw summation of all crimes, regardless of any differences in public views of the seriousness of harm across crime categories (Rossi et al. 1974). From Washington to London to Delhi, governments earnestly report whether there were more "crimes" this year than last. This count of the crime total treats every crime as equal.

All crimes are not created equal. Some crimes cause horrible injuries and deaths. Others cause scant meaningful harm to anyone, such as possessing a 1-inch joint of marijuana. Crime counts are dominated by such high-volume but minor crimes as shoplifting from large chain stores. In England and Wales, only 10 percent of reported crimes in the 12 months to October 2012 included injury or a sex offense. Yet public officials are pleased to report "success" when total crime drops and displeased when total crime counts rise—even if murders, rapes, and serious woundings stay constant. If EBP accepts this "fungibility fallacy" that all crimes are equal, its value will be greatly restricted.

Police professionals are acutely sensitive to the idea that the weight of harm from crime matters more than a raw count of incidents (Sherman 2011*a*). They know that a car theft is less harmful than a rape, that a £5,000 fraud is less harmful than a stabbing. What they lack is a way to account for these differences in combination across all crimes. This would allow police to know whether total harm from crime in one city is higher this year than last, higher in this part of town than that, or with this offender's record compared to that one's.

The best way to compile meaningful "evidence" about reported crime is to give each type of crime a weight that represents how harmful each type of crime is. Combining crime in this way would create what statisticians call an "index," yielding a single bottom line. The weight can be based on a variety of metrics. The simplest metric would be taken from any sentencing guideline recommendations of the number of days in prison for a first offender convicted of that offense. This would give an approximation of the "pure" weight of harm of the offense itself in contrast to the actual sentence length an offender may

receive—the latter being influenced by the number of prior convictions of the offender.

Summing the total weight for each type of any indicator (such as crime types), then adding the subtotal weights together, is the basis for a commonly used method of constructing an "index" of multiple indicators. A consumer price index (CPI), for example, takes the cost of consumer goods in different categories (food, housing, transportation) and then assigns a weight to those costs based on the average household's budget proportions for each category. If housing costs rise 10 percent but housing is only 33 percent of the family's budget, then the housing increase of 10 percent becomes a 3.3 percent increase in the total CPI. Similarly, a crime harm index (CHI) is a tool for creating just such a bottom line for the harm caused by crime (Sherman 2007, 2010, 2011*a*).

Weighting crimes on the basis of sentencing guidelines can be justified on good democratic grounds as reflecting the will of the people. Almost every sentencing guideline process in the United States and United Kingdom has reflected awareness of opinion polls, debates, and scrutiny by elected officials and news media. The scrutiny may not have been perfect. Yet it remains far closer to the will of the people than any theoretical or even empirical system of weighting that academics might develop. Most important, it is readily available to be applied to any set of crimes, whether for an individual, a community, or a nation.

Any police agency or government can construct its own index with the following seven steps:

1. counting up the number of crimes of each type in an area (or for one offender);
2. multiplying the count for each type by the median number of prison days recommended for crimes of that type by first offenders;
3. calling the product of that multiplication (crime count for a crime type × median days in prison) the HST for the crime type (for harm subtotal of days of prison for that offense type);
4. repeating steps 1, 2, and 3 for every type of crime recorded for the area or person;
5. summing up all HSTs to yield the total crime harm (TCH);
6. creating a standardized CHI for any *area or population* by dividing the estimated population size into the TCH to yield the

CHI in that time period for that area (where population size estimates should ideally include the average daily transient counts, such as [adjusted] arrivals minus departures by train and bus);

7. creating a standardized CHI for any *individual offender* by dividing the person's TCH by the number of years the person has been at risk of committing crime as an adult (usually since age 18), either removing days in prison or not, depending on data availability.

As an example of how a CHI works, consider a community (or an offender) with 100 crimes in the record. Assume that the recommended sentence for a shoplifting crime is 1 day in prison and that the recommended sentence for a first manslaughter is 10 years in prison, or 3,650 days. If all 100 crimes are shoplifting cases, the CHI value of those 100 crimes would be 100 days. If all 100 crimes are manslaughter cases, then the CHI value would be $10 \times 3{,}650 = 36{,}500$. In both cases the crime count is 100, but the range of CHI varies from 100 in the first example to 36,500 in the second. Which would tell voters more about how safe or dangerous their community is? By creating a more sensitive bottom line, a CHI can show much more transparently whether the harm from crime is higher or lower depending on the exact nature of the crimes.

To translate a CHI into an overall targeting plan, police leaders can simply use a checklist of possible targets to pick one or more ways of organizing a resource targeting strategy. Such a strategy can be based on hot spot places, convicted offenders, repeat victims, crime "recruiters" (see below), criminal networks, predicted crime patterns, times of day, days of the week, situations, or crime types. Any of these will do, as long as the analysis is comprehensive. That is, a CHI analysis of all of these units should encompass 100 percent of the CHI for the entire jurisdiction.

Perhaps the greatest value of a CHI would be in using it to forecast how much harm an individual offender, suspect, or defendant is likely to cause in the next 2 years. Such forecasts can help guide the massive use of police discretion to divert cases from prosecution into out-of-court penalties—constituting some 50 percent of all nontraffic penalties in England and Wales (Judge 2010). The legitimacy of policing could be enhanced by using these forecasts to reduce the risk of releasing a very dangerous offender who may, for example, soon murder

someone. Such forecasting can provide missing evidence for the existing system of "triage" in allocating scarce tax revenues for maximum public benefit.

The triage of targeting police resources embraces all police services. Better evidence can make the likely consequences of these hard choices clearer to those who must make them. A standard metric, such as a CHI, for predicting the benefit of one choice over another can make such evidence much easier to apply in practice.

2. *Good and Bad Predictions—or Forecasts.* No matter what currency police use to target police resources, the evidence they need to make good decisions is a reliable forecast of future events. They may prefer to have a prediction, but predictions are far less reliable than forecasts. The difference between a forecast and a prediction is crucial, as the scientists who study earthquakes know to their cost. Silver (2012, p. 149) reports that in seismology,

1. a prediction is a definitive and specific statement about when and where an earthquake will strike: *a major earthquake will hit Kyoto, Japan on June 28.*
2. a forecast is a probabilistic statement, usually over a longer time scale: *there is a 60 percent chance of an earthquake in Southern California over the next thirty years.*

Silver notes that the official position of the US Geological Survey (USGS) is that earthquakes can be forecasted, but they cannot be predicted.

Much the same can be said about evidence-based police targeting: it employs forecasting, not precise predictions, about when and where crimes are likely to occur. In the first published analysis of crime across every address in an entire city, for example, there were many instances in which there was a 100 percent chance of crime at a specific address every 10 days (Sherman, Gartin, and Buerger 1989). What the analysis did not do is to name a time and date when the next robbery would occur at a specific place—nor the name of the robber(s) or the victim(s). This same approach allowed Weisburd, Groff, and Yang (2012) to undertake their remarkable forecasting of crimes over every street segment in Seattle for a 16-year period. This work, for the first time, identified different trajectories of crime by place, showing most kinds of places to have highly stable levels of crime (high, medium, or low), while others were highly variable.

Most crime forecasting takes the form of projecting the future on the basis of the past, or what might be called "post-casting" based on long time periods with stable estimates of risk. The time periods used to make a forecast this way can be 1–15 years; the future time window in which the forecast applies could be 6 months or a year. These forecasts have high levels of accuracy but low levels of precision: they cannot tell police exactly where and when the next crime will occur.

A newer approach may offer both precision and accuracy, at least for high-volume crime types. It uses more short-term analysis of changing crime patterns to make short-term forecasts about where crime concentrations will soon pop up. The originators of this approach, Bowers, Johnson, and Pease (2004), developed ways to forecast the location of burglary hot spots before they had even developed. Yet the method remained within the definition of a forecast rather than claiming to predict the exact date, time, and location of each crime.

By 2010, a new tool for targeting claimed to offer far more exact predictions. PredPol, the predictive policing company, sells police agencies proprietary software that identifies extremely tight bounding of time and place in which crime is predicted to occur. A *New York Times* story on the Santa Cruz field test reported the following: "Based on models for predicting aftershocks from earthquakes, [our equation] generates projections about which areas and windows of time are at highest risk for future crimes by analyzing and detecting patterns in years of past crime data. The projections are recalibrated daily, as new crimes occur and updated data is fed into the program" (Goode 2011). The fact that the USGS says that accurate earthquake prediction is not even possible provides little comfort for PredPol's reliance on earthquake models. At this writing no evidence is available for the accuracy of the forecasts or the crime reduction benefits of using them. But if such predictions do prove to have high reliability, they could make the targeting of police resources more valuable than ever.

