This is the authors’ final peer reviewed (post print) version of the item published as:


Available from Deakin Research Online:

http://hdl.handle.net/10536/DRO/DU:30064410

Reproduced with the kind permission of the copyright owner.

Copyright : 2012, Emerald publishing
USING THEORY IN CRIMINAL JUSTICE EVALUATION

Brad Astbury

ABSTRACT

This chapter examines the nature and role of theory in criminal justice evaluation. A distinction between theories of and theories for evaluation is offered to clarify what is meant by ‘theory’ in the context of contemporary evaluation practice. Theories of evaluation provide a set of prescriptions and principles that can be used to guide the design, conduct and use of evaluation. Theories for evaluation include programme theory and the application of social science theory to understand how and why criminal justice interventions work to generate desired outcomes. The fundamental features of these three types of theory are discussed in detail, with a particular focus on demonstrating their combined value and utility for informing and improving the practice of criminal justice evaluation.

Criminal justice evaluators are confronted with a myriad of choices: What is the purpose of the evaluation? What are the key evaluation questions? What is the most appropriate design for the study? What methods should be employed? What role should I play in this evaluation? Which stakeholders should be involved? What should be done to ensure that the results are used? What about the legal, political and ethical challenges that may arise? The way in which evaluators respond to these kinds of questions seems to depend largely on particular features of the evaluation context, the information needs of decision-makers, as well as the tacit knowledge and experience of the evaluator.

There appears, though, to be little discussion regarding roles for ‘theory’ in criminal justice evaluation practice. There has been some debate among academic criminologists in Australia and the United Kingdom about political and methodological dimensions of evaluation work and some of this does touch on issues related to theory (see Farrington, 2003; Hope, 2004; Israel, 2000; Walters, 2003; Weatherburn, 2005). However, explicit
consideration of the nature and functions of theory in criminal justice
evaluation is rare. This could be one of the reasons why some
criminologists claim that evaluation research is unreliable, ‘lacks academic
rigour’ and has been corrupted by commercialism and political influence
(Travers, 2005, p. 39; White, 2001). A renewed emphasis on theory may
go some way towards promoting a more positive image of criminal justice
evaluation.

A major aim of this chapter is to unpack the relationship between theory
and evaluation practice, and in doing so demonstrate how and why theory
matters for practice. I begin by outlining briefly some of the reasons why
evaluation research appears to have adopted a primarily method-driven, a
theoretical orientation. Next I attempt to clarify what is meant by ‘theory’
in the context of contemporary evaluation practice by distinguishing
between theories of and theories for evaluation. Theories of evaluation
fall into the domain of what is commonly referred to as evaluation theory.
Theories for evaluation include programme theory and the application of
substantive social science theory to understand how and why criminal
justice programmes work to generate desired outcomes. The central
features of these various kinds of ‘practising theory’ are discussed in turn,
with particular emphasis on demonstrating the role each type of theory
can play in guiding the conduct of criminal justice evaluation.

HISTORICAL NEGLECT OF THEORY IN EVALUATION

The traditional neglect of theory in evaluation practice can be viewed, in
the main, as a by-product of the historical development of programme
evaluation (Chen, 1990). Compared to many other areas of social science,
programme evaluation is still a relatively young discipline. Although
systematic evaluation, particularly in the field of education and curriculum
assessment, had already been advanced to some extent in the 1930s and
1940s by scholars such as Ralph Tyler it was not until after World War II
that evaluation experienced a significant boom. Under the Kennedy and
Johnson Administrations of the late 1960s and 1970s social scientists from
a variety of disciplinary backgrounds were called upon to conduct
programme and policy evaluations of large-scale federal reforms.
Researchers studiously applied their analytical and technical skills to the ‘rational’ assessment of hundreds of new interventions designed to alleviate poverty and improve the human condition.

The general mood at this time was, in hindsight, overly optimistic and perhaps even a little naïve. It was widely believed that through solemn application of sound social science methodology, policy-makers would be able to identify and expand effective social programmes, while at the same time terminating and discarding programmes that did not work. However, distinguishing between programmes that failed because they were a bad idea and those that failed because of poor implementation proved to be much more difficult than anticipated. When clear results were available, many evaluators assumed that politicians and stakeholders, who had funded their work, would eagerly apply findings and recommendations. Experience soon revealed that direct, instrumental use of evaluation findings is frequently an exception rather than the norm.

Since that time, much has been learnt. While programme evaluation is still primarily a practice-driven field, it is also fair to say that it is becoming more sophisticated and theoretically mature (Shadish, 1998). Evaluation is starting to ‘crystallize and emerge as a distinct profession’ with its own unique and rich language, theories and logic of practice (Stufflebeam & Shinkfield, 2007, p. 41). Over the past two decades the number of articles and books about evaluation has grown dramatically, as has the number of professional evaluation societies. Significant developments in evaluation theory, methodology and practice continue to occur as the field addresses new challenges and becomes increasingly global in orientation.

