Realist Evaluation

Ray Pawson and Nick Tilley

2004[1]

[1] The preparation of this paper was funded by the British Cabinet Office.

See also Pawson, R. and Tilley, N., *Realistic Evaluation*, Sage, 1997

Realist Evaluation

Ray Pawson & Nick Tilley

Contents:

Introduction
1. The nature of programmes and how they work
2. Basic concepts in the explanation and understanding of programmes
3. Strategies and methods of realist evaluation
4. Realism's place in the policy cycle: formative, summative and synthetic approaches
5. The nature, presentation and use of findings from realist evaluation
6. Conclusion: strengths, limitations and relationships with other approaches

Appendix I – 'Thinking it through': an exercise in realist hypothesis making.
Appendix II – 'Varieties of realist evaluation': pocket illustrations of quantitative, qualitative, formative and synthetic applications.
Appendix III – 'Would it work here?': a grid to help decide on the feasibility of mounting a programme 'on your patch'.
References

Introduction

Realist evaluation is a species of theory-driven evaluation. Some of the differences between it and fellow members of the genus (programme theory evaluation, theories-of-change evaluation) will be noted in the course of the section. What should be stressed in the first instance, however, is the commonality. In all of these perspectives social programmes are regarded as products of the human imagination: they are hypothesis about social betterment. Programmes chart out a perceived course whereby wrongs might be put to rights, deficiencies of behaviour corrected, inequalities of condition alleviated. Programmes are thus shaped by a vision of change and they succeed or fail according to the veracity of that vision. Evaluation, by these lights, has the task of testing out the underlying programme theories. When one evaluates realistically one always returns to the core theories about how a programme is supposed to work and then interrogates it - is that basic plan sound, plausible, durable, practical and, above all, valid?

Realist evaluation has a distinctive account of the nature of programmes and how they work, of what is involved in explaining and understanding programmes, of the research methods that are needed to understand the workings of programmes, and of the proper products of evaluation research. The chapter will work though such specifics presently, but in these initial remarks it is appropriate to stress the underlying purpose of realist evaluation. What is in it for the policy analyst? What should you expect if you are commissioning or using a piece of realist evaluation? The short answer here is that such evaluation has an explanatory quest – programme theories are tested for the purpose of refining them. The basic question asked, and hopefully answered, is thus multi-faceted. Realist evaluations asks not, 'What works?' or, 'Does this program work?' but asks instead, 'What works for whom in what circumstances and in what respects, and how?'

Such questions drive the evaluator to inspect the reasoning of legions of programme stakeholders in a cavalcade of intervention contexts. In so doing, they dredge up in the

research conclusions many shades-of-grey about the then-and-there that will not always be welcome by policy makers needing make black-and-white decisions about the here-and-now. We will come to the matter of how to get to grips with complexity of interventions and the fallibility of findings towards the end of this essay. For now, we stress the reason for realist caution. Programmes are products of the foresight of policy-makers. Their fate though ultimately always depends on the imagination of practitioners and participants. Rarely do these visions fully coincide. Interventions never work indefinitely, in the same way and in all circumstances, or for all people. As it embarks on its explanatory quest realist evaluation is (realistically) panacea phobic.

'Realist' or 'Realistic' Evaluation - What's in a Name?
It is perhaps worth clearing up a little terminological confusion from the outset. This chapter uses the tag 'realist' to describe the preferred approach, though it describes a strategy first set down in Pawson & Tilley's *Realistic Evaluation* (1997). The reasons for the little terminological switch are, by the way, set down in the first page of that volume. Here, we settle on 'realist evaluation' because it has become the preferred nomenclature of other authors (Henry, Julnes and Mark, 1998; Mark, Henry and Julnes, 2000). This section thus attempts to distil the views of all fellow realists and presents a trans-Atlantic perspective. The three above-mentioned texts, incidentally, might be considered a preliminary reading list for those not acquainted with the approach.

1. The nature of programmes and how they work

The cornerstone of the realist project is a distinctive viewpoint on how intervention bring about change. It is only by understanding and probing its apparatus of change that one can evaluate a programme. According to realist evaluation programmes are 'theories', they are 'embedded', they are 'active', and they are parts of 'open systems'. Each one of these facets is described below, using illustrations from across the policy waterfront.

1.1 Programmes are theories

Programmes are theories incarnate. They begin in the heads of policy architects, pass into the hands of practitioners and, sometimes, into the hearts and minds of programme subjects. These conjectures originate with an understanding of what gives rise to inappropriate behaviour, or to discriminatory events, or to inequalities of social condition and then move to speculate on how changes may be made to these patterns. Interventions are always inserted into existing social systems that are thought to underpin and account for present problems. Changes in patterns of behaviour, events or conditions are then generated by bringing fresh inputs to that system in the hope of disturbing and re-balancing it.

For instance, some health education theories explain the unhealthy life styles of adolescents by the undue influence popular culture and poor examples created by film, soap and rock stars. This has led to the programme theory of trying to insinuate equally attractive but decidedly healthy role models (e.g. sport stars) into the pages and onto the airwaves of the teen media. Such a conjecture, known amongst denizens of health education as 'Dishy David Beckham theory', runs risks in both diagnosis

and remedy. Suffice to say that the evidence hereupon indicates the popularity of pouring over pictures of Beckham and friends in the teen magazines, but that as an activity it continues to exercise girls' minds rather than their bodies (Mitchell 1997).

We also note here (c.f. theories-of-change) that intervention theories are always multiple, mirroring the many decisions that have to made in drawing up and implementing an intervention. An example here is the chain of reasoning put in place to support the registration and community notification programme for released sex offenders in the US (Megan's Law). Decisions have to be made on who are the high risk cases, what information should be registered over what periods, how and to whom their identities should be released, how to monitor and regulate movement, how to control community reaction and encourage surveillance and so on. In each of these instances those responsible for the programme figure out what is likely to be best practice. So, for instance, there has to be a programme theory about the boundary of the community within which notification should occur. On the release of the offender, some authorities deposit posters according to a standard measure of the number of blocks or a yardage radius from the offender's dwelling. Some authorities prefer to 'eyeball' a map and make decisions *ad hoc*. Some use a piece of software called Megan's Mapper to make the decision for them (as well as printing address labels). And one official reports that his county draws the line on the basis of 'looking at how far the offender has to travel to buy cigarettes'. We hope that a rather jocular observation on to the weakness of this hypothesis for non-smoking offenders will not obliterate the crucial point that some of these hunches are probably more helpful than others, and that the effectiveness of programmes as a whole will depend of the combined efficacy of such theories.

1.2 Programmes are embedded

As they are delivered programmes are embedded in social systems. It is through the workings of entire systems of social relationships that any changes in behaviours, events and social conditions are effected. A key requirement of realist evaluation is thus to take heed of the different layers of social reality which make up and surround programmes. For instance, a programme of prisoner education and training may offer inmates the immediate resources to start on the road to reform. Whether the *ideas* transmitted will cement depends upon a further four I's: i) the *individual* capacities of trainees and teachers, ii) the *interpersonal* relationships created between them, iii) the *institutional* balance within the prison toward rehabilitation or containment, iv) the wider *infra-structural* and welfare systems that support or undermine the return to society.

This is an important principle that is duplicated across all policy domains: the success of job training programmes ultimately depends on whether there is work to be had; neighbourhood renewal depends on improving the lot of citizens but also on retaining them in the locality; sermons on recycling need waste management services to support them; welfare-to-work incentives can be overridden by black-market opportunities. Realism carries a profoundly sociological view on social change. In relation to individuals it is assumed that programme resources can be the spur promoting change, but whether and to what extent that transformation will hold is contingent on the social circumstances of that person. Commissioners of realist evaluations should thus

expect the research to take cognisance of the subjects' characteristics, their relationships, their organisational position, their economic conditions and so on.

1.3 Programmes are active

The triggers of change in most interventions are ultimately located in the reasoning and resources of those touched by the programme. Effects are thus generally produced by and require the active engagement of individuals. Take two dental health programmes – i) the fluoridation of water and ii) publicity on brushing twice-a-day. The former is a rare example of a passive programme. It works whenever water is swallowed and thus happens to whole populations. They are not required actively to engage with it. But in the health education intervention, the message is the medium and that message may not be so readily swallowed. The advice on the importance of dental hygiene may indeed be welcome, heeded and thus acted upon; or it may be missed, ignored, forgotten, found boring and thus overlooked; or it may be challenged on scientific grounds, regarded as paternalistic and thus disputed; or it may simply be overridden by the lure of sugar.

And so it is with the vast majority of programme incentives. This inevitability that social and public policy is delivered through active programmes to active subjects has profound implications for evaluation methodology. In trials with medical interventions human volition is seen as a contaminator. The experimental proposition under test is about whether the treatment (and the treatment alone) is effective. As well as random allocation of subjects, further safeguards such as the use of 'placebos' and 'double blinding' are also utilised to protect this causal inference. The idea is to remove any shred of human intentionality from the investigation of whether treatment brings about cure. Active programmes, by contrast, only work though the stakeholders' reasoning. And, as we shall see, this means that an understanding of the interpretations of programme participants is integral to evaluating its outcomes.