What all these approaches offer, at least in principle, are empirical tests of the accuracy of the forecasts. The evolution of evidence-based policing will draw increasing attention to the results of those tests and no doubt compare crime predictions in ways similar to Silver's (2012) own comparisons of election forecasting by different polling companies. The bright line for evidence-based targeting lies not between forecasting and prediction but between quantitative methods and subjective judgments—also known as informed "guessing."

a. Three Methods of Targeting. There are three basic ways to make statements about future harm, whether predictions or forecasts: subjective or "clinical" methods, a checklist of tested factors, and "supercomputer" data mining based on large sample sizes.

By far the most common method is the subjective, intuitive, system I approach based on experience with similar situations (Kahneman 2011). The second method is a "checklist" approach that moves past experience from qualitative to quantitative analysis. It translates prior experience into a more systematic "algorithm" (the word statisticians prefer for a formula or equation) based on a few consistent criteria, forcing the forecaster to answer a checklist of weighted questions to make decisions about predictive factors found in large samples of previous cases. Crime solvability factors (Eck 1979) provide an example of the checklist approach. The third method requires the use of a supercomputer performing massive calculations about each individual case in relation to tens of thousands of similar but slightly different cases. Developed initially by meteorologists, such "data-mining" predictions use massive amounts of data on previous events to predict future events. It is especially useful in forecasting very rare events, such as hurricanes or homicides.

The primary published example of data mining in crime forecasting is the classification of likely murderers with huge samples of previous cases: over 30,000 in Philadelphia (Berk et al. 2009) and 100,000 in England and Wales (Sherman, Cosma, and Neyroud 2012). Like weather forecasting, individual forecasting of crime risks uses the recent growth in supercomputers to find highly specific combinations of predictors that raise the odds of fairly rare events occurring. The Philadelphia forecast identifies the offenders on probation who are most likely to be among the 2 percent of offenders on probation who are charged with murder 75 times more often than the 60 percent of probationers who are classified as low risk. The England and Wales models can correctly classify low-risk offenders with 96 percent accuracy, or 4 percent error in forecasting that they will not be convicted of a serious crime (Sherman, Cosma, and Neyroud 2012).

b. Clinical versus Statistical Predictions. Since at least the advent of Freudian theory, professionals such as psychiatrists have been paid well to make predictions about human behavior. These predictions have been based on detailed analysis of an individual's personality and previous behavior. The method has been highly respected. It is still used

widely to decide whether violent people are insane or criminally responsible for their conduct. Many police agencies also use it to screen police recruits for mental health.

Since the advent of mechanical calculators, a far less expensive approach to prediction has been used by institutions dealing with large volumes of decisions. From insurance companies to university admissions offices, decisions have been made on the basis of classifying people by their numbers: how many accidents, how high a test score, what average grades, how many speeding violations, and so on. These characteristics can be recorded at much lower expense than the cost of having a professional personality assessment, while making exactly the same kind of predictions.

Until Meehl's (1954) comparison of these inexpensive statistical methods to the expensive clinical methods, no one had ever shown which method is more accurate. Since then, no research has ever contradicted his central conclusion. What Meehl found was that statistically validated predictions were always either more accurate than clinical prediction or—at worst—just as good, at a much lower cost. His book reviewed over 20 studies comparing subjective judgments of trained professionals to rather modest combinations of numerical information about the same people, organized according to a rule. Such rules are called algorithms. The statistical predictions based on algorithms in Meehl's comparison were very simple by modern standards, without extensive analysis. Nonetheless, they usually did better than predictions made on the same cases based only on open-ended interviews. Over 200 comparisons have now tested Meehl's conclusions on a wide range of behaviors and phenomena. In some 60 percent of comparisons, statistical forecasting does much better than clinical. In the rest, there is little difference, but statistics remain cheaper than qualitative data for the same accuracy rate (Kahneman 2011, p. 223). There is no convincing case in which clinical forecasting has beaten statistical forecasts.

Until recently this debate had no relevance to policing since police made little use of either kind of prediction in targeting their resources. Yet by the early twenty-first century, police on both sides of the Atlantic had invested in clinical prediction rather than in statistical forecasting. UK police developed national policies for clinical prediction of serious domestic violence cases, as well as risk assessments for sex offenders and other serious criminals. In the United States, problem-

oriented policing picked targets more subjectively than systematically (Goldstein 1990). UK police employed entirely subjective criteria for determining which people released from prison were dangerous enough to justify an expensive Multi-Agency Public Protection Agreement for preventing serious crime.

The cost of subjective methods of crime forecasting often becomes apparent when they are audited against what really happens. An analysis of cases of domestic murder or attempted murder in one major UK police force did not find one case of serious injury committed by someone who was on the list of dangerous people; the lists of thousands of "dangerous" people had excluded all of the fewer than 100 actual attackers (Thornton 2011). Some might say that proves the success of the monitoring program; other might see it as evidence of vast overprediction. Most important, there had been no investment in preventing the attacks that ultimately occurred.

3. *Solvability Factors.* In principle, evidence-based solvability factors could substantially increase detection rates (Eck 1979). At minimum, their use could achieve the same detection rates but at much lower cost. This approach offers a major opportunity for evidence-based targeting. Yet it lacks a crucial piece of evidence that is needed to recommend their widespread adoption.

The most important new evidence about statistical solvability factors would be a randomized trial comparing the overall detection rates of a team of detectives using tested solvability factors to target cases for investigation to those of a team using unstructured discretion to target cases. By randomly assigning all incoming cases to one or the other of the two teams, the experiment would yield a valid estimate of using one system or the other. Even better, such an experiment could use a CHI to weight the value of the cases in the two teams resulting in prosecution. This would create an incentive for the investigators to use harm weightings in combination with solvability factors to decide which cases to investigate. The investigative results could then be compared by both the CHI weight of the crimes solved and the criminal record of the criminals accused.

C. Issues in Testing for Outcomes

Evidence-based policing requires a clear understanding of the difference between "trying" and "testing." Many police agencies say they have "evidence" that a police practice "works" because they have

"tried" it. Yet in most cases, there is no good evidence gained from "piloting" a new practice or "trying it out." The minimal standards of evidence from the kind of test that should be used to guide policy require direct measurement of outputs and outcomes and direct comparison of those outcomes with a control group's outcomes. There is no better example than the case of the free chocolate.

1. *Trying versus Testing: The Chocolate Case.* The Dorset (UK) police tried a very interesting idea in Bournemouth in the early twenty-first century, when all pubs still closed at 11:00 p.m. sharp. What happened between 11:00 and midnight each evening was reportedly a big spike in the hourly number of fights. Many fights happened while people were waiting in line at fish and chips shops, hungry for their first food after drinking beer for 5 or 6 hours. A creative police leader reasoned that this pattern—or problem, as Goldstein (1979) defined it—was caused by low blood sugar, making people more irritable. The Dorset police then tried handing out free chocolates to people waiting in line for their takeaway food. Media reports at the time quoted police saying the idea "worked."

This brilliant idea appears, however, never to have been tested at the standard required by evidence-based policing. I found no report of any counts of fights between 11:00 and midnight, either before or after the free chocolate policy. The first standard of a scientific test is that there be consistently defined and reliable measurement of both the problem to be solved—fights, in this case—and the solution itself—in this case, chocolate. As a public health doctor might ask, what was the dosage of chocolate handed out? How many individuals took the chocolate? How many ate it? How many nights a week was the chocolate handed out? Was it handed out equally at every fish and chips shop? If not, were there fewer fights where there was more chocolate? Over a long enough time period to constitute a statistically "powerful" test, by what percentage did the average number of fights in Bournemouth between 11:00 and midnight go down?

Even if all those questions of measurement were answered, however, there would still be no assurance that the chocolates had "worked" unless the second standard of a test were to be met: a fair comparison to a control group that did not receive the free chocolates. A control group could be selected from the central entertainment area of a similar city nearby, which would constitute a "level 3" test on the Maryland Scale of Scientific Methods (Sherman et al. 1998), the minimum level

recommended for evidence on which to base policy. Even better would have been the identification of 100 fish and chips shops in Dorset's entertainment areas in multiple cities and towns, with 50 of them selected at random for the free chocolate treatment.