However, advances in the theoretical knowledge base of programme evaluation appear to have had a negligible impact on how evaluation is practised in the field (Christie, 2003; Shadish & Epstein, 1987; Smith, 1993). This is particularly so in many of the disciplinary areas where evaluations are regularly conducted. For example, in medicine the term ‘evaluation’ is often still synonymous with experimental impact studies. In psychology, education, health promotion and criminology, where a stronger interdisciplinary focus exists, a broader conception of evaluation is generally encouraged. Nonetheless, the extent to which knowledge of evaluation theory is communicated within and across disciplines appears to be limited. Although the reasons are likely to be varied, it is hard to argue that this low level of familiarity with theory is desirable, if for no
other reason than the fact that evaluators are deprived of knowledge of options they have available to use in their practice’ (Shadish & Epstein, 1987, p. 586).

THE MEANING AND USES OF THEORY IN EVALUATION

There is no universally accepted definition of theory in the natural or social sciences. Depending on the particular field of knowledge and context in which it is used, the term can mean different things. There are important consequences of this for the field of evaluation. The American sociologist Robert Merton, who devoted much of his career to advancing the role of theory in social science, describes the problem as follows:

Like so many words that are bandied about, the word ‘theory’ threatens to become meaningless. Because its referents are so diverse – including everything from minor working hypothesis, through comprehensive but vague and unordered speculations, to axiomatic systems of thought – use of the word often obscures rather than creates understanding. (1967, p. 39)

In light of these cautions it is important from the outset to be clear about what is meant by the rather abstract notion of ‘theory’ and to identify its various applications in evaluation. While it has been recognised for some time that evaluators make use of various kinds of theory in their work, Donaldson and Lipsey (2006) were perhaps the first to clarify some of the different meanings of theory in the context of contemporary evaluation practice. They make a useful distinction between three common forms of theory encountered in evaluation:

1. Evaluation theory
2. Programme theory
3. Social science theory

The first form of theory is best understood as a theory of evaluation because it offers a set of prescriptions about what constitutes ‘good’ evaluation practice. Programme theory and social science theory can be viewed as theories for evaluation in that they provide a methodology for understanding how and why programmes bring about change. Of course, the boundaries between these three forms of theory are blurry. Nevertheless, the distinction is helpful because it provides some
conceptual clarity regarding the nature and role of theory in evaluation practice.

THEORIES OF EVALUATION

Theories of evaluation attempt to provide a coherent set of principles to guide the study and practice of programme evaluation (Alkin, 2004). According to Shadish, Cook, and Leviton (1991), there are five main components of evaluation theory. These can be summarised briefly as follows:

1. Social programming: the nature, internal structure and functioning of social programmes particularly with respect to their role in ameliorating social problems.

2. Knowledge construction: philosophical assumptions about the nature of knowledge (epistemology), the nature of reality (ontology) and methods used to generate ‘credible’ knowledge about social programmes.

3. Valuing: the nature and role that various kinds of values play in the evaluation of social programmes.

4. Knowledge use: the ways social science information is used to modify programmes and policies and what evaluators can do to increase use.

5. Evaluation practice: the role of evaluators, the relationship they form with programme stakeholders and the tactics and strategies evaluators use in their professional work; especially given the constraints of time, money and expertise.

There is no single or ideal theory of evaluation. Evaluation theorists differ considerably in their conceptualisation of the five components of evaluation theory identified above. This has led to a vast (and now almost overwhelming) array of programme evaluation theories that are known more widely as ‘models’, ‘approaches’, ‘types’ or ‘forms’ of evaluation. Newcomers to the field of evaluation in the 1970s could perhaps be content with understanding the basic distinction between formative and summative evaluation approaches (see Scriven, 1967 and below for more detail). Today the novice evaluator faces a much more difficult prospect of
sorting through dozens of competing models of evaluation. One recent attempt at classification, for example, identified 26 alternative evaluation models (Stufflebeam & Shinkfield, 2007).

**Table 1** provides a select overview of several well-known and some not so well-known evaluation theories and theorists. The list is certainly not exhaustive, but is intended to demonstrate the depth and variety of theoretical influences on contemporary evaluation practice.

The diversity of evaluation theory is partly due to the influence that different philosophical orientations have had on the nature and development of intellectual thinking in evaluation. During the mid to late 1980s, evaluation, like many disciplines, experienced a heated debate between two prevailing perspectives or paradigms. These two paradigms have been labelled in various ways but most typically are referred to as ‘positivism’ and ‘constructivism’. The ‘paradigm wars’ have been fought vigorously across many key areas in the philosophy of social science, such as: the nature of reality (ontology); the nature of knowledge (epistemology); the role of values in inquiry (axiology); the possibility of causal linkages; and the ability to establish generalisations (Guba & Lincoln, 1981, 1989).

Traditionally, experimental and quasi-experimental theories of evaluation have dominated mainstream evaluation practice, providing a methodological template for how to do evaluation (e.g. Campbell & Stanley, 1966; Cook & Campbell, 1979). Experimentalism has historical roots in a positivist philosophy of social science. Briefly stated, this position maintains that evaluators should approach the study of social programmes in roughly the same way that the physical sciences such as physics and chemistry approach the study of natural phenomena (Martin & McIntyre, 1994).

Experimental evaluation follows a standard logic – construct two equivalent groups and undertake a ‘before’ measurement, then apply the intervention to only one of the two groups and then take an ‘after’ measurement. Any difference or change from first to second measurement in the intervention (experimental/treatment) group compared to the non-intervention (control) group is deemed to be attributable to the intervention. When cases are randomly assigned to each of the two groups, this research design is referred to as a randomised
control trial (RCT) or ‘true experiment’. When non-random methods are used to create experimental and control groups, or where a control group is absent the design is referred to as a ‘quasi-experiment’.