1.4 Programmes are open systems

Programmes cannot be fully isolated or kept constant. Unanticipated events, political change, personnel moves, physical and technological shifts, inter-programme and intra-programme interactions, practitioner learning, media coverage, organisational imperatives, performance management innovations and so on make programmes permeable and plastic. Such externalities always impact on the delivery of a programme and this entails that they are never quite implemented in the same way. Realism, however, goes a step further in understanding the changing nature of programmes. That is to say, they are regarded as self-transformational. Successful interventions can change the conditions that made them work in the first place.

The so-called 'arms race' in crime reduction programmes is a prime example. Having suffered the blows of the introduction of a new scheme, the criminal community is often able to figure out the intervention modus operandi and thus adapt its own modus operandi accordingly. A rather vivid example is the changing impact of town centre CCTV cameras. On installation, these were regarded with some foreboding by marauding youth. But once their positioning and range was understood, and as soon as it was gathered that impact depended on how the images are deciphered by an operator, and then on how quickly the police can mobilise in response to a call, a

different set of options opened up. The most bizarre twist is noted by Norris and Armstrong (1999) who observed youths staging mock-fights in front of city-centre cameras in order to prompt operator action for the pleasure of wasting police time. The result here and across the crime prevention field is that a constant stream of fresh initiatives is thus required to keep pace.

It must be stressed that this 'morphogenesis' of programme effects is no quirk of crafty criminals and chasing cops. Another example can be drawn from the area of welfare benefits. It is now widely recognised that such schemes are no longer concerned merely with alleviating the harshest edges of poverty. Rather they now aim to interpose within interlocking cycles of deprivation – for instance to balance the provision of assistance with the avoidance of welfare dependency, and to balance the goal of work preparation with the promotion of self-help. The pushes and pulls differ for welfare clients with varying impairments and different personal and family histories. The result is that welfare policy-making is always a matter of making adjustments to a scheme rather than creating programmes anew. Any learning we gain about such schemes has to be translated into minute adjustments performed upon a beam balance already loaded with benefit rates, dependant's allowances, earnings disregards, payment tapers, training grants and so on.

What we have tried to do in this section is to provide a portrait of how realism sees the process of change as instigated by policy innovation and programme intervention. Complex, differentiated and intertwined as it is, we know that it is a picture that will be recognised by policy makers and practitioners. We now turn to the implications for research and evaluation.

2. Basic concepts in the explanation and understanding of programmes

As should be clear from the previous section, realists regard programmes as rather sophisticated social interactions set amidst a complex social reality. Science deals with intricacy by using an analytic framework to break down systems into their key components and processes. Realist evaluation stresses four key linked concepts for explaining and understanding programmes: 'mechanism', 'context', 'outcome pattern', and 'context-mechanism-outcome pattern configuration'.

2.1. Mechanism

Mechanisms describe what it is about programmes and interventions that bring about any effects. Mechanisms are often hidden, rather as the workings of a clock cannot be seen but drive the patterned movements of the hands. This realist concept tries to break the lazy linguistic habit of basing evaluation on the question of whether 'programmes work'. In fact, it is not programmes that work but the resources they offer to enable their subjects to make them work. This *process* of how subjects interpret and act upon the intervention stratagem is known as the programme 'mechanism' and it is the pivot around which realist research revolves. Realist evaluation begins with the researcher positing the potential processes through which a programme may work as a prelude to testing them.

The concept is best grasped through an illustration. The 'primary school breakfast club' is a very popular measure used to boost early education performance, often

included within community regeneration initiatives. The key point here is that 'the measure' is not the basic unit of analysis for understanding causation. A measure may work in different ways or, in realist parlance, they may trigger different mechanisms (M_1, … , M_n). A breakfast club may aid classroom attentiveness by offering the kids a 'nutritious kick-start' (M_1) to the day, which they might not otherwise get. And/or it may act as a 'summoning point' (M_2) to prevent kids loitering or absconding or misbehaving in the chaotic period before school. And/or it may act as an 'energy diffuser' (M_3) to soak up gossip and boisterousness before formalities commence. And/or it may enable to school to present a more 'informal face' (M_4) to those uninspired by classroom and book learning. And/or it may act as a 'pre-assembly' (M_5) enabling teachers to troubleshoot potential problems and seed the day's schedules. And/or it might give parents and school staff an 'informal conduit' (M_6) to mix and offer mutual support. Mechanisms also explain a programme's failure, of course, so to this list we might add some adverse processes. It may act as an opportunity for 'messing about' (M_7) if only ancillary staff are on duty; it might provide an unintended 'den of iniquity' (M_8) for planning the day's misdeeds: or it might prove a 'cultural barrier' (M_9) because inappropriate food is served, and so on.

Having distinguished a measure from its mechanisms, it is perhaps appropriate to cover a couple of other potential misunderstandings of the notion of mechanism. Many programmes have multiple component interventions. A community regeneration programme, for instance, may contain a string of measures alongside the breakfast club, such as 'IT kiosks', 'neighbourhood wardens', 'one-stop employment-benefit shops' and so on. The term mechanism is not used to distinguish these components, each one of which will work through its own underlying processes. Interventions also often involve long sequences of steps before the outcome (e.g. the Megan's Law example presented earlier). Again, these are not what are referred to with the term mechanism. Rather mechanism refers to the ways in which any one of the components or any set of them, or any step or series of steps brings about change. Mechanisms thus explicate the logic of an intervention; they trace the destiny of a programme theory, they pinpoint the ways in which the resources on offer may permeate into the reasoning of the subjects.

2.2 Context

Identifying the crucial programme mechanisms is only the first step in a realist evaluation. It is also always assumed that they will be active only under particular circumstances, that is, in different contexts. Context describes those features of the conditions in which programmes are introduced that are relevant to the operation the programme mechanisms. Realism utilises contextual thinking to address the issues of 'for whom' and 'in what circumstances' a programme will work. In the notion 'context' lies the realist solution to the panacea problem. For realism, it is axiomatic that certain contexts will be supportive to the programme theory and some will not. And this gives realist evaluation the crucial task of sorting the one from the other.

We can return to the earlier example of prisoner education programmes for an example, since it demonstrates the wide compass of the notion of contextual constraints. Let us assume that within the programme lurk mechanisms that may aid rehabilitation such as increases in cognitive skills, social dexterity, qualifications and so on. Whether these can be acquired and cashed in depends on circumstances both

within and without the prison classroom. Contexts both enable and constrain and the individual capacities of the learner are obviously relevant and so characteristics such as being 'drug and alcohol free' (C_1), 'prison weary' (C_2), and (obviously) having some 'aspiration' to go straight (C_3) might be significant in terms of using educational resources. The culture of the classroom is likely to make a difference and the education department's ability to steer a path between the violent/macho culture of the wings (C_4) and the containment/surveillance culture of the prison (C_5) may be crucial in sustaining the learning process. Rehabilitation is only achieved on the 'outside' and the activation of learning instincts may count for nought without a stable home to return to on release (C_6) and further education and training support (C_7). The wider community itself is crucial in 'permitting' rehabilitation and so opportunity afforded by employment havens (C_8) might be crucial, as will be wider norms about toleration/retribution (C_9) of/against 'ex-cons' in neighbourhoods and workplaces.

Context must not be confused with locality. Depending on the nature of the intervention, what is contextually significant may not only relate to place but also to systems of interpersonal and social relationships, and even to biology, technology, economic conditions and so on. Standard measures of demographic difference in social science, in terms of sex, age, ethnicity, and class, are in themselves unlikely to capture what is contextually important, but may at best be rough indicators. The salient conditions must also be identified as part of the programme theory. These generally suppose that certain types of subjects are in with a better chance and that certain institutional arrangements are better at delivering the goods. Contextual knowledge is absolutely crucial to the policy maker. The best programmes are well-targeted programmes and the notion of context is a crucial entrée to that goal.

2.3. Outcome patterns

Programmes are almost always introduced into multiple contexts, in the sense that mechanisms activated by the interventions will vary and will do so according to saliently different conditions. Because of relevant variations in context and mechanisms thereby activated, any programme is liable to have mixed outcome-patterns.

Outcome-patterns comprise the intended and unintended consequences of programmes, resulting from the activation of different mechanisms in different contexts. Realism does not rely on a single outcome measure to deliver a pass/fail verdict on a programme. Nor does it make a hard and fast distinction between outputs (intermediate implementation targets) and outcomes (changes in the behaviour targeted). Outcome patterns can take many forms and programmes should be tested against a range of output and outcome measures. Much is to be learned by monitoring programmes across a range of such measures. We may find an influence at point A or in respect of characteristic B. But no change may be discernible at time C or in relation to property D. Then again, we may find a quite unexpected movement at E and an unwanted outcome at F. Deciphering the reasons for such a variegated pattern can give us vital clues to the workings of programmes.

For instance, it may be instructive to learn that CCTV installation increases car park 'turnover' (O_1) as well as observing a fall in 'crime rate' (O_2). This might prompt the hunch that 'public presence' as well as 'deterrence' or 'detection' is causing the

change. This secondary hypothesis about the importance of 'natural surveillance' could be further checked out by comparing crime rates at 'busy' (O_3) and 'slack' (O_4) times of the day Moving to an accident prevention intervention provides another useful example of the need for multiple measures of outcome. Consider a programme based on the free distribution of smoke alarms. Monitoring how many had been properly installed (O_1) and then maintained in the medium and long term (O_2) may give a better indication of how and why and for whom they are effective, rather than relying on long-term changes in death or injury by fire (O_3).