The distinction between "trying" and "testing" has become increasingly clear to police leaders, if not the rank and file, in the United States and United Kingdom over the past four decades. Yet the absolute comprehension level remains low. It is hard to teach these concepts to senior police leaders in a short conversation, even though they now feel obliged to discuss them as if they understand them perfectly. The concept of testing has not, until recently, been a part of police training, for leaders or street officers. But since 1996, the Cambridge University Institute of Criminology has been teaching these ideas to a large proportion of all the leaders who became chief constables. The success of that teaching has been far from perfect. But it is no accident that the UK police leadership spends far more time talking about the evidence for policies than US police leaders do.

What may work better than training is to have a much greater volume of experiments going on in police agencies than ever before. Since 2010, this has been happening in the United Kingdom if not in the United States. The downside of this development may be less rigor and care in the conduct of experiments, for reasons discussed below. These problems may be outweighed, however, by the increased awareness they promote of the difference between trying and testing: between evidence-free and evidence-based policing.

2. *What Works: Testing Practices, Using Tests.* There are two ways for police to find out whether a specific police practice is effective. One is to have someone look for relevant evidence in the growing police research repositories (see the Appendix). The other is for police to conduct tests in their own agencies.

In an ideal world, there would be so many tests of police services (as there are of medical services) that EBP would operate primarily by "looking up" tests rather than doing them. The EBP Matrix (Lum, Koper, and Telep 2010) and other tools already make it easy to do this. Other tools for easy "lookups" are under development, some of them designed for access via smart phones. The College of Policing may even develop the ideas of computerized medicine for instant access to relevant evidence (Hafner 2012).

In the foreseeable future, however, many expensive police services

will remain untested. The costs of conducting tests are minimal in comparison to the costs of delivering untested services. And every time a police agency contributes a high-quality cost-effectiveness test to the global police literature, the entire world will gain.

Unlike such agencies as the UK National Health Service, police are not yet required to have randomized trials justify large portions of the police budget. But it would be major progress if at least 10 or 20 percent of a police operating budget could be based on good evidence. The most relevant question for any police leader is, "which of our services is so expensive and of such doubtful value that it is worthwhile to do our own local randomized controlled trial?"

a. Generalizability of Evidence. There is also a scientific question of how reliably research in one police agency (or more) will predict effectiveness in any other agency. The present state of science cannot say whether one, 10, or even 20 strong tests are enough to generalize to all democracies, or even across the country where the tests were done. This problem is confounded by the wide range of results reported in such systematic reviews as the meta-analysis of the effects of hot spot policing (Braga, Papachristos, and Hureau 2012). While the average effect is beneficial, the range of effects is very great. Whether another agency implementing some form of hot spot policing will achieve a large or small effect remains highly uncertain from the available research.

The external validity of results is also challenged by the wide variety of practices tested, even under a common rubric like "hot spots policing." The range of specific police methods of policing included in that review is almost as great as the range of effect sizes (Braga, Papachristos, and Hureau 2012). While the review is nonetheless a major contribution to the literature, the wide range of research designs unavoidably strains the scientific definition of "replication" on which the statistical procedures are premised. For example, while the hot spot review found 11 randomized trials of hot spot policing (and more have been finished since), only one of them tested a substantial increase in marked car patrols in street corner hot spots. Other designs are of interest. But for a police leader considering a 100 percent increase in hot spot patrols at street corners, any other design is a different research question.

For the foreseeable future, the best evidence on outcomes police agencies can get will come from conducting their own experiments.

Most readers may wonder whether that idea is feasible or is a fantasy. The answer depends entirely on who does the experiments: who designs them, leads them, and analyzes them. Those tasks have been the traditional domain of police scholars, with police practitioners as their partners. Yet a rapidly growing appetite for good evidence will soon swamp the available numbers of experimental criminologists unless they reinvent their partnerships with policing.

b. Who Shall Test: The Police Themselves? From 1970 until 2010, police testing was largely orchestrated by academics in universities who had found police chiefs to partner with them. The experiments typically involved large grants from foundations or national governments' research agencies (US National Institute of Justice, UK Home Office, Australian Research Council). These grants usually provided for well-trained and closely supervised doctoral students to work with police in the delivery and measurement of the experimental variables, ensuring independent quality control and monitoring of the data collection. Senior academics would then analyze the data and write the report. Since the number of academics interested in doing all this work was limited, the maximum number of experiments was limited by human resources as much as by the tiny overall funding stream for such experiments.

Since around 2010, both funding and leadership barriers have started to fall. Funding has been increasingly provided by police agencies, subsidized by universities teaching police executives. The leadership of police experiments has been moving from professors to "pracademics," as police practitioner-academics increasingly describe themselves (e.g., Mitchell 2012). While professional research scientists still play a strong role in designing and analyzing police experiments, more of the massive labor of organizing people to deliver good experiments is being shouldered by experienced police (e.g., Telep, Mitchell, and Weisburd 2012). Many pracademics are launched on a path to earn a doctoral degree, some in midcareer and some retired after 30 years in policing. Whatever their background, they are rapidly increasing the numbers of people who can produce good evidence on policing.

This change is especially visible in the United Kingdom, where Cambridge graduates form a critical mass to pioneer the new model for testing police practices. This model has made possible a far greater volume of experimentation because police officers from sergeant to police chief have designed and delivered their own experiments. This

benefit must be weighed against important issues of quality control in the integrity of the random assignment, the measures of outcomes, and other important issues. Even innovative solutions like an online "randomizer" (Ariel, Vila, and Sherman 2012) may pose surprising errors unless the experiment is closely monitored by a site manager (Neyroud and Slothower 2013).

D. Issues in Tracking Police Outputs and Outcomes

Three key issues shape the growth of tracking police outputs and outcomes. One is how police leaders measure what police officers do. A second is how they compare measures of performance across officers and units. The third is how they use tracking data to deliver better policing by evidence-based management.

1. *Measuring What Police Do.* For all the progress that COMPSTAT has brought policing, it is striking how little measurement it has used of what police do. In 1999–2005, for example, the Philadelphia Police Department's COMPSTAT never reviewed data on where police patrolled, where they made arrests, where they conducted stop-and-frisks, or even how many police were scheduled to work by time or day in relation to the hourly frequency of crime in any police district. In social science terms, COMPSTAT was a discussion of dependent variables without any review of independent variables. They discussed crime patterns endlessly each week, with rarely any evidence on policing.

Since then, technologies such as GPS have made such measurement even easier. All that is required is a commitment to tracking policing along with crime and an investment in information technology to produce the data and graphics. Mapping police presence in relation to crime harm, for example, would produce an algorithm that could identify outliers. Wherever a patrol district deployed its patrols (or arrests) in too great a departure from the occurrence of crime, a list of such "exceptions" can be generated for police managers. If they fail to correct the discrepancies, the lists can be reviewed in aggregate at COMPSTAT meetings.

2. *Comparing Police Performance.* One reason police activity is undermeasured may be the risk of making inappropriate comparisons with disciplinary implications. The number of arrests per officer, for example, is likely to be much higher in certain areas than in others, if only because there is more crime per square foot in some areas. Yet

there are many ways to adjust for these differences in order to create fair comparisons across officers and units.

The critical tool for fair comparisons is a *risk model for each policing area*. Such a model can be computed on the basis of any large sample, such as the hundreds of local policing areas in England and Wales. The models can address both police outputs and outcomes. Using social, economic, and demographic data, the models can forecast how much work police will be asked to do reactively by citizens and how much work they may choose to initiate proactively in relation to social conditions. The models can estimate how many calls will be dispatched, how many arrests may be made, and even how much force police will use with how much injury to citizens and police.

Once a risk model is found to be reliable in forecasting outputs or outcomes, it can detect major changes in delivery of those measures over what was predicted. If police produce more proactive work than predicted or less, the model can say so. This could be done in a way that could confirm or disprove the success of an area commander in delivering specific changes in police patrols, arrests, problems solved, or any objective for police outputs.

The greatest benefit of a fair comparisons model would be to link changes in outputs to changes in crime harm (CHI). Within a range of error, these models can forecast the likely CHI (or CHI per capita) in each area for each year. If a change in outputs appears linked to changes in CHI outcomes, then police leaders can begin to develop skill in relating these two tools at a strategic level. In contrast, if there is no change in outputs but the CHI gets worse, that can be taken as evidence of sudden changes in social conditions rather than as a failure of policing. If police outputs worsen and the CHI rises, that can be taken as evidence that police failures actually caused an increase in harm.