Over the past three decades the orthodoxy of the experimental approach has been challenged by the development of anti-naturalist approaches to evaluation, such as: responsive evaluation (Stake, 1983); fourth-generation evaluation (Guba & Lincoln, 1989); and empowerment evaluation (Fetterman, 1996).
<table>
<thead>
<tr>
<th>Type</th>
<th>Basic Description</th>
<th>Key Theorist(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Objectives-based evaluation</td>
<td>The objectives-oriented approach to evaluation was first formulated in the 1940s, but is still popular today. In essence, it involves operationalising programme goals into measurable objectives that then serve as a benchmark for determining success.</td>
<td>Ralph Tyler, William Popham</td>
</tr>
<tr>
<td>Experimental evaluation</td>
<td>An evaluation that involves random assignment or the use of matching procedures to allocate participants to treatment and control groups. Quantitative estimates of programme effect are achieved by comparing outcome between groups.</td>
<td>Donald Campbell, Tom Cook, William Shadish</td>
</tr>
<tr>
<td>CIPP model</td>
<td>The essential features are captured in the acronym CIPP – which stands for evaluations that focus on context, inputs, processes and products. Many evaluation taxonomies are based on the four types of evaluation identified in this model.</td>
<td>Daniel Stufflebeam</td>
</tr>
<tr>
<td>Goal-free evaluation</td>
<td>This approach is based on the premise that evaluators should be concerned with what a programme is actually doing rather than official statements of what was intended. The goal-free evaluator deliberately avoids programme goals and attempts to uncover all effects, including unintended positive or negative outcomes.</td>
<td>Michael Scriven</td>
</tr>
<tr>
<td>Responsive evaluation</td>
<td>Emphasis is given to holistic observation of programmes in order to support local improvement efforts. The responsive evaluator prefers qualitative case study methods that allow questions and issues to emerge during the evaluation, focuses on programme activities rather than goals, incorporates diverse values and responds to audience information needs.</td>
<td>Robert Stake</td>
</tr>
<tr>
<td>Meta-evaluation</td>
<td>Meta-evaluation is a systematic process for evaluating an evaluation, typically against standards such as utility, feasibility, propriety and accuracy. Can be used to improve an ongoing evaluation or inform judgements about the strengths, limitations and overall merit of a completed evaluation.</td>
<td>Michael Scriven, Daniel Stufflebeam</td>
</tr>
<tr>
<td>Utilisation-focused evaluation</td>
<td>Epitomised by the maxim 'intended use by intended users' An eclectic and collaborative approach to evaluation that involves identifying stakeholders who have a special interest in an evaluation, obtaining their buy-in, and then orienting the study so that it meets their information needs.</td>
<td>Michael Patton</td>
</tr>
<tr>
<td>Adversary evaluation</td>
<td>A concept of evaluation that borrows from the judicial process, where a programme is effectively 'put on trial'. The adversarial or judicial evaluation model seeks to provide a comprehensive, impartial and balanced overview of programme strengths and weaknesses through the presentation of evidence from opposing perspectives. Briefly</td>
<td>Thomas Owens, Robert Wolf</td>
</tr>
<tr>
<td>Type</td>
<td>Basic Description</td>
<td>Key Theorist(s)</td>
</tr>
<tr>
<td>----------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Connoisseurship and criticism</td>
<td>This perspective on evaluation is influenced by art criticism and aesthetic theory. It is concerned fundamentally with the process of making value judgements. The role of the evaluator is somewhat analogous to a good wine connoisseur or art critic who draws on their professional expertise, experience and insights to provide an assessment of programme quality and worth.</td>
<td>Elliot Eisner</td>
</tr>
<tr>
<td>Fourth-generation evaluation</td>
<td>Based on constructivist philosophy, fourth-generation evaluation grew out of skepticism towards traditional objectives-based and experimental evaluation approaches. The basic process involves identifying stakeholders, eliciting stakeholders' constructions of the programme and related claims, concerns, and issues, and then facilitating consensus among stakeholders via discussion, negotiation and exchange.</td>
<td>Egon Gube Yvonna Lincoln</td>
</tr>
<tr>
<td>Empowerment evaluation</td>
<td>A highly participatory model that uses evaluation methods, concepts, techniques and findings to foster improvement and self-determination. The origins of this approach are based in community psychology and action research. Empowerment evaluators perform the role of a 'coach', providing technical assistance and support to build the capacity of participants to improve their programmes through critical reflection and self-evaluation.</td>
<td>David Fetterman</td>
</tr>
<tr>
<td>Deliberative democratic evaluation</td>
<td>An approach to evaluation that draws on the democratic principles of inclusion, dialogue and deliberation to reduce bias and arrive at defensible evaluative conclusions. It involves the use of traditional evaluation methods, but emphasises the need to consider all relevant interests, values and perspectives, engage in stakeholder dialogue and promote reflective deliberation.</td>
<td>Ernest House Kenneth Howe</td>
</tr>
<tr>
<td>Developmental evaluation</td>
<td>Patton's most recent contribution to evaluation theory involves the application of systems thinking and complexity theory to the field of evaluation. Developmental evaluators are an integral part of the innovation design team. They assist organisations to develop programmes in complex, dynamic environments.</td>
<td>Michael Patton</td>
</tr>
</tbody>
</table>

Source: Adapted from Mathison (2005).
These are based, to varying degrees, on constructivist philosophies of social science. This position holds the view that evaluators should not investigate social programmes in the same way as natural scientists study nature. The basic argument is as follows: people construct their own subjective reality of social programmes through complex processes of human interaction. As a consequence, knowledge about social programmes cannot exist external to human interactions which occur in specific contexts, spaces and times.