Policy makers are often besotted and sometimes bewildered by performance measures. This notion of 'outcome patterns' allows for a more sensitive evaluation of complex programmes. Hunting down outcome patterns may involve implementation variations, impact variations, socio-demographic sub-group variations, temporal outcome variations, personal attribute outcome variations, regional outcome variations, biological make-up outcome variations and so on.

2.4. Context mechanism outcome pattern configuration

By now it should be clear that realist evaluation has little use for a 'find-the-intervention-X-that-cures-problem-Y' notion of programme building. All interventions involve multiple perturbations of pre-existing regularities in behaviours, events or social conditions, leading to the creation of many new regularities. Such outcome-variations are found routinely within programmes of all types. Any programme rolled out nationally will have winners galore and losers in abundance, and such differences will occur within and between each programme trial. The nature and source of these internal differences is a key focus of attention in realist evaluation.

Realist evaluation is about theory testing and refinement. Context-mechanism-outcome pattern configurations (CMOCs) comprise models indicating how programmes activate mechanisms amongst whom and in what conditions, to bring about alterations in behavioural or event or state regularities. These propositions bring together mechanism-variation and relevant context-variation to predict and to explain outcome pattern variation. Realist evaluation thus develops and tests CMOC conjectures empirically. The sign of a good evaluation is that it is able to explain the complex signature of outcomes (Mark et al, 2000).

The 'findings' of realist evaluation thus always try to pinpoint the configuration of features needed to sustain a programme. Let us take the example of a very simple device used to try to reduce domestic burglary, namely 'property-marking' (using indelible, immovable tags). For it to work optimally and efficiently, however, requires a complex alignment of implementation and contextual factors. Property-marking works better in reducing domestic burglary if overall levels of marked property are high; when crime targets are concentrated with few alternatives; in small, well-defined communities providing plausible conditions for tracing stolen property; with attendant persuasive publicity demonstrating increased risk of being caught; and thus only over a limited time period (Laycock, 1997).

Configurational thinking – the key to programme building
The concept 'context-mechanism-outcome-configurations' describes how different components of a programme need to be harmonised. However, it is an ugly pug of a term, which may detract from the crucial idea it seeks to insinuate into evaluation. The basic notion, nevertheless, is commonplace in social explanation. The logic utilises a 'configurational' approach to causality, in which outcomes are considered to follow from the alignment, within a case, of a specific combination of attributes. Such an approach is used, for example, by historians attempting to figure out why a social movement had 'worked'. Explanations for why England experienced an early industrial revolution turn on identifying the *combination* of crucial conditions such as 'technological innovation', 'weak aristocracy', 'commercialised agriculture', 'displaced peasantry', 'exploitable empire' and so on (Moore, 1966, ch1). For a more homely metaphor we turn to the kitchen. Recipes 'work' by assembling the right ingredients in the correct proportion to suit the tastes of the diner. Think too of bridges, cars, aeroplanes, computers and gardening. Though physical, they all 'work' (when they do) through configurations. Programme building is also a matter of getting the right ingredients in place in the right setting to suit the needs of particular sets of consumers. A configurational evaluation of programme components is a necessary prerequisite to such decision making.

3. Strategies and methods of realist evaluation

3.1. Scope of realist inquiry

Realist inquiry can be located in every social science discipline. For example, it has found a home in philosophy (Collier, 1994), law (Norrie, 1993), psychology (Greenwood, 1994), economics (Lawson, 1997), sociology (Layder, 1998: Archer 1995), management studies (Ackroyd and Fleetwood, 2000), and geography (Sayer, 2000). Given this miscellany of topics, it should be clear that realism is not a research technique as such. It is a 'logic of inquiry' that generates distinctive research strategies and designs. And so it is with realist evaluation. It may be used prospectively (in formative evaluations), concurrently (in summative evaluations) or retrospectively (in research synthesis). Realist evaluation, moreover, has no particular preference for either quantitative or qualitative methods. Indeed it sees merit in multiple methods, marrying the quantitative and qualitative, so that both programme processes and impacts may be investigated. The precise balance of methods to be used is selected in accordance with the realist hypothesis being tested, and with the available data.

3.2 The realist research cycle

Realist research is absolutely conventional, and pleased to be so, in utilising the time-honoured 'research cycle' of hypothesis testing and refinement. In evaluation terms the wheel of science turns as in figure one:

Figure one about here

3.21 Realist evaluation normally begins by eliciting and formalising the programme theories to be tested. There can be various sources of these including documents,

programme architects, practitioners, previous evaluation studies and social science literature. Hence documentary analysis, interviews and library searches may all be involved. Interviews with programme architects and documentary analysis can help articulate the formal or official programme theory in CMOC terms. Interviews with practitioners are deemed especially important as discussions of apparent programme successes and failures can lead to fine-grained hypotheses about what works for whom and in what circumstances and respects. This stage is the launching pad for realist evaluation and is, in many ways, its most distinctive phase. What is involved is bringing the imagination to bear in 'thinking though' how a programme works.

This is a research skill that can be nurtured and developed and we devote a small exercise to this at appendix 1.

3.22 The next stage is to collect data that will allow interrogation of these embryonic hypotheses. Programme documents and/or practitioners will have suggested that a particular resource is vital, and that a particular way of interpreting it is the key to success. The initial hypotheses will also often cover the type of client who is better placed to succeed and some institutional locations where this is most likely to happen. Data gathering has the task of trying to match information to these various leads. Given the preliminary theories cover mechanisms and contexts and outcomes, data collection has to be both qualitative and quantitative. The evaluator has, quite literally, to scavenge for the best data to test out the theories. Existing administrative records might be put to use, stakeholders of all type might be interviewed and shadowed, dedicated before-and-after measures might designed and put in place, focus groups might be assembled to unearth reasons for crucial choices, and so on.

3.23. The third stage is to subject a whole package of CMOC hypotheses to systematic test, using data sets assembled as above. The purpose of the analysis is to see if the model will explain the complex footprint of outcomes left by the programme. There is no single analytic method suitable for this purpose and the design of data analysis is a matter of the subtlety of the proposed theories and the availability of data. Realism's primary expectation is that there will be a nuanced outcome pattern of successes and failures within and across interventions. The primary tactic is thus to interrogate these hypotheses by making sub-group comparisons. Overall, the explanatory theory is investigated by devising and testing out multiple comparisons identifying winners and losers amongst subjects and pros and cons in programme delivery. Will the emerging theory explain implementation variations, impact variations, socio-demographic sub-group variations, temporal outcome variations, personal attribute outcome variations, biological make-up outcome variations and so on?

3.24. The final stage is the assessment and interpretation of the analysis. Have the theories about how the programme worked been supported or refuted by the proceeding analysis? Judgement on this score is invariably mixed, with some output and outcome variations being clear and intelligible, whilst others remain quite puzzling. Just as with programme building itself, quite unanticipated effects can be uncovered in the sub-group analysis and these require a revisit to the hypothesis drawing board. Stage four of the process is thus an ever-repeating cycle, the purpose being to draw closer to explaining the complex signature of outcomes left behind by an intervention. This may be attempted in further rounds of analysis on the same

programme within the same evaluation, or by picking up the same theories in other evaluations in the same family of programmes.

To illustrate the potential variations in design, data and method within the research cycle, two fragments of realist analysis (one quantitative and one qualitative) are presented in appendix 2 (A & B).

3.3 The research relationship

Realist research is distinctive in its understanding of the research relationship between evaluators and stakeholders. There is a longstanding debate in evaluation about whether the evaluator should:

i) take an insider perspective, viewing the knowledge of stakeholders as paramount in both understanding a programme and making it work, and thus engaging with them in developing a shared understanding about programme improvements.
ii) take an external perspective, relying on objective methods to make the judgement about the efficacy of the programme, thus treating stakeholders as sources of data to input into these standard research designs.

On the realist approach, stakeholders are regarded as key sources for eliciting programme theory and providing data on how the programme works. But it is not assumed that they are all-knowing, nor that they will necessarily agree on how, for whom and in what circumstances a programme will work. Stakeholders generally have experience of and thus expertise in particular phases and process within an intervention. Realist evaluation requires data on process and outcome, and on individuals, interrelationships, institutions and infra-structures. In order to assemble this bricolage of data, there needs to be a division of labour of information and informants.

The realist interview recognises the theory-testing purpose of evaluation and it is this that shapes the research relationship. Subjects are thus understood to be trying to respond to what they deem the interests of the interviewer. Collecting data that are relevant to evaluation thus involve teaching (often in more or less subtle ways) the respondent the particular programme theory under test in order that subjects can summon responses which speak in relevant ways to CMO configuration at issue. The respondent, having learned the theory under test, is able to teach the evaluator about those components of a programme in a particularly informed way.

Further details and examples of what realists call the 'teacher-learner relationship' or 'assisted sensemaking' may be found in Pawson and Tilley (1997, chapter six) and Mark et al (2000, part two). The final point to note here is that evaluation commissioners often seek to steer a course between the external and the internal, between appraisal and development, between audit and support, and that 'theory-testing' is a particularly useful pathway through which to steer this middle course.