This kind of macro-level strategic management is already the substance of discussion in many COMPSTAT meetings. What it lacks is any kind of evidence base. There are a few published studies, for example, of the effect of the ratio of arrests to crime (output) on crime rates (outcome) themselves (Tittle and Rowe 1974; Brown 1978; Chamlin 1991), but they would quickly be replaced by current studies within large police agencies. Better evidence could lead to better strategic decisions, with less crime harm as a result.

3. *Delivering Policing with Evidence-Based Management.* The great-

est value of measuring police performance comes from leaders taking immediate corrective action. All too often, police leaders are flying blind about whether policies or operations are being implemented. Tracking evidence provides independent audits on a timely basis about whether the plans are being delivered. If they are not, then commanders can be transferred and tighter control can be imposed. If they are, then rewards can be given and medals can be pinned.

The use of tracking to focus on implementation can have major consequences. For example, there is much controversy about whether stop-and-questioning in minority areas reduces crime. By using tracking evidence for deciding whether stop-and-question has been reducing robberies in minority group areas, a police leader can ask to have stop-and-question encounters plotted against robberies. If the answer is that very few stops have actually been done, the leader can take stronger action to ensure that the policy is carried out. If the policy has been delivered, the leader can ask what level of crime reduction is needed to be cost-effective. At the same time, tracking may raise the question of police legitimacy. Are complaints up since stops have increased? Are public surveys needed? Should people who have been stopped be interviewed to assess their reactions? Should training be refreshed on how to speak to suspects? Tracking can often raise questions that would otherwise never be asked. Asking them could, in turn, substantially improve service delivery.

Tracking methods can even be subjected to testing for outcomes. In a Dallas experiment, for example, Weisburd, Groff, and Yang (2012) randomly assigned GPS information to some middle managers (and not others) showing how much time officers were spending in their assigned hot spots. The experiment suggested that the assignment of that information led to less crime in the high-crime hot spots. Other experiments with tracking systems could profitably use similar designs across police districts, across patrol beats, or even (in large police agencies such as India's or Chile's with hundreds of police stations) across police districts themselves.

E. Connecting the Three Ts

Thinking of triple-T as a triangle may allow police leaders to think about evidence in a more integrated way. Agencies can use the triangle to design checklists of all the numbers that must be examined simultaneously in EBP. Like the pilots who look at all the dials on the

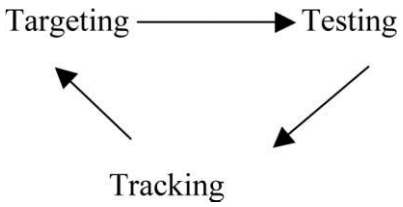

FIG. 2.—The dynamic process of triple-T

dashboard of a B-747 cockpit, police leaders must be aware of multiple sources of threat to smooth performance. When all three Ts are able to be seen at once, the combined triple-T can become a dynamic process: targets can be adjusted as crime goes down on some targets and rises on others, tested treatments can change as the types of targets shift, and tracking may look at different delivery indicators as both targets and practices shift (fig. 2). No police agency stands still for a day, let alone for a 3-year plan. Triple-T is a *plan for a plan*, for a constantly updated plan, one that keeps changing as the facts change (O'Connor 2012).

Connecting the Ts may be the best response to a major challenge to EBP: quantitatively derived conclusions are often hard to sell. When the conclusion directly challenges conventional thinking, acceptance of evidence requires nothing less than a culture change. This phenomenon has been observed repeatedly in medicine (Giluk and Rynes-Weller 2012), baseball (Lewis 2003), and politics (Issenberg 2012). Yet these cases also offer much hope that experience and evidence can work together (Silver 2012, chap. 3). That is what can happen to make culture change more likely to succeed. The open window may be a need to cut budgets and to do more with less.

By connecting policy decisions on tested services with policy decisions on what to target, police leaders can develop a clear dialogue about hard choices. Operational police may accuse them of "robbing Peter to pay Paul," as if all priorities are of equal importance. But the use of evidence to make those decisions can disprove that claim. By combining the seriousness of the harm being targeted with the effectiveness of the police practice being considered for a target, police leaders can compare two different choices for their likely cost-benefit ratios. If spending $10,000 can yield a 25 percent reduction in the harm

of a robbery pattern but only a 10 percent reduction in the harm of a burglary pattern, the better choice may seem obvious. It is not. The key evidence in such a choice would be the absolute levels of harm associated with the burglary pattern and the robbery pattern. Creating or using such a metric is not routinely done in EBP—yet. But it could be the key to pulling triple-T into a viable operating system. That key is the idea of a crime harm index, which may ultimately be the key to sustaining EBP.

IV. Future Prospects for Evidence in Policing

> The best way to predict your future is to create it.
> (Attributed to Abraham Lincoln)

By 2013, more people than ever before were trying to create the future of evidence-based policing. The Society of Evidence-Based Policing had over 600 members. The new professional College of Policing had a mandate from the UK government to develop and share "evidence on what works in policing." The George Mason University Center for Evidence-Based Crime Policy joined with the Scottish Institute of Police Research to communicate EBP to policy makers in Washington. Newly elected police and crime commissioners in England and Wales organized an all-day seminar at Cambridge University to gain an introduction to the evidence on policing. Over 75 percent of senior leaders at Scotland Yard said they had at least heard of EBP (Stanko 2013).

Yet the same study at Scotland Yard found that 20 percent of its leaders said they would not change a policy if the evidence showed it did not work (Stanko 2013). Many, if not most, operational police officers remained skeptical that research could help improve policing. Laws in the majority of US states, as of 2013, still required police to make arrests when evidence shows that they would only increase crime rather than prevent it (Sherman 1992*a*). Evidence-based policing in 2013 still had a long way to go. As Kahneman (2011) and others have documented, our intuitive preference for "fast thinking" with an emotional system I will always cause resistance to conclusions based on more comprehensive "slow thinking" by our system II reasoning. Which system will win depends on a larger cultural struggle for the "branding" of policing as a complex profession.

For those who take Lincoln's advice seriously, it may be useful to

suggest 10 ways to create a stronger future for using evidence in policing. The suggestions are speculations, informed by a deep engagement in the context: the kind of experience that is essential for making the best use of evidence. The suggestions aim at institutional decisions as the outcomes. But they are also aimed at the individuals whose actions can make institutions change their policies.

The key institutions are the College of Policing in the United Kingdom, the state Police Officer Standards and Training Boards in the United States, the Australian Institute for Police Management, and the Australia New Zealand Policing Advisory Agency. These are all professionally driven institutions with governmental funding capable of making substantial changes in the incentives and organizational culture for better use of research evidence. Until recently they were largely unconnected to the academic community. But the invention of a new professional body in the cradle of Anglo-Celtic policing is committed to forging that connection, with a working agenda for police-university relationships. It is not impossible that professional bodies in other countries may follow suit. Unless otherwise indicated, each suggestion below is aimed at the institutions named above.

1. *A graduate-level education in evidence-based policing should be required for all chief police executives.* Starting at the top is a path of least resistance from police labor organizations and an ideal first step toward a "total evidence" strategy of the kind Scotland Yard has adopted. Police Commissioner Sir Bernard Hogan-Howe, who approved that plan, not only has a master of business administration and a law degree but also completed the graduate-level Cambridge Police Executive Programme that teaches evidence in policing.

2. *A graduate-level knowledge of evidence-based policing should be required for all police managers at ranks equivalent to chief inspectors in UK policing and captains in the New York or Los Angeles police.* An early second step toward a culture of evidence in policing would be to focus ambitious officers on the first major step to police command. Knowing that research evidence is constantly growing could foster a habit of keeping up, a habit that may continue even after completing a degree or taking an exam, or doing whatever is established as the requirement for promotion to that level. Most important may be the substantial impact this requirement would have on the demand for graduate education.

3. *Promotion to first-line supervision* (*e.g.*, *sergeant*) *should be based in equal parts on knowledge of the law and of the evidence on what works in*

policing. Perhaps some time after the first two suggestions are adopted, a cultural familiarity with EBP will ease a transition to changing the content of promotional examinations. Once knowledge of evidence becomes essential for any promotion at all in policing, it will greatly broaden the incentive to study evidence and use it in police operational decisions. Research could even be cited in arrest reports and court testimony, thereby promoting a stronger public image of the complex knowledge police must master.