It is difficult to summarise the design ‘logic’ of constructivist evaluation, as there are many different approaches that fit broadly into the constructive camp. At a fundamental level though, there is usually a tendency to employ qualitative approaches such as case studies, in-depth interviewing and prolonged periods of field observation. This is deemed necessary to understand the complexity and richness of human interactions that occur between various programme stakeholders. Constructivist evaluators also tend to adopt a ‘negotiator’ role, work closely with stakeholders to facilitate consensus among multiple interpretations of the programme (see Guba & Lincoln, 1989).

**The Evolution of Evaluation Theory**

The emphasis placed on the five components of evaluation theory has changed over time. Early developments in evaluation theory focused primarily on knowledge construction and valuing (see the work of Donald Campbell and Michael Scriven). In the late 1970s and throughout the 1980s conceptual development turned to theories of use and strategies for responding to the needs of policy-makers (e.g. Carol Weiss, Joseph Wholey and Robert Stake). Evaluation theorists such as Peter Rossi and Lee Cronbach then tried to integrate the work of previous scholars into tailored, systematic evaluation approaches (Shadish et al., 1991).

Over the past decade there have been some notable attempts to further consolidate different perspectives on evaluation. For example, Mark, Henry, and Julnes (2000) present a coherent and practical theory of evaluation, grounded in an emergent realist philosophy of science. They offer a useful distinction between four key purposes of programme evaluation. These are: (1) to assess the merit and worth of programmes; (2) to ensure programme compliance with regulatory mandates, (3) to
improve programmes and organisations; and (4) to build theoretical knowledge and cumulate evidence for future programmes. They also identify four modes of evaluative inquiry: description, classification, causal analysis and values inquiry. For practitioners, this integrative framework can guide decisions regarding the most appropriate evaluation approach and methods, given specific circumstances and constraints.

As the field of evaluation continues to grow there has also been a noticeable rise in the number evaluation ‘textbooks’, as well as new journals, encyclopaedias and handbooks targeted specifically at professional evaluators. In 1972 Carol Weiss published one of the very first evaluation textbooks, and in 1976 the first journal, Evaluation Review, was launched by Sage Publications. Today there are over a dozen dedicated evaluation journals that publish scholarly articles reporting on evaluation studies, evaluation methods, theory, use, values and philosophy.

Another indication of growth is the worldwide expansion of professional evaluation associations. In 1980 there were only three regional or national evaluation societies. This grew to nine by the late 1990s and just one decade later there were more than 50 professional evaluation societies. Today, there are over 125 societies and a global network, the International Organisation for Cooperation in Evaluation, has been established to build evaluation capacity and foster the exchange of ideas, practices and insights (http://ioce.net/index.shtml).

These trends and developments suggest that evaluation is becoming much more commonplace in modern democracies as well as in developing countries. We are now experiencing what might be called a ‘Second Boom’ period in evaluation (Donaldson & Lipsey, 2006). The idealism and heated debates associated with philosophical prescriptions of the past have been tempered by a growing acceptance of the complementary roles that different forms of data can play in evaluation research. A planned eclecticism of approach holds sway. Today’s evaluation practitioner typically adopts a pragmatic stance; mixing-methods and tailoring the evaluation to meet the needs of intended users.
Roles for Evaluation Theory in Criminal Justice

Having outlined, very briefly, the nature and evolution of evaluation theory one might ask, why is evaluation theory worth knowing? One way of answering this is to consider the relationship between evaluation theory and methodology. No single research design, methodology or method of data collection is appropriate always and everywhere. Strict adherence to one approach is likely to result in data that is of little use. It may also lead to inaccurate conclusions about the merit or worth of the programme or policy that is being evaluated. The decline of the rehabilitative ideal following Martinson’s (1974) now infamous review of 231 quasiexperimental evaluations of offender treatment studies is an instructive example of how the methodological ‘law of the hammer’ approach to evaluation can dangerously misinform judgements regarding programme effectiveness. Shadish et al. (1991) explain that knowledge of evaluation theory encourages appreciation of the range of methods and techniques that can be employed by practising evaluators. Most importantly though, it provides guiding principles for their application across the range of scenarios that criminal justice evaluators are confronted with:

Evaluation theories are like military strategies and tactics; methods are like military weapons and logistics. The good commander needs to know strategy and tactics to deploy weapons properly or to organise logistics in different situations. The good evaluator needs theories for the same reasons in choosing and deploying methods. Without thorough grounding in evaluation theory, the evaluator is left to trial and error or to professional lore in learning about appropriate methods. (p. 34)