4. Realism's place in the policy cycle: formative, summative and synthetic approaches

4.1 Realism and pluralism

In this section, we reach the cusp of the evaluator's task and the moment when the results of evaluation are realised. There are a great many issues to be confronted here, the first of which is about timing and how realist evaluation fits into the various processes and stages of policy making and implementation. As is well known, the tempos of policy and of research are very different. From either perspective, pot is inclined to call kettle black. Researchers, used to piloting, triangulating, checking and balancing, are prone to complain about being shoed into 'quick and dirty' evaluations. Policy makers, in the teeth of impending and perpetual policy change, are likely frown upon many evaluations as 'slow and ambivalent'.

There is an obvious solution here – namely, harmonising research findings to each turn of the policy cycle. Despite the caricatures, policy-making is hardly momentary and unitary. Typically, there will be the following phases: preliminary analysis of the problem, the speculative first-spark of a potential solution, the mulling over of the plausibility of the intervention; the spelling out of implementation details, the contemplation of pilots and demonstrations, the move to full stream, and in the longer run, the decision on whether to continue, expand or curtail. None of these judgements, moreover, is arrived at in isolation or in the abstract. They are always made on the back of limited resources and thus are made as decisions-taking-alternatives-into-consideration.

Given this series of incremental judgements, it is apparent that the supportive research could and should be commissioned and assimilated throughout the policy cycle. Evidence can be targeted at big policy ideas or small implementation details. It can be aimed prospectively at delivery, trying to figure out the best way to marshal together a programme or service. It can be placed concurrently with a programme, asking the traditional question about whether and in what respects it is working. It can be put in place retrospectively, calling on all past evidence about former incarnations of an interventions in order to inform whether and what guise it might be targeted at an impending problem.

Evaluation, in short, can be formative or summative or synthetic. These orientations have in the past, alas, been associated with the bun-fight between evaluation paradigms and painted as rivals rather than as sources of complementary information. The general thrust of the *Magenta Book* is towards methodological pluralism and thus of pursuing a varied diet of policy analysis, scoping studies, plausibility studies, developmental evaluations, impact evaluations, audit, cost-benefit analysis and so – *as appropriate to the policy purpose*. Realism sits conformably enough with this outlook because it too can be performed prospectively, concurrently or retrospectively. The linking theme is, of course, the development, testing and refinement of programme theory. Just as programme theory can be interrogated with both qualitative and quantitative data, evidence can also be compiled by looking backwards to bygone studies, by taking snapshots of an unfolding programme, or by working forwards fine-tuning a programme by a process of trial and error.

To illustrate the potential variations in data sources and their origin within the research cycle, two further fragments of realist evaluation (one formative and one research syntheses) are presented in appendix 2 (C & D).

4.2 Realism and Knowledge Management

One of the most interesting consequences of the current tryst with evidence-based policy has been the development of 'knowledge management strategies' in the analytic services divisions of many government departments. The need has arisen out of the insight described in the previous section. Interventions are becoming more complex and, if they are to be understood and improved, they require underpinning by a multifaceted body of evidence. Matching the one to the other requires a strategy.

Realism's vision of programmes as configurations of theories provides a rationale that may be helpful in formulating that strategy. The logic can be understood by drawing a contrast with what might be thought of as the 'standard evaluation regime' as in Figure 2. The conventional arrangement is for an evaluation to be commissioned as and when a new intervention X is mounted (Fig 2a). The research is thought of, and becomes known, as the 'evaluation of X'. This one-to-one relationship has survived through the commissioning of even the largest of the current national programmes. Sometimes research teams may be broken down into regional (e.g. 'north', 'midlands', 'south') and/or substantive groups (e.g. 'crime', 'health', 'education'). And sometimes, as in Fig 2b, there may be a methodological division (e.g. 'process' and 'outcome' evaluation teams). The key linkage remains however, in as much as evaluation activities are firmly attached to the *current* intervention.

This direct connection has become broken somewhat in the trend towards review and synthesis in evidence-based policy. Knowledge management logic now tends to reckon (quite correctly) that much is to be learned from research on previous incarnations of bygone interventions. The assumption is that much the same programmes get tried and tried again and researched and researched again, and learning accumulates by pooling the previous evidence.

A realist knowledge management strategy takes this dealignment of intervention and research on board, but proceeds a step further (Figure 2c). The starting point, as ever, is the disaggregation of interventions into their component theories. If we take an example a complex programme like the *New Deal for Communities* it contains a whole raft of hypotheses and decision points. For instance it assumes that: a) concentrated, multi-agency working will harness a more effective set of services than will conventional one-issue-at-a-time delivery, b) that community leaders can be identified to help manage a much more user-oriented intervention, c) that 'quick wins' are needed to galvanise community support, d) that community members can make gains without significant mobility and displacement out of the locality, e) that there is no 'intervention fatigue' or cynicism associated with multiple past and present regeneration efforts, etc, etc. In addition there are a whole range of theories associated with the specific package of programmes put in place such as: f) breakfast clubs, g) fitness centres h) warden patrols etc. etc. We draw the list to a halt here and represent the whole body of theories with the flow of darker arrows in Figure 2c.

Figure 2 about here

The basic realist evaluation strategy is to test these component theories and, to return to a previous theme, the basic expectation is that such interventions require a portfolio of different evaluation methods. It might be that theory (a) requires a process

evaluation; theory (d) utilises geographic modelling; theory (e) needs an historical case study; theory (g) has a systematic review at the ready; theory (h) requires a before-and-after crime survey, and so on. These are depicted as the run of lighter arrows in Figure 2c.

The realist knowledge management strategy comes into play in the third element of Figure 2c. It is no more and no less than a 'recycling process'. There is a regular bus service of new programmes and, to be sure, a new one will always come along (Programme 2) that has some of the components of previous efforts. Note that the new incarnation does not have to be a 'replication' elsewhere of the 'same' intervention. Initiatives aimed at quite different problems often rely on common components and suffer similar setbacks. Hypotheses about joint-working are common in trying to improve services across the policy waterfront – why not learn from existing research on the same point? Problems of gentrification and displacement are a familiar feature of community regeneration – why not drawn upon previous studies? Once they are decoupled from a specific intervention, evaluations can find extended use. The dashed lines in Figure 2c depict this process of joined-up thinking in the utilisation of evaluations. If one imagines the diagram extended to scores of evaluations of scores of programmes, then the opportunity to re-use, make more of, and make savings on, evaluation becomes evident.

Our final remarks on this point are addressed to those who commission evaluative research. The idea is to give evaluation a history and to stop commissioning the same pieces of work. Our view is that policy makers can always expect a little learning on the basis of previous studies. But there is also a responsibility to share and prolong the utility of a new inquiry. This can be done by ensuring that the researchers pay attention to the theory underlying an intervention, for it is an appreciation of the scope of that theory that offers transferable knowledge, which can then be picked up by future decision makers. More details on a realist evaluation strategy for dealing with complex programmes may be found in Pawson (2004).

5. The nature, presentation and utilisation of findings from realist evaluation

Realist evaluation is applicable in principle to all forms of programme evaluation, and to all areas of social and public policy. In every case, the goal is to produce a tested theory about what works for whom in what circumstances and in what respects. This end product is never a pass/fail verdict on an intervention but an understanding of how its inner workings produce diverse effects. Strong realist evaluations are thus intended to lead to better-focused and more effective programmes. This section describes the limits to that endeavour, beginning with a close consideration of the nature of the findings of realist evaluation (5.1, 5.2), then going on to consider matters of presentation (5.3) and utilisation (5.4)

5.1 The provisional nature of findings

The findings of realist evaluation mirror the nature of programmes. As we have tried to demonstrate in previous sections, programmes are complex social systems

introduced amidst complex social systems. Both of these systems are open. Programmes are composed of an intricate, reverberating sequence of decisions, which will shift this way and that over the duration of the programme. The systems in which they are introduced – organisations, localities, welfare regimes, moral communities, belief configurations – are also in permanent transition. Furthermore, the relationship between the two systems is dynamic. A powerful programme can alter the conditions that made it work in the first place, thus changing its effectiveness over time (recall section 1.4).

This state of affairs places profound limitations on what evaluation can achieve. Realism also means pragmatism, of course, and this is another feature of the perspective. Having a subject matter composed of double and triple doses of complexity does not mean that we are forever blind to the consequences of interventions. It does mean, however, that our understanding will always be partial and provisional and, consequently, that we must be somewhat modest in our ambitions for evaluation. The purpose of this section is to explore these restrictions in a little more detail. We begin with a pragmatic summary of what evaluations can achieve.

What can evaluations tell us?
Evaluating in open systems is a profoundly uncertain business. What are the limitations and what can the decision-maker reasonably expect? It should be possible to detect outcome changes over the course of a programme. These are likely to be complex, with certain attributes and behaviours shifting rather more than others. It should be possible to detect some processes activated within the programme that may be responsible for and make sense of the changes observed. It should be possible to detect something about the conditions and circumstances in which the intervention is mounted which allow for and make sense of the observed process and outcomes.

We emphasise these imperfect products of an evaluation in this deliberately stark synopsis, firstly to show that the 'findings' of evaluation are inevitably equivocal, but also to point out that they are still profoundly useful. The caution of realist evaluation applies, inevitably, in respect of the issue of 'attribution'. In its classic guise, this refers to the problem of how and how safely we may infer that a programme is responsible for outcomes observed. Realism rejects this particular formulation arguing that that programmes are active, and thus it is the operation of particular mechanisms acting in context that brings about change. All well and good, but one still has the problem of attribution. From the preceding examples it is clear that a programme may operate through many different mechanisms. So, how can we be sure that a particular mechanism or set of mechanisms is in fact responsible for bringing about change in a given context?