4. *Completion of probation at the end of police recruit training should be based on an examination that includes equal parts on knowledge of the law and of the evidence on what works in policing.* Once EBP becomes an expected part of promotional examinations, it will be less difficult to add it to the final requirement for permanent police officer status (if that concept itself survives). It will also enhance the use of evidence in police operations since officers at all rank levels would then be tested on their knowledge. It would also mean that research evidence would have to be taught in police recruit schools.

5. *Police professionals at all ranks should be able to achieve international certification of advanced knowledge by an exam-based Fellowship of the College of Policing* (*to be designated an FCP*). The practice of advanced testing of highly skilled practitioners is well established in medicine and other professions; US doctors have an equivalent process of becoming board certified. Since the sixteenth century, British professional colleges have been administering such examinations and certifying knowledge for people from all over the world. Online educational preparation will make such examinations even easier to pass, providing free access to knowledge worldwide.

6. *Universities should create faculties of policing that would eventually operate like medical schools.* The growth of demand for police evidence requires far greater commitment by universities to this kind of public service. The rise of "pracademics" makes it even more important since the technical support they need to do experiments at a high standard will grow rapidly.

7. *Academic positions in faculties of policing should include "clinical" lectureships and professorships modeled on those in medical schools, in which academics split their time equally between working on police operational matters and teaching-research activity in the university.* By investing in partial funding of clinical professorships, police agencies or the institutions

named above could speed up the process of creating more police faculties.

8. *Promotion to senior police posts, as with senior medical positions, should require demonstrated achievements in research and teaching within the profession.* This suggestion builds on the recommendations of the UK Home Secretary (May 2012). It could be achieved with evidence of publications in peer-reviewed journals or teaching in university faculties or police training institutions.

9. *The College of Policing (UK) should create a global example of a process for approving and recommending evidence-based police practices, engaging police leaders at all ranks with researchers, victims, and others affected by police policies in deciding what works, when, and for whom.* The transparency of evidence is a hallmark of science, just as openness to public demands is a hallmark of policing. The model of the process used by the United Kingdom's National Institute for Health and Clinical Evidence is one that can be widely adopted (Sherman 2009).

10. *Every nation should develop its own inventory of evidence-based practices, based on its own domestic police research, international research, approved practices of the College of Policing, and a list of similar institutions in other nations.* Once a process is adopted to review evidence for policy decisions, the dearth of such evidence in most countries will become clear. There is no point in American police research providing the only evidence for policing in New Zealand, India, or Hong Kong. Each and every country may have its own social and cultural context that affects the transportability of evidence of what works. The only way to be certain of how police practices work in each country is to test those practices in that country.

Since the twenty-first century began, the pace at which police use evidence has grown geometrically. That pace could still slow, reverse, or be crippled by a hijacking of the meaning of "evidence" in police practice. But if even half the suggestions offered here are adopted, evidence-based policing will continue to grow. That trend may ultimately provide many tangible benefits to police officers, making their work more legitimate and meaningful. A majority of neither police nor the public may ever accept or even understand the benefits of evidence, but popularity is not the right test for the success of EBP. The sole test is whether it benefits the public. All the evidence suggests that it will.

APPENDIX

The resources for locating external tests of police services are growing rapidly. In the past decade, the Campbell Collaboration (hosted by the Norwegian government) has completed over 20 systematic reviews of multiple tests of practices relevant to crime and policing. A "systematic review" in this context is a comprehensive inventory of all possible tests of a specific practice, which then examines the pattern of results to conclude whether the program is effective where it has already been tested. The results in other countries do not necessarily predict results in all countries. But the information in these reviews is highly detailed, reporting exactly where the tests were done and how. Reviews are in PDF files that can be downloaded for free from http://www.campbellcollaboration.org. As of late 2012, the reviews included such police services or issues as problem-oriented policing, hot spots policing, focused deterrence, second responders to domestic violence cases, gun crime prevention patrols, effects of DNA testing on detection rates, community-oriented policing, legitimacy in policing, "broken windows" policing, and crime displacement from place-targeted crime prevention programs.

A second resource is less rigorous and comprehensive within subjects but has broader coverage of topics and tested police services: the US Department of Justice website at http://www.crimesolutions.gov. This site provides an accessible, test-by-test summary of what was tested and how.

A bolder model of retrieving police field tests is the EBP Matrix posted at George Mason University in Virginia, with over 100 tests that can be retrieved by a point-and-click scan of the study summaries represented as dots on the matrix (Lum, Koper, and Telep 2010). This matrix can be found at http://gemini.gmu.edu/cebcp/matrix.html. It is accessible on three dimensions: scope of target (individuals to nation), specificity of police service (general to specific), and proactivity of police service (from citizen-initiated to police-initiated).

REFERENCES

Angel, Caroline M. 2005. "Crime Victims Meet Their Offenders: Testing the Impact of Restorative Justice Conferences on Victims' Post-traumatic Stress Symptoms." PhD dissertation, University of Pennsylvania, Department of Criminology and School of Nursing.

Ariel, B., J. Vila, and L. Sherman. 2012. "Random Assignment without Tears: How to Stop Worrying and Love the Cambridge Randomizer." *Journal of Experimental Criminology* 8(2):1–16. DOI: 10.1007/s11292-012-9141-4.

Banton, Michael. 1964. *The Policeman in the Community*. London: Tavistock.

Bennett, Trevor, Katy Holloway, and David P. Farrington. 2006. "Does Neighborhood Watch Reduce Crime? A Systematic Review and Meta-Analysis." *Journal of Experimental Criminology* 2:437–58.

Berk, Richard A., Alec Campbell, Ruth Klap, and Bruce Western. 1992. "The Deterrent Effect of Arrest in Incidents of Domestic Violence: A Bayesian Analysis of Four Field Experiments." *American Sociological Review* 57:698–708.

Berk, Richard A., Lawrence W. Sherman, G. Barnes, E. Kurtz, and L. Ahlman. 2009. "Forecasting Murder within a Population of Probationers and Parolees: A High Stakes Application of Statistical Learning." *Journal of the Royal Statistical Society: Series A, Statistics in Society* 172:191–211.

Berkow, Michael. 2011. Lecture to the National Police Academy, Hyderabad, India, June 10.

Blumstein, Alfred, Jacqueline Cohen, Jeffrey A. Roth, and Daniel Nagin. 1986. *Criminal Careers and "Career Criminals."* Washington, DC: National Academies Press.

Bottoms, Anthony, and Justice Tankebe. 2012. "Beyond Procedural Justice: A Dialogic Approach to Legitimacy in Criminal Justice." *Journal of Criminal Law and Criminology* 102:119–253.

Bowers, Kate, Shane Johnson, and Ken Pease. 2004. "Prospective Hot-Spotting: The Future of Crime Mapping?" *British Journal of Criminology* 44(5): 641–58.

Braga, Anthony A., Andrew V. Papachristos, and David M. Hureau. 2012. "The Effects of Hot Spots Policing on Crime: An Updated Systematic Review and Meta-Analysis." Campbell Systematic Reviews 2012:8. DOI: 10.4073/csr .2012.8.

Braga, Anthony A., and David L. Weisburd. 2010. *Policing Problem Places*. New York: Oxford University Press.

———. 2012. "The Effects of Focused Deterrence Strategies on Crime: A Systematic Review and Meta-Analysis of the Empirical Evidence." *Journal of Research in Crime and Delinquency* 49(3):323–58.

Bratton, William J. 1998. *Turnaround: How America's Top Cop Reversed the Crime Epidemic.* New York: Random House.

Brown, Don W. 1978. "Arrest Rates and Crime Rates: When Does a Tipping Effect Occur?" *Social Forces* 57:671–82.

Buerger, Michael E. 1993. "Convincing the Recalcitrant: An Examination of the Minneapolis RECAP Experiment." PhD dissertation, Rutgers University, School of Criminal Justice.

Chamlin, Mitchell B. 1991. "A Longitudinal Analysis of the Arrest-Crime Relationship: A Further Examination of the Tipping Effect." *Justice Quarterly* 8:187–99.

Christopher, Warren, ed. 1991. *Report of the Independent Commission on the Los Angeles Police Department*. Los Angeles: Diane Publishing.