In a similar vein, Stufflebeam and Shinkfield (2007) remind us that evaluation theory is worth knowing because poor theories of evaluation can mislead evaluation practice, resulting in serious negative consequences. They use as an example, the development of objectives-based evaluation theory. This approach was developed by Ralph Tyler in the 1930s, and grew out of his work on the effects of large-scale educational innovation. Objective- or goal-based evaluation narrowly defines the purpose of evaluation as the process of determining whether or not a programme has achieved specified outcomes. This ‘theory’ of evaluation misled evaluators for decades as there was a strong tendency to focus only on intended outcomes, not on the actual outcomes (positive or negative) that a particular programme might be generating.
Another informative case-in-point comes from the area of evaluation commonly referred to as performance measurement or monitoring (Whooley, 2001). Performance data typically comes in the form of quantitative output measures such as: number of offenders enrolled in a programme; percentage of programme completions; crime ‘clearance’ rates, recidivism rates and so on. Criminal justice administrators regularly develop and use ‘indicators’ such as these to track implementation and measure programme impact over time. The application of a measurement or indicator-based approach to evaluation can, however, alter the delivery of an intervention in destructive ways. In the evaluation theory literature this phenomena is commonly referred to as ‘Campbell’s Law’ of goal displacement:

The more any quantitative indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor. (Campbell, 1979, p. 85)

Consider, for example, the commonly used ‘number and rate of completion’ measure for correctional programmes. In many jurisdictions new initiatives have been introduced far too rapidly, resulting in perverse effects. In one study, Raynor (2004) found that the three-year time frame for implementing offender programmes in probation agencies across England and Wales was not realistic. Pressure from government and Treasury officials to introduce change quickly occurred mainly as a result of unreasonable targets for programme completions. These targets ‘drove the pace of the roll-out of offending programs’ but were established ‘without any systematic prior assessment of the characteristics of offenders under supervision and their suitability for programmes’ (p. 316).

Aside from wasting time and resources, the presence of an unenthusiastic offender in a group programme could have a disruptive or harmful affect on other participants. Even the use of higher-order ‘outcome’ measures such as recidivism can undermine the intent of correctional programmes. Programme staff, for instance, may recruit lower-risk offenders to make a new programme appear more successful or to gain additional funding (a phenomena known as ‘creaming’). While it is difficult to prevent these kinds of scenarios from occurring altogether, prior knowledge of relevant
evaluation theory can assist evaluation practitioners to minimise the dysfunctional side-effects of quantiative performance measures.

The benefit of understanding evaluation theory can be demonstrated further through a final illustration. Scriven (1967) first distinguished between the formative and summative roles of evaluation. Formative evaluation asks the question ‘How are we doing?’ It focuses on generating information about programme processes and implementation. This information is then used primarily by internal staff to inform decisions about programme development and improvement. Summative evaluation asks the question ‘How have we done?’ It focuses on generating evidence about the worth of the programme in terms of outcomes and ultimate impact of the intervention. Summative evaluation is typically conducted after a programme has had time to settle and mature, and often informs decisions concerning programme continuation, termination, expansion, replication and so on.

The formative–summative typology, although a simplification of reality, is now almost universally accepted in the field and has been instrumental in shaping the way people think about evaluation. Evaluators who lack an understanding of the importance of tailoring an evaluation to the stage of programme development may fall into the trap of conducting a summative evaluation too soon. In the socio-political context of criminal justice programming this can be problematic. For example, where an experimental study conducted on a poorly implemented programme leads to findings of ‘no effect’, this may result in the premature termination of a potentially effective intervention.

To conclude, evaluation theory serves several purposes. Perhaps most critically, it functions as a guide to practice. It does this in many ways: by informing decisions about when and why to use particular models of evaluation and methods of data collection; by steering evaluators away from narrow or faulty conceptions of what evaluation is and how it should be conducted; and by directing the practising evaluator to common issues, problems and challenges that are likely to arise in particular evaluation contexts.
THEORIES FOR EVALUATION

This section introduces two further notions of theory, which can also be incorporated usefully into the practice of criminal justice evaluation. These two types of theory are referred to here as programme theory and social science theory. Both can be considered theories for evaluation because they provide the evaluator with a powerful way of understanding how and why social programmes generate outcomes. This is somewhat different from a theory of evaluation, which as we have seen, focuses on the body of principles, prescriptions and ideas about how to conduct evaluations.

Programme Theory

Many discussions of programme theory seem to imply that it is a new approach – something which has only recently been discovered by the evaluation community. In fact, the concept of ‘programme theory’ has a relatively long lineage, particularly when one considers the short history of the field of evaluation. The roots of theory-driven evaluation can be traced back to the work of Edward Suchman, who in his 1967 book Evaluative Research, was perhaps the first evaluator to argue for the importance of investigating a programme’s implicit theory of action (at the time he referred to it as a programme’s ‘chain of objectives’).

Importantly, Suchman suggested that one of the primary purposes of evaluation is ‘testing’ a programme hypothesis:

The evaluation study tests some hypothesis that activity A will attain objective B because it is able to influence process C which affects the occurrence of this objective. An understanding of all three factors – program, objective and intervening process – is essential to the conduct of evaluative research. (p. 177)

Perhaps because of Suchman’s untimely death shortly after the publication of his book, the importance of focusing evaluation on the theory underlying the action of social programmes was not recognised fully until some years later. Subsequently, evaluation theorists such as Carol Weiss (1972, 1997), Joseph Wholey (1979, 1987), and, in particular, Peter Rossi’s student Huey-Tsyh Chen (1990, 2005), have all helped to establish the prominence of programme theory.
There are some important differences in the various approaches to programme theory-based evaluation, but the common underlying theme is that programmes are theories. That is, all social programmes contain an assumption (or more likely a set of assumptions) about how and why programme resources and activities will bring about change in the reasoning and behaviour of participants. In this sense, programmes are not ‘dosages’ given to passive recipients; they are active interventions that only ‘work’ if people choose to make them work and are placed in the right conditions to enable them to do so’ (Pawson & Tilley, 1994, p. 294).