The answer is captured in a provocative little phase we have used, quite deliberately, in the boxed insertion above, namely 'make sense of'. Thus, attribution is dealt with when we accept that action of a mechanism *makes sense of* the particular outcome pattern observed. Now, sense-making is not the platform on which other traditions of evaluation are built but it is the cornerstone of the realist approach (and we would say, for good measure, that of science itself). Hence, rather more needs to be said about its basic operation.

What we have in mind is a continuous programme of weeding out of alternative theories about how a programme works. This procedure, which also carries the rather grander title, namely 'theory adjudication' (ref), can be illustrated as follows. Suppose a CCTV system is installed in a car park and suppose researchers are able to observe a subsequent fall in car crime. Now, there are a couple of theories that could easily 'make sense of' this outcome pattern. It might be that the system catches thieves red-handed and that increased detection and arrest follow from images of crimes committed being caught on camera. Or, it might be that the thieves get nowhere near the car park and that thefts fall because of increased trepidation about the risks of being caught by the new devices. So is detection or deterrence the causal mechanism? In this instance there is a rather simple means of adjudicating between the two theories. The evaluator can examine the number of criminals apprehended and arrested as a result of their CCTV exposure, and calculate just how much of the overall fall in crime is accounted for in this mode. Readers interested in this particular adjudication may like to consult Tilley (1993) and contributions to Painter and Tilley (1999) and to Gill (2003).

Good empirical work in the realist tradition should always carry this strategy of developing and adjudicating between rival explanations for programme outcomes. As a research design, of course, it does have a clear limitation. Having sorted out a couple of alternatives, as per our example, does not of course preclude further potential explanations. Returning to the car park, it may be that the arrival of the cameras has resulted in much greater usage. This turnover of customer results in much greater 'natural surveillance' and it may be that nosey-parkers constitute the vital mechanism protecting against theft. Again, it is not too difficult to imagine a test of this theory, a comparison of crime rates when the car park is full (thus having no natural surveillance) against periods when there is considerable entry and exit.

Of course, alternative explanations no more end at three than they do at two. And this conundrum spells out the ineluctable limitation of realist evaluation. Programmes modify and transmute and are constantly being tried out in fresh circumstances, so there are always more and new potential explanations of their efficacy. At this point, realist evaluation resorts to the pragmatism principle, which says simply - go as far as you can in sorting and sifting the rival explanations. All eventualities cannot be anticipated but, importantly, knowledge in considerably improved on each adjudication. If, for instance, policy makers know that CCTV works significantly more through deterrence rather than detection, they are in a better position to plan its next application and modification.

In general, there are an infinite number of explanations for why, when and how a programme works but there are only so many ways in which a programme might be improved. Realist evaluators need not wait to figure out the totality of explanations but should concentrate, therefore, on those programme ideas and variations, which are 'on the table'.

5.2 The 'middle-range' nature of findings

Realist evaluation steers a path between making universal claims about what works, and focusing on the particulars of specific measures in specific places relating to specific stakeholders. Thus it places no faith in black-and-white (or even red, amber

and green) policy prognostications of the kind that suppose that street-lighting works to reduce crime, or that mentoring programmes for disaffected youth are harmful, or that 5-fruit-and-veg-portions-a-day health education initiatives have a null effect. And although it acknowledges the actuality that a particular programme may indeed have failed because of Fred Bloggs's recalcitrance or succeeded because ACME had the spare managerial capacity to support it, it supposes that such personal and institutional details do not help when it comes to the transferability of evaluation's findings.

Realism operates at middle range, using concepts that describe interventions at a level between the big policy ideas and the day-to-day realities of implementation. Such 'middle range theory' has a time-honoured place in social science (Merton, 1968). No exposition is attempted here (see Pawson, 2000), since the key idea is readily intelligible through examples. Mechanisms are the engine of change in realist thinking and these describe how programme resources seek to influence their subject's reasoning. Such programme theories often have much in common, and it is the parallels between such ideas that are the prime focus of leaning in evaluation.

One can comprehend similarities in programme theory in several ways. Policy ideas are contagious. Ideas like 'public disclosure of performance data', 'zero tolerance', 'shared working', 'public-private partnership', 'mentoring', 'learning collaboratives' and so on develop in waves through government. Another way of viewing intervention commonality is to consider the nature of the 'proposal' made to programme subjects. Policy makers and practitioners are, in truth, able to offer relatively few ways of inducing change (the assertion here is not meant to reflect upon the lack of imagination of programme architects but on the limited nature of the interventions they may draw upon). The result is that the same programme theories repeat themselves from initiative to initiative and jump from domain to domain. The bravest rendition of this brave idea by Bemelmans-Videc et al (1997), who argue that, if one scrapes away programmes to their elemental bones, there are only three types of mechanism on offer, namely 'carrots', 'sticks' and 'sermons'.

Without necessarily going this far, realism supposes that evaluation can learn lessons from diverse programmes by operating at the middle range. For instance, consider an intervention like the free two-week distribution of nicotine patches to help poorer smokers quit (ref). It can be considered an entirely new idea, or it can be considered yet another incentive-based programme (carrot theory revisited). It is pretty well established that incentives work better according to the worth of the deal (benefit) and how readily it is realised (ease of access). And, lo and behold, it turns out that savings anticipated in the short NRT subsidy are not so great if one considers that poor smokers often avail themselves of black-market, duty-unpaid imports. Benefits are also reaped in a complex, down-the-line manner under the programme theory. The hypothesis is about subsidising a brief treatment so that there are funds available for its continued purchase, which will in turn pay off to allay future tobacco-costs. The rational calculation involved here is not so straightforward given the chaotic demands of debt-ridden budgets. The study once again points to the need to fine tune benefit and access in incentive packages. The methodological point here is that by operating at the middle-range, there is a much greater opportunity for realising and transferring the findings of evaluations.

5.3 The presentation of findings

Realist evaluation seeks to discover what it is about programmes that works for whom in what circumstances and in what respects, and why. Unsurprisingly, its results also seek to elucidate propositions of this kind (known formally as CMO propositions). It is often impossible to attend to everything in this explanatory ensemble in a single evaluation, with the result that findings are likely to concentrate on sub-sets of mechanisms, contexts and outcomes. The nature of these findings are illustrated throughout this text and, especially, in the four examples in Appendix 1. In general terms they might indicate:

- that a particular intervention works in quite separate ways
- that it gets implemented in different ways
- that it is more effective with some groups rather that others
- that it will find more use in one location rather than another
- that it has intended and unintended consequences
- that its effects are likely to be sustained or taper off

This body of findings is intended to help with the business of targeting and tailoring programmes. Distinctions of these kinds are often in policy-makers' minds when evaluations are commissioned and realist evaluation is assisted if such questions are well specified and forefronted.

In addition, the results of realist evaluation are also conditioned by the nature of the programmes they investigate and this means that findings are also 'configurational', 'middle-range' and 'adjudicationist'. These terms have already found explanation in the text, but their ramifications in terms of findings, results and conclusions can be summarised briefly. In addition to the above list, researchers should strive for, and commissioners should expect, findings that:

- Show how combinations of attributes need to be in place for a programme to be effective. Optimal alignments of implementation schedules, target groups and programme sites should be suggested.
- Have the potential for transferability on the basis using concepts that link to other programme theories and thus rest on further bodies of findings. Conclusions should evoke similarities and differences with existing evidence.
- Bring alternative explanations to the fore in order to sift and sort them. Programme building involves making choices between limited alternatives, and findings should be addressed at and help refine those choices.

Bringing these three features of the findings of realist evaluation together imparts a fourth and crucial characteristic – namely, that findings are always 'provisional'. Realist evaluation begins and ends with theory. It develops and tests theory about what works for whom in what context and in what respect. The tests may support the emerging theory on these inter-linkages but can not prove it. The realist approach is particularly keen that one evaluation should learn from another. A sequence of realist evaluations will lead to more powerful CMOCs, which are increasingly refined and better tested - but not ultimately proven. We can thus add to the desiderata for the presentation of findings, that they should:

- Synthesise the best evidence available without pretending that the evaluation or review has covered every conceivable programme process, output or outcome. Evaluators should not protest that 'more research is needed', but should always make clear the scope and boundaries of their research.

5.4 Utilisation: The payoff of realist evaluation

We reach the point where the evidence arrives at the policy-maker's desk. In common with all modern modes of evaluation, the vision would be of an arrival that was expected and welcome, because of thorough consultation on the progress of research from commissioning onwards. And the findings, moreover, will usually arrive containing presumptions and recommendations (both informal and formal) about how they may be put to use. So what should the policy makers expect of realist recommendations, and will they indeed be welcome?

As we have seen, realist evaluation produces results that are aimed at a relatively complex question – what works for whom in what circumstances and in what respects? The findings, if they are true to the method, are also likely to be 'configurational', 'middle-range', 'adjudicationist' and 'provisional'. Does this explanatory ensemble bode well for utilisation? The answer is, of course, both yes and no and the purpose of this section is to steer mutual aspirations to the former.