Cohn, Ellen G. 1993. "The Prediction of Police Calls for Service: The Influence of Weather and Temporal Variables on Rape and Domestic Violence." *Journal of Environmental Psychology* 13:71–83.

Collins, Randall. 2004. *Interaction Ritual Chains.* Princeton, NJ: Princeton University Press.

Croft, Elizabeth B. 1985. "Police Use of Force: An Empirical Analysis." PhD

dissertation, State University of New York at Albany, School of Criminal Justice.
Davis, Robert C., David Weisburd, and Bruce Taylor. 2008. "Effects of Second Responder Programs on Repeat Incidents of Family Abuse." Campbell Systematic Reviews 2008:15. DOI: 10.4073/csr.2008.15.
Dunford, Franklyn W. 1990. "System-Initiated Warrants for Suspects of Misdemeanor Domestic Assault: A Pilot Study." *Justice Quarterly* 7(4):631–53.
Eck, John E. 1979. *Managing Case Assignments: The Burglary Investigation Decision Model Replication.* Washington, DC: Police Executive Research Forum.
Englefield, Ashley E. 2013. M.St. thesis, Cambridge University, Institute of Criminology.
Eyle, Alexandra. 2001–2. "Mayor Rocky Anderson Talks about What It's Like to . . . Drop the D.A.R.E. Program." *Reconsider Quarterly* 1(4):12–13.
Farrar, Tony. 2013. M.St. Thesis, Cambridge University, Institute of Criminology.
Farrell, Graham. 1995. "Preventing Repeat Victimization." In *Building a Safer Society: Strategic Approaches to Crime Prevention*, edited by Michael Tonry and David P. Farrington. Vol. 19 of *Crime and Justice: A Review of Research*, edited by Michael Tonry. Chicago: University of Chicago Press.
Ferraro, Kathleen J. 1989. "Policing Woman Battering." *Social Problems* 36:61–74.
Foa, Edna. 1997. "Trauma and Women: Course, Predictors, and Treatment." *Journal of Clinical Psychiatry* 58:25–28.
Fyfe, James J. 1980. "Administrative Interventions on Police Shooting Discretion: An Empirical Examination." *Journal of Criminal Justice* 7(4):309–23.
———. 1982. "Blind Justice: Police Shootings in Memphis." *Journal of Criminal Law and Criminology* 73(2):707–22.
Giluk, Tamara L., and Sara Rynes-Weller. 2012. "Research Findings Practitioners Resist: Lessons for Management Academics from Evidence-Based Medicine." In *The Oxford Handbook of Evidence-Based Management*, edited by Denise M. Rousseau. New York: Oxford University Press.
Glaser, Barney, and Anselm Straus. 1967. *The Discovery of Grounded Theory.* New York: Aldine.
Goldstein, Herman. 1979. "Improving Policing: A Problem-Oriented Approach." *Crime and Delinquency* 25:236–58.
———. 1990. *Problem-Oriented Policing.* New York: McGraw-Hill.
Goode, Erica. 2011. "Sending the Police before There's a Crime." *New York Times*, August 15. http://www.nytimes.com/2011/08/16/us/16police.html?_r=0.
Greenberg, Bernard, and Karen Lang. 1973. *Enhancement of the Investigative Function*. Vol. 4, *Burglary Investigative Checklist and Handbook*. Stanford, CA: Stanford Research Institute.
Greenwood, Peter, and Joan Petersilia. 1975. *The Criminal Investigation Process.* Santa Monica, CA: RAND.
Hafner, Katie. 2012. "For Second Opinion, Consult a Computer?" *New York*

Times, December 3. http://www.nytimes.com/2012/12/04/health/quest-to-eliminate-diagnostic-lapses.html?pagewanted=all&_r=0.

Her Majesty's Inspectorate of Constabulary. 2010. *Valuing the Police: Policing in an Age of Austerity*. London: Her Majesty's Inspectorate of Constabulary.

Howarth, David M. P. 2010. Address to the Third Cambridge-NPIA International Conference on Evidence-Based Policing. Cambridge University, July 4.

Hume, David. 1748. *An Enquiry Concerning Human Understanding*. http://www.gutenberg.org/.

Issenberg, Sasha. 2012. *The Victory Lab: The Secret Science of Winning Campaigns*. New York: Crown.

Jackson, Jonathan, and Ben Bradford. 2009. "Crime, Policing and Social Order: On the Expressive Nature of Public Confidence in Policing." *British Journal of Sociology* 60(3):493–521.

Johnson, Shane D., and Kate J. Bowers. 2004. "The Burglary as Clue to the Future: The Beginnings of Prospective Hot-Spotting." *European Journal of Criminology* 1:237–55.

Johnson, Steven Berlin. 2001. *Emergence: The Connected Lives of Ants, Brains, Cities, and Software*. New York: Scribners.

———. 2010. *Where Good Ideas Come From: The Natural History of Innovation*. New York: Riverside.

Judge, Igor. 2010. "Summary Justice In and Out of Court: The Police Foundation's John Harris Memorial Lecture." Lecture presented at Draper's Hall, London, July 7. http://www.judiciary.gov.uk/Resources/JCO/Documents/Speeches/lcj-speech-john-harris-lecture.pdf.

Kahneman, D. 2011. *Thinking, Fast and Slow*. New York: Farrar, Straus & Giroux.

Kansas City Police Department. 1977. *Response Time Analysis: Executive Summary*. Kansas City: Kansas City Police Department.

Kelling, George L., Antony Pate, Duane Dieckman, and Charles E. Brown. 1974. *The Kansas City Preventive Patrol Experiment: A Summary Report*. Washington, DC: Police Foundation.

Kelling, George L., Antony Pate, Amy Ferrara, Mary Utne, and Charles E. Brown. 1981. *The Newark Foot Patrol Experiment*. Washington, DC: Police Foundation.

Kennedy, David M. 2008. *Deterrence and Crime Prevention*. London: Routledge.

———. 2011. *Don't Shoot: One Man, a Street Fellowship, and the End of Violence in Inner-City America*. New York: Bloomsbury.

Koper, Christopher S. 1995. "Just Enough Police Presence: Reducing Crime and Disorderly Behavior by Optimizing Patrol Time in Crime Hot Spots." *Justice Quarterly* 12:649–72.

Koper, Christopher S., and Evan Mayo-Wilson. 2006. "Police Crackdowns on Illegal Gun Carrying: A Systematic Review of Their Impact on Gun Crime." *Journal of Experimental Criminology* 2:227–61.

Kubrin, Charis E., Steven F. Messner, Glenn Deane, Kelly McGeever, and

Thomas D. Stucky. 2010. "Proactive Policing and Robbery Rates across US Cities." *Criminology* 48:57–97.
LaFave, Wayne. 1965. *Arrest: The Decision to Take a Suspect into Custody*. Boston: Little, Brown.
Larson, Richard C. 1976. "What Happened to Patrol Operations in Kansas City? A Review of the Kansas City Preventive Patrol Experiment." *Journal of Criminal Justice* 3:267–97.
Lewis, Michael. 2003. *Moneyball: The Art of Winning an Unfair Game*. New York: Norton.
Lum, Cynthia, Christopher S. Koper, and Cody D. Telep. 2010. "The Evidence-Based Policing Matrix." *Journal of Experimental Criminology* 7:3–26.
Manning, Peter K. 1992. "Information Technologies and the Police." In *Modern Policing*, edited by Michael Tonry and Norval Morris. Vol. 15 of *Crime and Justice: A Review of Research*, edited by Michael Tonry. Chicago: University of Chicago Press.
Maple, Jack. 1999. *The Crime Fighter: How You Can Make Our Community Crime-Free*. New York: Doubleday.
Martin, Susan E., and Lawrence W. Sherman. 1986. "Selective Apprehension: A Police Strategy for Repeat Offenders." *Criminology* 24:155–73.
May, T. 2012. Speech by the Home Secretary to the Police Superintendents' Association Conference, September 11. http://www.homeoffice.gov.uk/media-centre/speeches/supers2012?version=2.
Mazerolle, Lorraine, Emma Antrobus, Sarah Bennett, and Tom R. Tyler. 2013. "Shaping Citizen Perceptions of Police Legitimacy: A Randomized Field Trial of Procedural Justice." *Criminology* 51:33–63.
McEwen, Thomas, Edward F. Connors, and Marcia Cohen. 1986. *Evaluation of the Differential Police Response Field Test*. Washington, DC: National Institute of Justice.
Meehl, Paul E. 1954. *Clinical vs. Statistical Prediction: A Theoretical Analysis and a Review of the Evidence*. Minneapolis: University of Minnesota Press.
Millenson, M. 1997. *Demanding Medical Excellence: Doctors and Accountability in the Information Age*. Chicago: University of Chicago Press.
Mitchell, Renee. 2012. Presentation to the Cambridge Police Executive Programme, April 17.
Monkkonen, Eric H. 1992. "History of Urban Police." In *Modern Policing*, edited by Michael Tonry and Norval Morris. Vol. 15 of *Crime and Justice: A Review of Research*, edited by Michael Tonry. Chicago: University of Chicago Press.
Murray, Alexander. 2009. M.St. thesis, University of Cambridge, Institute of Criminology.
Nevin, Rick. 2000. "How Lead Exposure Relates to Temporal Changes in IQ, Violent Crime, and Unwed Pregnancy." *Environmental Research* 83:1–22.
Neyroud, Peter, and Molly Slothower. 2013. *Operation Turning Point: Progress Report*. Cambridge: Jerry Lee Centre of Experimental Criminology, Institute of Criminology, University of Cambridge.