Although there is currently no accepted definition, programme theory based evaluation has been usefully described by Bickman (1987) as ‘the construction of a plausible and sensible model of how a program is supposed to work’ with whom, how, why and under what circumstances (p. 5). This definition depicts nicely the idea that programme theories are best thought of as ‘small theories’ associated with specific social programmes, policies and interventions (Lipsey, 1993).

Unfortunately, the term ‘programme theory’ all too often invokes images of some grand theory that attempts to account for the functioning of human behaviour in a law-like fashion. This image does not capture the kind of theorising that evaluators apply to the study of social programmes. Theory oriented evaluations are much more modest. This is one of the reasons why programme theories are often likened (after the sociologist Robert Merton) to middle-range theories which position themselves between universal social laws and description (Pawson, 2000).

Programme theory can be expressed in many ways – a graphic display of boxes and arrows, a table, a narrative description and so on. The methodology for constructing programme theory, as well as the level of detail and complexity, also varies significantly. Some examples include: path analysis and causal modelling; observations of the programme in action; interviews with staff to uncover implicit assumptions about how the programme works; concept mapping exercises; formal (argumentational) analyses of programme and policy documents; and detailed examination of research on similar programmes (Connell, Kubish, Schorr, & Weiss, 1995; Donaldson, 2007; Leeuw, 2003; Smith, 1990; Trochim, 1989). Regardless of the way it is developed and expressed, programme
theory should be grounded in the empirically realities of the particular programme, its implementation context, staffing arrangements, funding levels and social and political environment.

As well as helping to build knowledge about how and why further examples of the benefits of programme theory for evaluation. For example, one early advocate of programme theory, Joseph Wholey (1983), used a form of programme theory known as ‘evaluability assessment’ in his work with the US government to determine whether or not a programme was ready to be evaluated. A good evaluability assessment can save considerable resources by forestalling a premature and potentially costly impact evaluation that would almost certainly yield biased estimates of programme effect.

As interest in programme theory increased throughout the 1990s, evaluators also discovered the importance of programme theory for guiding evaluations and engaging stakeholders. For example, contributors to the *New Directions for Evaluation* compilation on programme theory (see Rogers, Hacsi, Petrosino, & Huebner, 2000) discuss the role of programme theory in formulating and prioritizing evaluation questions, directing data collection activities; and augmenting experimental evaluations and meta-analysis. Evaluators also report that the process of programme theory development can assist goal clarification, enhance stakeholder buy-in and provide information to improve programme design and delivery.

It is common, however, to discover at the beginning of an evaluation that the theory of change underlying the programme has never been articulated or tested systematically. The programme theory or ‘theories’ remain hidden, typically in the minds of policy architects and practitioners. Failure to surface assumptions about how programmes are ‘supposed to work’ can be problematic because if a programme is based on a faulty theory, then it will not bring about desired change, no matter how well it is implemented.

Consider, for example, delinquency prevention programmes such as ‘Scared Straight’. These programmes involve young people visiting prison to experience prison life firsthand. The purpose of these visits is usually
framed within the context of reducing crime and re-offending. A theory-driven evaluation might start by asking: ‘What is it about exposure to prison life and prisoners that is supposed to help reduce crime and re-offending? One of the (faulty) assumptions underlying ‘Scared Straight’ programmes is that criminal activity can be prevented by giving wayward teenagers a taste of what it would be like to be locked-up in prison. However, this hypothesis is not supported by evaluation findings which consistently demonstrate that programmes such as ‘Scared Straight’ appear to do more harm than good. For example, one comprehensive meta-analysis found that young people who complete these types of programmes have higher rates of re-offending than young people who do not go through the programme (Petrosino, Turpin-Petrosino, & Buehler, 2003). It seems that despite considerable popularity and intuitive appeal ‘attempts to scare teenagers into better behaviour [are] not a successful enterprise’ (McCord, 2003, p. 26).

A theory-driven evaluation would dig further in an attempt to explain negative outcome patterns. Why is it that ‘Scared Straight’ programmes produce criminogenic effects? A quick survey of the literature suggests that a number of alternative programme theories seem to be at play. Consider the following three examples that may account for the harmful effects of ‘Scared Straight’ programmes:

*Alternative programme theory #1:* Rather than ‘detering’ crime, ‘Scared Straight’ programmes may inadvertently promote deviance among impressionable youths who view prisoners as positive (rather than negative) role models (Finckenauer, 1982).

*Alternative programme theory #2:* Participants may react to attempts at intimidation by rising to the challenge and ‘proving’ to themselves, their peers and others that they are not scared of prison (Finckenauer, 1982).