The school of theory-based evaluation, of which realist evaluation is an affiliate, has always described its appointed task as 'enlightenment' as opposed to 'political arithmetic' (Weiss and Bucuvalas 1980). The metaphor of enlightenment describes rather well the working relationship between research and policy (slow dawning - sometimes staccato, sometimes dormant, and sometimes antagonistic). Endless studies of research utilisation have described these chequered liaisons and the 'realistic' assumption remains that politics, in the last analysis, will always trump research. However, enlightenment's positive prospect, for which there is a great deal of empirical evidence, is that the influence of research on policy occurs through the medium of ideas rather than of data. Research is unlikely to produce the thumping 'fact' that changes the course of policy making. Rather, policies are born out of clash and compromise of ideas and the key to enlightenment is to insinuate research results into this reckoning.

On this score, realist evaluation has considerable advantages. Policy-makers may struggle with and find difficulty in using data revealing, for instance, the comparative statistical significance of an array of mediators and moderators in meta-analysis. They are more likely to be able to interpret and put utilise an explanation of why a programme mechanism works better in one context than another. These two research strategies are, as a matter of fact, serving to answer rather similar questions; the crucial point being that the one that focuses on sense-making has the advantage. This is especially so if the investigation has the task of checking out rival explanations (i.e. adjudication), which then provides justification for taking one course of action rather than another (i.e. politics). Here, then, is the positive message on research utilisation. Explanatory evaluations bring power to the decisions in decision-making.

Now, what of the bad news? An evaluation that asks a relatively complex question is bound to fetch up with a relatively complex answer. And if that answer is also 'configurational', 'middle-range', 'adjudicationist' and, horror of horrors, 'provisional' then things do not appear to bode well for utilisation, which is said to precede most smoothly if the advice is 'simple', 'immediate', 'unequivocal' and 'valid'. Once again, there is good evidence from the investigations of the 'enlightenment' school (Weiss, 1987) supporting this view that key policy-makers

spend precious little time on research. And, when consulted, the preference is for digests, executive summaries, potted histories and potted historians.

So can this circle be squared – is there some realist alchemy available to turn complex evidence into simple policy advice? In many cases and on many issues, the answer here is no. There will always be tension between certain policy questions and the answers that realist evaluation can supply and there is little use pretending otherwise – for a selection see the following box.

The policy question	The realist response
• Did that intervention work?	• It depends (in what respects?)
• Does that intervention work?	• It depends on the conditions.
• Does that programme work?	• Parts only, in some places and at some times
• Should we fund X rather than Y?	• Check first to see if they are commensurable
• Will it have a lasting effect?	• Unlikely, but you'd have to wait and see
• The pilot was great, should we go large?	• No, play only to its strengths
• Can you let us known before the next spending round?	• Sorry, not in all honesty

Whilst there are elements of reality in all the viewpoints depicted in this box of tensions, it paints an overly glum picture of the possibilities of utilising realist research. But as soon as one changes the nature of the query, then realism comes into its own. The alternative set of questions that makes best use of realist evaluation is listed in the following box. In this case we set down only the agenda, and not the response (which is delivered throughout the text).

Policy questions for realist research?

- What do we need to know in formulating programmes in this area?
- What are likely to be the key decisions in implementing it?
- What pointers can you give us in making these decisions?
- Would it work here?*
- Should the programme be targeted and if so how?
- Should the intervention be adapted to local needs?
- Are we likely to need to adapt the programme over time?
- How can we track the programme and keep it on track?

As an illustration of how the realist evaluator would tackle this vital question, we reproduce a useful grid in appendix 4.

So what should we expect a programme of theory-testing to reveal? What is enlightenment's content? Perhaps the best metaphor for the end-product is to imagine the research process as producing a sort of *'highway-code'* to programme building, alerting policy-makers to the problems that they might expect to confront and some of the safest measures to deal with these issues. An evaluation highway-code could never provide the level of prescription or proscription achieved in the real thing, the point of the parallel being that the highway-code does not tell you how to drive but how to survive the journey by knowing when, where and for what to keep your eyes peeled.

What the realist (theory-driven) approach initiates is a process of *'thinking though'* the tortuous pathways along which a successful programme has to travel. What is

described are the main series of decision points through which an initiative has proceeded and findings are put to use in alerting the policy community to the caveats and considerations that should inform those decisions. For each decision point, the realist evaluators should be able to proffer the following kind of advice: 'remember A', 'beware of B', 'take care of C', 'D can result in both E and F', 'Gs and Hs are likely to interpret I quite differently', 'if you try J make sure that K has also been considered'.

Programmes are theory incarnate. And in general terms and as a final summary, one can say that the realist viewpoint is that the most durable and practical recommendations that evaluators can offer come from research that begins with programme theory and ends with a refined programme theory.

6. Conclusion: strengths, limitations and relationships with other approaches

Realist evaluation has some clear strengths. It draws its foundations from the methodology of the natural sciences, and translates this into the world of policy and practice, with a view to bringing to that area of human endeavour the kinds of achievement that are manifest in the applied natural sciences. In taking its lead from the ways of the natural sciences, realistic evaluation stresses theory and the scope for generalisation that comes from attention to explanatory theory - generalisation that is critical in moving progressively from one programme experience to another.

The tendency to generalisation in realist evaluation invites attention to forms of underlying mechanism and forms of context across the substantive concerns of government and across departmental boundaries. Realist evaluation promises thereby to maximise learning across policy, practice and organisational boundaries. In steering evaluations towards transcendence of conventional divisions, it chimes well with discourse on partnership, cross-cutting issues and avoidance of silo thinking. With its insistence that context is critical and that agents interact with and adapt to policies and programmes, realist evaluation is sensitive to diversity and change in programme delivery and development. It provides a principled steer from failed one-size-fits-all ways of responding to problems.

Realist evaluation provides a coherent and consistent framework for the whole range of orders of engagement with programmes in which evaluation plays a part: the formative and summative and synthetic moments in the policy cycle. Finally realist evaluation is alive and alert to the importance of stakeholders to programme development and delivery. It steers a course between disregard for stakeholders on account of their self-interested biases and their craven treatment as omniscient and infallible on account of their inside knowledge. Stakeholders are treated as fallible experts whose understanding needs to be formalised and tested.

We have also to acknowledge important shortcomings in realist evaluation. It is intellectually enormously challenging. There can be no simple formula book that can provide tick-box recipes for delivering findings, any more than this is possible in research in the natural sciences. Realist evaluation requires sustained thinking and imagination to work through programme theory, to define expected outcome patterns,

and to figure out exactly what footprints or data signatures to look for and where to find them to test or arbitrate between theories. None of this is easy. It requires advanced theoretical understanding, abilities to design research and techniques for data analysis.

Even when undertaken well realist evaluation promises no certitude and it eschews efforts at being comprehensive. Again as in the natural sciences findings are tentative and fallible. And understanding relates to analytically defined mechanisms rather than to lumpy and disparate whole programmes. With its emphasis on contextual contingency and on temporal changes in the ways programmes are implemented and interacted with by their participants, realist evaluation is chary about serving up those stable net effects conclusions that are understandably beloved by economic modellers anxious to help wrest greatest utility from our precious and finite resources.

We turn finally to the fit between realist evaluation and other approaches. Realist review adopts an open-door policy. It can draw in and draw on studies using any of a wide range of research and evaluation approaches. This is not to say that studies are treated indiscriminately. Indeed they are raided for specific, realist purposes – for the potential they have to identify, test or arbitrate between promising context-mechanism-outcome pattern configuration hypotheses. Likewise, realist evaluation of initiatives has no particular predilection for privileging quantitative or qualitative methods. There is space for both. The choice of technique is subordinate to the theoretical task of working through expected programme outcome footprints however these may exhibited, in the light of mechanisms activated according to contextually relevant sub-group.

Realist evaluation is content to be pragmatic where data sources or resources are limited – to try to find out whether specific expected programme footprints can be identified in light of available data and data types. Generally, realist evaluation prefers, in ways that are now quite conventional, to combine quantitative and qualitative methods. Qualitative methods are often crucial to the elicitation of promising theory amongst programme architects and workers. Equally, they are often important in checking participants' means of interacting with programmes. These do not, nevertheless, exhaust sources of theory or active ingredients of programmes or sources of information on programme outcomes. Documents, official records of various kinds, observational material, survey-findings and so on can all find their legitimate place. In that sense realist evaluation is an inclusive approach.