O'Connor, Denis. 2012. Lecture to the Cambridge Police Executive Programme, Institute of Criminology, University of Cambridge, September.
Parkinson, John. 2013. M.St. thesis, University of Cambridge, Institute of Criminology.
Pate, Antony M., and Edwin E. Hamilton. 1992. "Formal and Informal Deterrents to Domestic Violence: The Dade County Spouse Assault Experiment." *American Sociological Review* 57:691–97.
Pease, Ken. 1998. "Repeat Victimisation: Taking Stock." Crime Detection and Prevention Series Paper 90. London: Home Office.
Perry, Cheryl L., Kelli A. Komro, Sara Veblen-Mortenson, Linda M. Bosma, Kian Farbakhsh, Karen A. Munson, Melissa H. Stigler, and Leslie A. Lytle. 2003. "A Randomized Controlled Trial of the Middle and Junior High School DARE and DARE Plus Programs." *Archives of Pediatrics and Adolescent Medicine* 157:178–84.
Petrosino, Anthony, Carolyn Turpin-Petrosino, and Sarah Guckenburg. 2010. "Formal System Processing of Juveniles: Effects on Delinquency." Campbell Systematic Reviews 2010:1. DOI: 10.4073/csr.2010.1.
Pittet, Didier, and John M. Boyce. 2001. "Hand Hygiene and Patient Care: Pursuing the Semmelweis Legacy." *Lancet Infectious Diseases* 1:9–20.
Police Executive Research Forum. 2008. *Violent Crime in America: What We Know about Hot Spots Enforcement*. Washington, DC: Police Executive Research Forum.
President's Commission on Law Enforcement and Administration of Justice. 1967. *Task Force Report: The Police*. Washington, DC: US Government Printing Office.
Ratcliffe, Jerry H., Travis Taniguchi, Elizabeth R. Groff, and Jennifer D. Wood. 2011. "The Philadelphia Foot Patrol Experiment: A Randomized Controlled Trial of Police Patrol Effectiveness in Violent Crime Hotspots." *Criminology* 49:795–831.
Reaves, Brian. 2010. *Local Police Departments, 2007*. Washington, DC: Bureau of Justice Statistics, Office of Justice Programs, US Department of Justice.
Reiss, Albert J., Jr. 1971. *The Police and the Public*. New Haven, CT: Yale University Press.
———. 1988. "Co-offending and Criminal Careers." In *Crime and Justice: A Review of Research*, vol. 10, edited by Michael Tonry and Norval Morris. Chicago: University of Chicago Press.
———. 1992. "Police Organization in the Twentieth Century." In *Modern Policing*, edited by Michael Tonry and Norval Morris. Vol. 15 of *Crime and Justice: A Review of Research*, edited by Michael Tonry. Chicago: University of Chicago Press.
Reiss, Albert J., Jr., and David P. Farrington. 1991. "Advancing Knowledge about Co-offending: Results from a Prospective Longitudinal Survey of London Males." *Journal of Criminal Law and Criminology* 82:360-95.
Robin, Gerald D. 1963. "Justifiable Homicide by Police Officers." *Journal of Criminal Law, Criminology and Police Science* 54:225–31.
Roman, John, Shannon E. Reid, Aaron J. Chalfin, and Carly R. Knight. 2009.

"The DNA Field Experiment: A Randomized Trial of the Cost-Effectiveness of Using DNA to Solve Property Crimes." *Journal of Experimental Criminology* 5:345–69.

Rossi, Peter H., Richard E. Berk, and Bettye Eidson. 1974. *The Roots of Urban Discontent*. New York: Wiley.

Rossi, Peter H., Emily Waite, Christine E. Bose, and Richard E. Berk. 1974. "The Seriousness of Crimes: Normative Structure and Individual Differences." *American Sociological Review* 39:224–37.

Sackett, David L., William M. Rosenberg, J. A. Gray, R. Brian Haynes, and W. Scott Richardson. 1996. "Evidence Based Medicine: What It Is and What It Isn't." *British Medical Journal* 312:71–72.

Sagan, Leonard. 1987. *The Health of Nations*. New York: Basic Books.

Sampson, Robert J., and Dawn Jeglum Bartusch. 1998. "Legal Cynicism and Subcultural Tolerance of Deviance: The Neighborhood Context of Racial Difference." *Law and Society Review* 32:777–804.

Saunders, Charles. 1970. *Upgrading the American Police: Education and Training for Better Law Enforcement.* Washington, DC: Brookings Institution.

Shaw, James W. 1995. "Community Policing against Guns: Public Opinion of the Kansas City Gun Experiment." *Justice Quarterly* 12:695–710.

Skogan, Wesley. 1990. *Disorder and Decline: Crime and the Spiral of Decay in American Neighborhoods.* New York: Free Press.

Skogan, Wesley, and Kathleen Frdyl, eds. 2004. *Fairness and Effectiveness in Policing: The Evidence.* National Research Council, Committee to Review Research on Police Policy and Practice. Washington, DC: National Academies Press.

Skogan, Wesley, and Susan Hartnett. 1997. *Community Policing, Chicago Style*. New York: Oxford University Press.

Sherman, Lawrence W. 1976. *The Control of Police Misconduct in Louisville, Kentucky.* Louisville: Office of the Public Safety Commissioner.

———. 1978. *Scandal and Reform: Controlling Police Corruption*. Berkeley: University of California Press.

———. 1980. "Execution without Trial: Police Homicide and the Constitution." *Vanderbilt Law Review* 33:71–100.

———. 1983. "Reducing Police Gun Use: Critical Events, Administrative Policy, and Organizational Change." In *Control in the Police Organization*, edited by Maurice Punch. Cambridge, MA: MIT Press.

———. 1984. "Experiments in Police Discretion: Scientific Boon or Dangerous Knowledge?" *Law and Contemporary Problems* 47:61–81.

———. 1986. "Policing Communities: What Works?" In *Communities and Crime*, edited by Albert J. Reiss Jr. and Michael Tonry. Vol. 8 of *Crime and Justice: A Review of Research*, edited by Michael Tonry and Norval Morris. Chicago: University of Chicago Press.

———. 1987. *Repeat Calls to Police in Minneapolis.* Crime Control Reports 4. Washington, DC: Crime Control Institute.

———. 1990. "Police Crackdowns: Initial and Residual Deterrence." In *Crime*

and Justice: A Review of Research, vol. 12, edited by Michael Tonry and Norval Morris. Chicago: University of Chicago Press.

———. 1992*a*. "Attacking Crime: Police and Crime Control." In *Modern Policing*, edited by Michael Tonry and Norval Morris. Vol. 15 of *Crime and Justice: A Review of Research*, edited by Michael Tonry. Chicago: University of Chicago Press.

———. 1992*b*. *Policing Domestic Violence: Experiments and Dilemmas*. With Janell D. Schmidt and Dennis P. Rogan. New York: Free Press.

———. 1993. "Defiance, Deterrence, and Irrelevance: A Theory of the Criminal Sanction." *Journal of Research in Crime and Delinquency* 30:445–73.

———. 1998. *Evidence-Based Policing.* Ideas in American Policing Series. Washington, DC: Police Foundation. http://www.policefoundation.org.