*Alternative programme theory #3:* Delinquent-prone participants who may already feel alienated from society could form a belief that prison is a place of sanctuary where they can establish new friendships and derive a sense of belonging (Greater Egypt Planning & Development Commission, 1979).
Of course, further investigation would likely reveal several additional, competing ‘Scared Straight’ programme theories. These would need to be teased out in a similar fashion and empirically tested using a variety of methods of data collection. The important point to emphasise is that programme theory-driven evaluations re-shift the focus of evaluation away from the view that programmes are molar treatments. According to the programme theory perspective, all programmes are based on a set of propositions regarding what goes on in the ‘black box’ to transform inputs into outcomes. Thus, when conducting programme theory evaluation, the theory or theories underlying programmes (even well-established programmes with high levels of support) are an important focus for the evaluator’s scrutiny, not just observable programme effects.

Social Science Theory

A major task of the social sciences is to understand the nature, extent and causes of social problems. Of course, this is not the only task but it is arguably one of the most important. Theories of social problems are based on various kinds of observations, facts and events. When empirical data are packaged into a set of sensible, logical, and interrelated propositions, they form the basis for a verifiable and generalisable theory. Social science theories can be usefully defined as an attempt to explain ‘recurrent patterns or regularities in social life’ (Blaikie, 2009, p. 124).

Social science theory is relevant for evaluators because programmes are planned efforts to solve social problems. Programmes ‘work’ by attempting to disrupt social regularities, altering the conditions, behaviours or processes that give rise to recurrent problems like crime, poverty, unemployment, poor educational achievement, substance misuse and so on. Although the evidence base for some social phenomenon is sparse, existing knowledge on prior efforts to prevent or solve social problems often exists. Of course, the knowledge base may not be directly applicable to a specific programme or evaluation, but ‘operating within a vacuum when useful information exists, can be a very inefficient way to practice evaluation’ (Donaldson & Lipsey, 2006, p. 63).
In the context of criminal justice evaluation, there is an enormous variety of research and empirically based theories that programme planners and evaluations can usefully draw upon. For example, the conceptual basis for offending behaviour programmes is typically based on cognitive social learning theory (McGuire, 2006), and in the area of restorative justice and youth conferencing, Braithwaite’s (1989) theory of reintegrative shaming has been applied extensively. Tilley’s (1993) realistic evaluation of the way in which surveillance tactics, like closed-circuit television (CCTV), might influence (i.e. ‘block’) opportunities for criminal behaviour in and around car parks is a clear demonstration of how situational crime prevention theory can be instructive for evaluation.

Social science theory also plays a critical role when evaluators are involved in assessing the need for a programme. Needs assessment is a fundamental stage of criminal justice evaluation ‘because a program cannot be effective at ameliorating a social problem if there is no problem to begin with or if the program services do not actually relate to the problem’ (Rossi, Lipsey, & Freeman, 2004, p. 102). The Australian criminologist Don Weatherburn provides an insightful critique of car-theft policy in Western Australia that highlights the importance of utilising existing research and theory, especially during needs assessment and programme design.

Weatherburn (2004) explains that in the late 1980s, Western Australia was experiencing a significant rise in the number of officially recorded car thefts. In what was to become the first of two misguided policy responses, law enforcement officials decided to deal with the ‘problem’ by engaging in high speed pursuits of stolen cars. The consequences of this policy decision were disastrous (16 deaths during an 18-month period). The Labour Government then introduced severe penalties for repeat juvenile offenders to appease public reaction, which spuriously linked the problem of car theft to a lenient juvenile justice system. A subsequent evaluation revealed that the introduction of new punitive legislation had no effect on the rate of car theft in Western Australia. In a bid to win law and order votes politicians and policy-makers had misdiagnosed the nature of the problem. Weatherburn (2004) writes:

The crowing irony is that the car theft problem in Western Australia was eventually brought under control, not as a result of tougher penalties, but
because a subsequent government passed a law requiring engine immobilisers to be fitted to all vehicles being registered [an example of situational crime prevention theory at work]. (p. 29)

Another significant role for social science theory is to assist evaluators in the development of programme theory. As noted earlier, there are a variety of methods available for reconstructing programme theory. Many popular approaches utilise some form of visual ‘logic model’ to depict programme operations and outcomes. Typically the programme logic is based on information generated by some kind of elicitation technique, where the evaluator collaboratively develops the model with stakeholders (Leeuw, 2003). However, this does not mean that programme theories always need to be based solely on practitioner wisdom, as stakeholder theories could in fact be wrong (as some of the above examples demonstrate). Where possible, programme theory should also incorporate information from previous research and knowledge of mechanisms derived from social and behavioural science (Astbury & Leeuw, 2010).

From an evaluation perspective, social mechanisms are not programme activities or variables. Mechanisms are the underlying causal processes that generate change in participants, given certain conditions. Put another way, mechanisms focus on the way in which participants respond or react to programme services. A first test of the relevance and ‘validity’ of a programme or intervention theory can be undertaken by confronting the science with existing knowledge about mechanisms.