Figure 1: evaluation as hypothesis testing

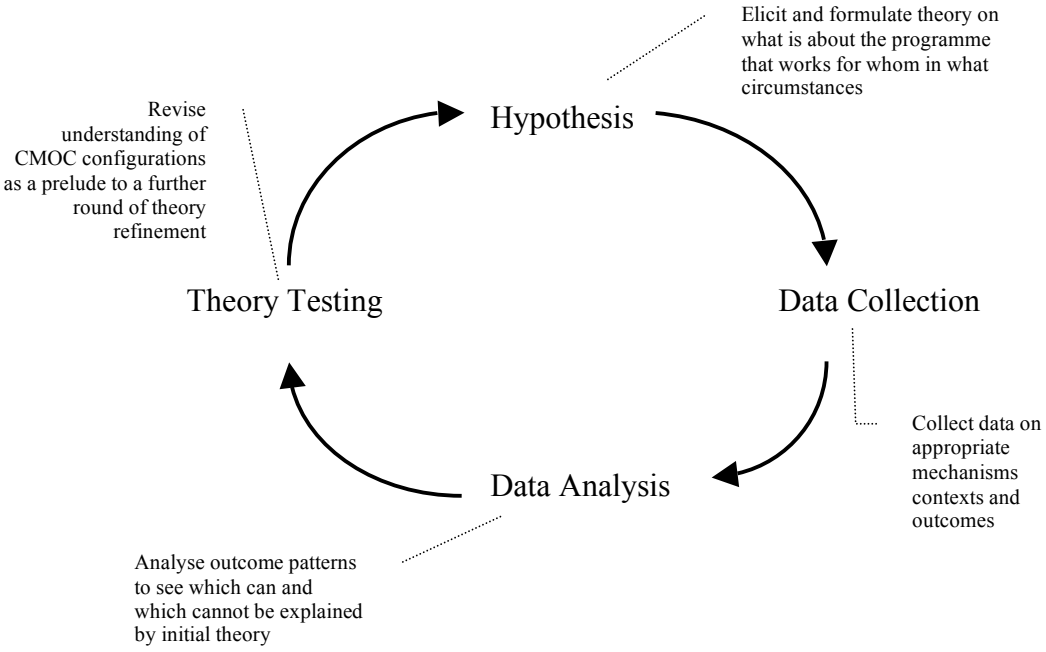
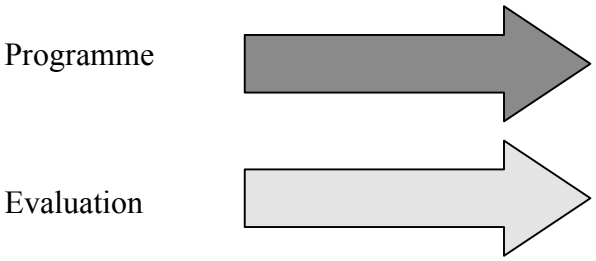
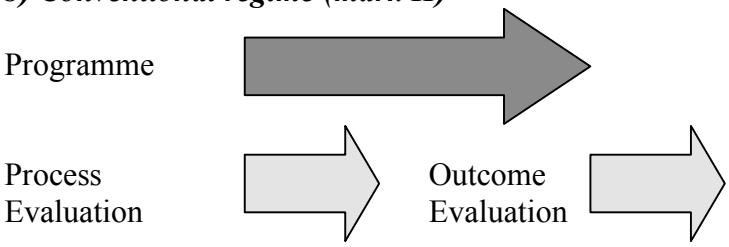


Figure 2: Matching evaluation(s) to programme theory

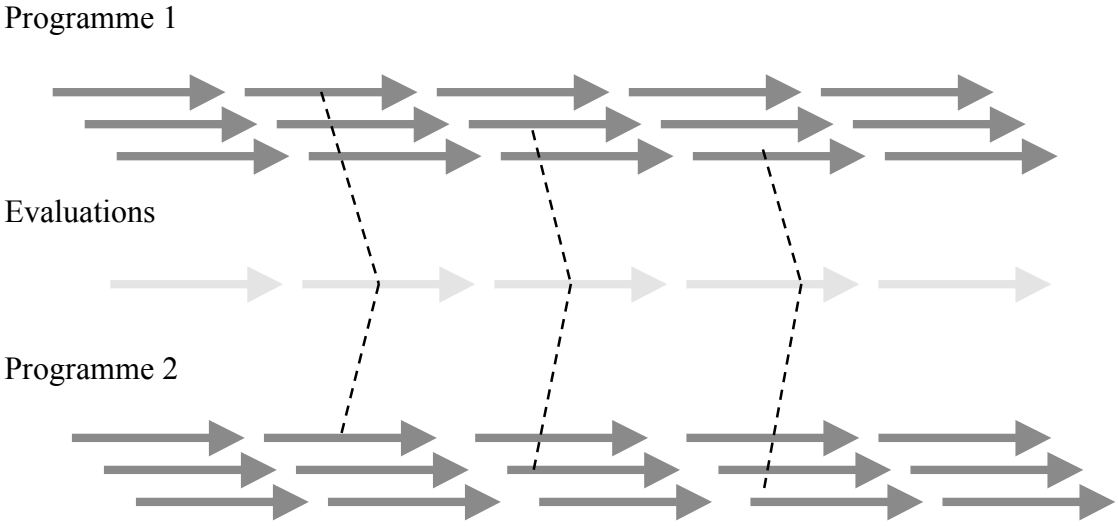
a) The conventional knowledge management regime



b) Conventional regime (mark II)



c) Realist regime



Appendix 1: 'Thinking it through'

Realist evaluation is a hypothesis driven approach and it is a truism to say that it can be only as good as the hypotheses that drive it. As is suggested in the main body of the text, these hypotheses can be derived from several sources including key stakeholders, the programme's administrative materials, its legislative framework, and social science literature in general. In addition to all of these, a touch of what Mills (1959) calls 'the sociological imagination' is also required.

In dozens of example in the main text, we have tried to tried to illustrate the breaking of programmes down into their component mechanisms (M_1, M_2, M_3, M_4, etc.), their surrounding contexts (C_1, C_2, C_3, C_4, etc.), and their potential outcomes (O_1, O_2, O_3, O_4, etc.). There follow two brief exercises for newcomers to help in thinking about *how* social interventions operate. The idea is to spend 10-15 minutes 'brainstorming' the following programme, in order to come up with some mechanisms through which it might work, and to highlight some differing contexts which might shape which mechanisms are activated, and thus to suggest an outcome pattern of potential successes and failures. You should be able to come up with a whole list of different reasons 'why' and 'for whom' and 'in what circumstances' such programmes might work and (just as important) might not work. Tax your imaginations and try to fill the hypothesis grid.

1. A Smoking Cessation Programme. Your task is to develop the hypotheses to evaluate a 'shock campaign' aimed at persuading habitual smokers that they *'should never give up on giving up'*. The campaign takes the form a series of hard-hitting, 20-second promotions shown in TV commercial breaks. Each one shows a smoker (in fixed frame, with a gradually closing focus) in a state of intense physical pain. Each patient attributes their terminal illness directly to smoking and each one acknowledges that they had always assumed it 'could never happen to them'. Each film ends with the slogan about it being never too late give up smoking and gives a help-line number to call. The final frame reveals that 'XY died n weeks after this interview was filmed.' The campaign is run intensively in the evening across all channels for two-months and then repeated one year later.

Think of some of the different potential reactions to this campaign. Think of what different viewers might think. Think of the different circumstances (smoking habits, health status, family, lifestyle, other health advice) that might prompt a different reaction. Think of what subjects might go on to do as a consequence of the message.

2. A Domestic Violence Programme. This time, your job is to formulate some testable conjectures around an initiative to address repeat domestic violence by arresting perpetrators even where the incident is relatively 'minor'. In this programme discretion is taken from the police officer called to incidents reporting domestic violence. In all cases the alleged perpetrator is arrested though they may not necessarily subsequently be charged.

Think, in this example, about variations in community setting, family composition, sub-culture, and economic and employment circumstances. Think too about

differences in the men and women encountering this programme. How might reactions vary in differing contexts? How might behaviour patterns then differ?

If interested in this case you might care to compare your theorising with that of others given the same task (Tilley 2000).

The realist hypothesis grid

Some plausible mechanisms	Some potential contexts	Some possible outcomes
M_1	C_1	O_1
M_2	C_2	O_2
M_3	C_3	O_3
M_4	C_4	O_4
M_5	C_5	O_5
M_7	C_6	O_6

Appendix 2: Varieties of realism: Four pocket illustrations of the approach

A. A quantitative realist analysis

This example is taken from Duguid's (2000) evaluation of a higher education course in Canadian prisons. The core of the statistical analysis is made up of a comparison between the predicted rate of recidivism of men who had undergone the programme and their actual rate. Readers are referred to the book for details of the definitions, measures, time-intervals involved in these calculations. It may be worth pointing out that reconviction predictors are commonplace in correctional services. The revolving door of re-entry to prison turns at different rates for different groups and so there is already a constant outcome pattern according to whether the offender is a school drop out from a broken home, has an addiction problem, committed violent offence, is of certain age, has a family home to return to after release, and so on. The reconviction predictor captures and weighs these factors providing a probability score for the recommittal/rehabilitation of any inmate.

Duguid's basic hypothesis is that undergoing the course will not be beneficial to all but will impact significantly on quite different groups of offenders. The table below (Duguid, p. 216) is but one of dozens and dozens of comparisons within the programme's subject group. It looks at what he calls 'hard cases', those prisoner-students with 'serious convictions', from 'broken homes', who had 'dropped out' of school. The table is subdivided into age categories, which identify further sub-groups according to their age of admission on the current conviction.

Sub-group by age	Predicted rehabilitation rate	Actual rehabilitation rate	Difference Gain under programme
16-21	43	69	+26
22-25	43	41	-2
26-30	40	41	+1
31-35	40	57	+18
36+	47	89	+42

There are indeed successes and failures unearthed in the analysis as can be seen in the comparisons of actual and expected non-return as above. These are all men deeply involved the dismal cycle of crime, yet the impact of the course on them is hugely different. But this is not just a simple tale of age demographics, because the standard imprint of age is already taken into account in the reconviction prediction. These subgroups thus mark out different contexts in which education might light a different spark. So why might the youngest of these toughest cases benefit significantly from the programme? Duguid posits a 'shelter' mechanism. It might be that the course offers an 'immediate second chance' - *before* these young men have to confront and become drawn into the macho culture of the wings and a continuing criminal career. The benefit for the two older groups might be due, by contrast, to 'last chance' and 'maturation' mechanisms. And for the group in the middle, it may be that education cannot penetrate a set of 'twenty-somethings', for whom criminal status might be might be *the* badge of honour recognised for survival on the inside.