———. 2002. "Trust and Confidence in Criminal Justice." *National Institute of Justice Journal* 248:23–31.

———. 2007. "The Power Few Hypothesis: Experimental Criminology and the Reduction of Harm." *Journal of Experimental Criminology* 3:299–321.

———. 2009. "Evidence and Liberty: The Promise of Experimental Criminology." *Criminology and Criminal Justice* 9(1):5–28.

———. 2010. "Less Prison, More Police, Less Crime: How Criminology Can Save the States from Bankruptcy." Lecture presented to the National Institute of Justice, April 21. http://www.nij.gov/multimedia/presenter/presenter-sherman/data/resources/presenter-sherman-transcript.htm.

———. 2011*a*. "Al Capone, the Sword of Damocles, and the Police-Corrections Budget Ratio." *Criminology and Public Policy* 10:195–206.

———. 2011*b*. "Professional Policing and Liberal Democracy." Benjamin Franklin Medal Lecture delivered to the Royal Society for the Encouragement of Arts, Manufactures, and Commerce, London, November 1. http://www.thersa.org/events/audio-and-past-events/2011/professional-policing-and-liberal-democracy.

Sherman, Lawrence W., and Richard A. Berk. 1984. "The Specific Deterrent Effects of Arrest for Domestic Assault." *American Sociological Review* 49(2): 261–71.

Sherman, Lawrence W., Ioana Cosma, and Peter Neyroud. 2012. "Forecasting Serious Crime among Convicted Offenders in England and Wales: A Random Forests Approach." Paper presented to the American Society of Criminology, Chicago, November.

Sherman, Lawrence W., Patrick R. Gartin, and Michael E. Buerger. 1989. "Hot Spots of Predatory Crime: Routine Activities and the Criminology of Place." *Criminology* 27:27–55.

Sherman, Lawrence W., Denise Gottfredson, Doris MacKenzie, John E. Eck, Peter Reuter, and Shawn D. Bushway. 1998. *Preventing Crime: What Works, What Doesn't, What's Promising.* Research in Brief. Washington, DC: National Institute of Justice.

Sherman, Lawrence W., and Heather Harris. 2012. "Life-Course Effects of Arrest for Domestic Violence: Informal and Formal Control over 24 Years."

Paper presented to the American Society of Criminology, Chicago, November.
Sherman, Lawrence W., and Robert H. Langworthy. 1979. "Measuring Homicide by Police Officers." *Journal of Criminal Law and Criminology* 70:546–60.
Sherman, Lawrence W., and the National Advisory Commission on Higher Education for Police Officers. 1978. *The Quality of Police Education.* San Francisco: Jossey-Bass.
Sherman, Lawrence W., and Heather Strang. 2012. "Restorative Justice as Evidence-Based Sentencing." In *The Oxford Handbook of Sentencing and Corrections*, edited by Joan Petersilia and Kevin Reitz. New York: Oxford University Press.
Sherman, Lawrence W., Heather Strang, Geoffrey C. Barnes, and Daniel J. Woods. 2006. "Race and Restorative Justice." Paper presented to the American Society of Criminology, Los Angeles, November.
Sherman, Lawrence W., and David Weisburd. 1995. "General Deterrent Effects of Police Patrol in Crime Hot Spots: A Randomized, Controlled Trial." *Justice Quarterly* 12:635–48.
Silver, Nate. 2012. *The Signal and the Noise: Why So Many Predictions Fail—but Some Don't.* New York: Penguin.
Sparrow, M. 2011. *Governing Science*. Cambridge, MA: Harvard University. http://www.nij.gov/pubs-sum/232179.htm.
Spelman, William. 1990. *Repeat Offender Programs for Law Enforcement.* Washington, DC: Police Executive Research Forum.
Spelman, William, and Dale K. Brown. 1981. *Calling the Police: Citizen Reporting of Serious Crime.* Washington, DC: Police Executive Research Forum.
Spelman, William, and John E. Eck. 1989. *Sitting Ducks, Ravenous Wolves and Helping Hands: New Approaches to Urban Policing*. Austin, TX: Lyndon B. Johnson School of Public Affairs, University of Texas.
Stanko, Betsy. 2013. Paper presented to the Society of Evidence-Based Policing, University of Edinburgh, February 12.
Strang, Heather, Lawrence Sherman, Evan Mayo-Wilson, Daniel J. Woods, and Barak Ariel. 2013. "Restorative Justice Conferencing (RJC) Using Face-to-Face Meetings of Offenders and Victims: Effects on Offender Recidivism and Victim Satisfaction." Campbell Systematic Reviews, forthcoming.
Tankebe, Justice. 2009. "Public Cooperation with the Police in Ghana: Does Procedural Fairness Matter?" *Criminology* 47:1265–93.
Tapscott, Don, and Anthony D. Williams. 2006. *Wikinomics: How Mass Collaboration Changes Everything.* London: Portfolio.
Telep, Cody W., Renée J. Mitchell, and David Weisburd. 2012. "How Much Time Should the Police Spend at Crime Hot Spots? Answers from a Police Agency Directed Randomized Field Trial in Sacramento, California." *Justice Quarterly* 2012:1–29.
Tennenbaum, Abraham N. 1994. "The Influence of the 'Garner' Decision on Police Use of Deadly Force." *Journal of Criminal Law and Criminology* 85(1): 241–60.

Thompson, James D. 1967. *Organizations in Action*. New York: McGraw-Hill.
Thornton, Sara D. 2011. M.St. thesis, University of Cambridge, Institute of Criminology.
Tittle, Charles R., and Allan R. Rowe. 1974. "Certainty of Arrest and Crime Rates: A Further Test of the Deterrence Hypothesis." *Social Forces* 52:455–62.
Trojanowicz, Robert C., and Robert Baldwin. 1982. *An Evaluation of the Neighborhood Foot Patrol Program in Flint, Michigan*. East Lansing: Michigan State University.
Weisburd, David. 2012. "Can Automobile Vehicle Locator Systems Be Used to Manage Police Presence and Reduce Crime? The Dallas, TX AVL Experiment." Paper Presented to the Fifth Cambridge University International Conference on Evidence-Based Policing, July 10. http://sms.cam.ac.uk/media/1277185.
Weisburd, David, Elizabeth R. Groff, and Sue-Ming Yang. 2012. *The Criminology of Place: Street Segments and Our Understanding of the Crime Problem*. Oxford: Oxford University Press.
Weisburd, David, and Cynthia Lum. 2005. "The Diffusion of Computerized Crime Mapping in Policing: Linking Research and Practice." *Police Practice and Research* 6:419–34.
Weisburd, David, Stephen D. Mastrofski, Rosann Greenspan, and James J. Willis. 2004. *The Growth of Comp-Stat in American Policing*. Police Foundation reports. Washington, DC: Police Foundation.
Weisburd, David, and Peter Neyroud. 2011. *Police Science: Toward a New Paradigm*. Cambridge, MA: Harvard University. https://www.ncjrs.gov/pdffiles1/nij/228922.pdf.
Weisburd, David, Cody W. Telep, Joshua C. Hinkle, and John E. Eck. 2010. "Is Problem-Oriented Policing Effective in Reducing Crime and Disorder?" *Criminology and Public Policy* 9:139–72.
Weisburd, David, Laura A. Wyckoff, Justin Ready, John E. Eck, Joshua C. Hinkle, and Frank Gajewski. 2006. "Does Crime Just Move around the Corner? A Controlled Study of Spatial Displacement and Diffusion of Crime Control Benefits." *Criminology* 44:549–92.
White, Susan O., and Samuel Krislov. 1977. *Understanding Crime: An Evaluation of the National Institute of Law Enforcement and Criminal Justice*. Washington, DC: National Academy of Sciences.
Willis, J. J., Stephen D. Mastrofski, and David Weisburd. 2007. "Making Sense of COMPSTAT: A Theory-Based Analysis of Organizational Change in Three Police Departments." *Law and Society Review* 41:147–88.
Wilson, James Q. 1968. *Varieties of Police Behavior*. Cambridge, MA: Harvard University Press.
Wilson, Orlando W. 1950. *Police Administration*. New York: McGraw-Hill.
Zimring, Franklin E. 1976. "Field Experiments in General Deterrence: Preferring the Tortoise to the Hare." *Evaluation* 3:132–35.
———. 2012. *The City That Became Safe: New York's Lessons for Urban Crime and Its Control*. New York: Oxford University Press.