Although there are not yet repositories on mechanisms, several scholars have provided useful summaries of the research literature on mechanisms. Jon Elster (2007), for example, draws on insights from psychology to economics and political science to identify and discuss some twenty mechanisms that underlie a range of social phenomena. Similarly, Farnsworth (2007) takes legal arrangements like laws and contracts as a starting point and dissects which types of mechanisms play a role when one wants to understand how these arrangements work. He discusses mechanisms such as the ‘slippery slope’, the ‘endowment effect’ and ‘framing effects’. Theoreticians within the social sciences have also contributed to knowledge about mechanisms, as work by Festinger (1954), Merton (1968) and Olson (1971) has shown.
If an intervention is based on a mechanism that is known to not work well (in a given context), this is a signal that the programme will probably not be very effective. For example, Pawson (2006) demonstrates that many policies and programmes that appear on the surface to be very different actually rely on the same mechanism to change behaviour. School league tables, sex offender legislation, and car safety indices, for instance, can be conceptualised as a type of public disclosure intervention that taps into the same underlying mechanism (in this case, ‘naming and shaming’). Evidence suggests, however, that naming and shaming does not work particularly well in the first two contexts. Drawing on Merton’s celebrated reference group theory, Pawson (2010) shows that sex offender notification systems often lead to ‘naming and evading’, driving offenders underground and jeopardising public safety. Meanwhile, school league tables can result in ‘naming and gaming’, with administrators responding to ranking exercises by deploying tactics such as ‘teaching to the test’ in order to improve performance data. In contrast, vehicle companies seem to respond much more positively to naming and shaming efforts because they want to protect their position and reputation.

Lastly, a careful examination of social science theory and research literature can be beneficial for guiding decisions about design and method options, such as the selection of constructs and variables, measurement techniques and analysis procedures. Prior research can inform and provide context for interpretation of findings, such as standards and expectations regarding effect sizes for particular kinds of criminal justice interventions. Examination of relevant evaluation studies can also help to identify the types of conceptual and methodological strategies that have worked well, or not so well in the past. Of course, careful adaption for the programme under investigation is likely to be required, as concepts and techniques may not always generalise across different contexts. Nevertheless, the value of drawing on extant theory to anticipate likely pitfalls and leverage from accumulated knowledge should not be underestimated as it can save evaluators and funding agencies considerable time and resources (Donaldson & Lipsey, 2006).
In the 1960s, during the first boom-period in social programme evaluation, researchers from a variety of disciplinary backgrounds found themselves conducting evaluations. At this time there was virtually no theoretical foundation to guide practice. As we enter the second boom-period of evaluation, social scientists are again increasingly being called upon to undertake complex and often large-scale evaluations. For the many criminologists (and psychologists, sociologists and economists) now entering the field of evaluation, one of the major messages of this chapter is that evaluation researchers no longer need to rely so heavily on a ‘learning-by-doing’ or ‘trial-and-error’ approach.

Of course, good criminal justice evaluation is still good social science. Even so, history has taught us that while technical proficiency in the application of research methodology is important, an understanding of how theory can be incorporated into evaluation practice is also necessary. Drawing on the work of Donaldson and Lipsey (2006), this chapter has identified two broad types of theory and examined some of the ways in which these are useful for evaluation practice in criminal justice settings. First, theories of evaluation inform decisions regarding evaluation purpose, selection of data collection methods; identification of evaluator roles; techniques for engaging stakeholders and strategies for enhancing use.

There is no such thing as an ideal or perfect theory of evaluation, and practitioners should not adhere rigidly to one approach. Knowledge of evaluation theory tells us how, where, when and why to deploy methods, given the circumstances and constraints. Learning evaluation theory is a professional obligation because it can help evaluators to avoid inadvertent, but potentially harmful practices. For example, selecting an evaluation approach that does not match the stage of programme development or adopting a managerial bias by narrowly defining the success of a programme according to predetermined goals and objectives.

Second, theories for evaluation (i.e. programme theory and social science theory) help to build knowledge about social problems by identifying the causal mechanisms by which criminal justice programmes work (or fail to work). When used as an organising framework, programme theory can
also play a valuable role in several stages of an evaluation. During planning, programme theory can inform decisions about whether a programme is even ready to be evaluated. During evaluation design, it assists in determining appropriate questions and can focus the evaluation on critical components of the programme, saving time and resources.

When evaluators collaborate with staff, seeking their input into programme theory development, this can enhance stakeholder engagement in the evaluation process and facilitate use of findings. When combined carefully into evaluation practice, these three forms of theory provide a powerful set of concepts, methods and tools for guiding the conduct of criminal justice evaluation and generating new knowledge about how, why, where and for whom criminal justice interventions work to ameliorate social problems.
1. Christie (2003) suggests that this can be explained in part by the fact that many people who conduct evaluations have not received any formal training in evaluation and most are not evaluators; rather their primary identity is with some other field:

   today’s evaluators may not use any existing evaluation theory to guide their work…many people conducting evaluation rely on informal theory, derived from their experiences and ideas about evaluation…or use theories offered by their own professional disciplines rather than those offered by the evaluation community. (p. 2)

2. When applied as the dominant framework for organising the conduct of an evaluation, programme theory could be viewed as a distinct model or theory of evaluation. Other forms of theory in evaluation that are sometimes distinguished include implementation theory and domain theory; although these are arguably subsets of social science theory.

3. This is a reference to Kaplan’s (1964) ‘Law of the Instrument’ – ‘give a small boy a hammer, and he will find that everything he encounters needs pounding’ (p. 28). Kaplan is describing the similar tendency for researchers to employ their favoured methods and techniques irrespective of the nature and circumstances of the research task.

4. The term ‘social science’ is used here in a very general sense to refer to the collective wisdom of each of the various substantive disciplines (e.g. history, sociology, economics, psychology, criminology, politics, biology, etc.).
REFERENCES