It must be stressed that this snapshot of data reveals hypotheses *in development*. 'Shelter', 'maturation', the grabbing of 'second' and 'last chances' and so on are potential explanations for the above outcome pattern. These various claims are hardened and refined in the rest of Duguid's quantitative analysis in which he compares the fate of differently composed sub-groups, and also in qualitative work in which practitioners and subjects are able to voice their reactions to opportunities and constraints offered by an education-in-prison regime.

B. A qualitative realist analysis

This example is taken from Campbell and MacPhail's (2002) evaluation of a peer-education programme used as an HIV-prevention strategy in South Africa. In this instance, a well-developed programme theory is opened up for inspection using qualitative data on the mechanisms by which this particular intervention was delivered and the contexts that surrounded it. The programme theory starts with the idea that young people established norms about sexual conduct in a process of collective negotiation within group settings. Peer education settings thus might provide an ideal context in which they might come together to forge identities that might challenge existing relationships and behaviours that put their sexual health at risk. It takes a degree of consciousness raising and mutual empowerment to resist dominant gender and sexual norms - and the peer support context, rather than HIV-information alone, was thus a favoured vehicle.

Interviews and focus groups were held with both peer-educators and subjects, gathering information on group activities and the perceived challenges in mounting the intervention. These responses were then taped, translated and transcribed. Transcripts were then analysed using the NUDIST software for processing large qualitative data sets. Here the realist formula, *context* + *mechanism* = *outcome* was used was used to identify and code the process and circumstances that were deemed to influence the success (of failures) of the programme in achieving its desired outcome.

The overall finding is that the programme theory outlined above has little chance of being developed and sustained (and thus having an impact) in the particular circumstances of this South African trial. Just a few of the of the critical contexts uncovered are outlined here:

- *The highly regulated nature of the school environment.* Peer educators reported that it was difficult to overcome the tradition of didactic teaching and rote learning. Free discussion and argument had not hitherto been encouraged. Peer leaders would drift back into chalk and talk; students automatically raised hands to ask questions.
- *Teacher control of the programme.* The goal of empowerment requires that the school should act in an advisory and non-directive capacity. Peer educators reported that they fell under strict supervision of guidance teachers and principals, leading to disputes on the content of the programme.
- *Biomedical emphasis of the programme content.* There was relatively little focus on the social content of sexuality. Discussion was more likely to be focused on the transmission of the virus and on simple behavioural advice on resisting sexual advances.

- *Negative learner attitudes to the programme.* As levels of HIV have rocketed many South Africans have responded with high levels of denial. Accordingly, some potential students ridiculed the programme suggesting that peer leaders were themselves HIV positive.
- *Adult role models of sexual relationships.* Almost half of the informants had absent fathers and many referred to domestic violence in their households. Expectations about the quality of sexual relationships were not high.

The authors argue that, as it is presently implemented, this programme has little chance of overcoming a whole configuration of such personal, family, school and community constraints that mediate against safer sex relationships. They go on to suggest the need for further measures that might form a more comprehensive HIV-prevention strategy, capable of combating some of these deeply-rooted contextual impediments.

C. A formative realist evaluation

This example is taken from Clark and Goldstein's (2002) collaboration with Charlotte-Mecklenburg Police Department to address problems of theft from construction sites in their Charlie One district. Clarke and Goldstein worked with the local police to formulate a plausible strategy to identify a specific aspect of the problem where a preventive mechanism could plausibly be activated. Though not badged realist, the local strategy is realist in logic. Moreover its formulation accords with a realist approach and methodology in formative evaluation.

Of all the property crimes occurring at construction sites (which also include, for example, thefts of tools, materials, and plant), Clarke and Goldstein, alongside the local police officers and crime analyst, specifically identified theft of electrical goods that had been installed in yet-to-be occupied houses. They note that these accounted for 22% of commercial burglaries in the area. They also found that plug-in rather than hard wired appliances were most often taken, suggesting that opportunist rather than organised and determined thieves were at work. Clarke and Goldstein provide a realist account of the problem. In the context of new housing developments the disparate new residents were poorly placed and little motivated to watch over unoccupied new dwellings. Moreover with the comings and goings at a construction site they were unlikely to notice suspicious movements of goods. This created opportunities for thieves to take high value disposable electrical goods that could be readily sold. Hence there was a relatively high rate of their theft.

Clarke and Goldstein helped the local police think through which opportunity-reducing mechanism would be most promising in these circumstances. 'Reward removal' comprised the obvious choice. The postponement of the installation of plug-in electrical goods until the dwelling was occupied would deprive the prospective thief of the key reward and incentive to the burglary that had hitherto been available.

Lest such a conclusion might seem to the reader just too obvious to be worth stating, it should be emphasised that this was not the strategy being adopted by builders or promulgated by police. Indeed, less sharply focused, or less realistically thought through responses, had been tried but had not been found effective. These included,

for example, police and private security patrol, provision of tips and contact numbers to builders, establishment of watch schemes, efforts at targeted enforcement and so on.

Clarke and Goldstein catalogue some of the implementation issues that faced the preferred strategy: most importantly persuading building firms to alter their routine installation practices at some cost in terms of convenience. There was only partial success in this. Clarke and Goldstein then track the specific expected effects: the rates of loss amongst builders participating in the scheme, in particular in relation to the appliances whose risk of theft was reduced – those that plugged in. They also examine changed rates in the most plausible substitute offences to try to gauge whether any crimes saved have simply been switched to alternatives.

The outcome focused findings were that theft of targeted appliances amongst the 12 participating builders fell from 4.0 per 100 houses to 1.6 per 100, and of all appliances from 5.7 to 3.5 (631 houses in test year). Amongst the 47 non-participating builders the rate amongst targeted appliances fell from 4.1 to 1.8 and amongst all appliances from 5.1 to 2.0 (1131 houses in test year). Overall rates of target appliance burglary in Charlie 1 fell from 4.0 to 1.6 and for all appliances from 5.3 to 2.5 per 100 houses. There were slight increases in the rest of the Charlotte-Mecklenburg police districts. There was no evidence of displacement, but some of diffusion of benefits – the production of positive effects beyond the operational reach of measures brought about in this case by a generalised reduction in the expected benefit from theft.

D. A realist synthesis

This example is taken from Pawson's (2002) review of 'public disclosure' initiatives. Such interventions involve some aspect of 'under-performance' or 'failure' or 'misdemeanour' being drawn to the attention of a wider public, with the idea that wider pressures can be brought to bear in order to chasten miscreants and bring them into line. The underlying theory is a venerable one, with more than a sprinkle of scholarly overtones:

'Sunlight is the best of disinfectant: electric light the most efficient policeman' (Brandeis)
'A good name is better than precious ointment'. (Ecclesiasticus)

The theory is also a common one, such 'naming and shaming' schemes being brought to bear across all policy arenas. The efficacy of the programme theory is explored by examining its success in five different domains:

- The Car Theft Index revealing makes and models most easily stolen
- Newspaper publicity revealing those households defaulting on Poll-Tax payments
- Community notification of sex-offender (Megan's Law)
- Hospital rating systems (US Mortality Report Cards)
- Public notification of environmental damage (US Toxic Releases Inventory)

The synthesis used available evidence from a variety of primary sources. There is no space here to report on the whole gamut of outcomes, suffice to say they were typically mixed, with the 'adverse' publicity sometimes being countermanded, overlooked, ignored, resented, and even enjoyed in different measure in different

circumstances. The most successful intervention appeared to be the Car Theft Index (at least in its very first incarnation). One manufacturer performed particularly badly (seven ranges of cars were identified as 'high risk' in the '92 report and Ford made six of them) and subsequently made rapid steps to transform its security system.

It is important to make the conclusions of the review clear. The assertion is *not* that public disclosure works only to combat car crime and in no other domains (indeed in recent times the Car Theft Index seems to create little public interest and all new cars come well-protected electronically). The point of working though the five comparisons is that they reveal the contextual conditions that assist the operation of the disclosure mechanism and the public response. The conclusion is thus a theory about the conditions necessary to sustain the naming and shaming and sanction sequence. The paper hypothesises that CTI worked well in its 1992 outing because:

- the shamed party was an 'aspirational' insider, with a reputation to protect (C_1)
- the shaming mechanism could be dovetailed with 'market sanctions', with loss of sales bringing the company to heel as well as loss of reputation (C_2)
- the disclosure carried intense 'media interest', with the 'one key fits all' headline proving irresistible (C_3)
- the public data allowed for 'unambiguous' response, it being much easier to purchase a rival model than choose a different hospital (C_4)
- the shamer (the Home Office) had exemplary 'watchdog' credentials, which gave authority to the index (C_5)

The odd one or two of these characteristics applied in the other intended applications but they do not occur in concert. It is the configuration of characteristics that seems to be crucial. Note also that the theory produced here is transferable, the contexts are described at a middle level of abstraction and so can be examined as part of the evaluation of any future public disclosure intervention.

Appendix 3: 'Would it work here?'

This is one of the most crucial, and most difficult, questions confronting evaluation. It is a vital issue because it is what practitioners, especially, really want to know. Innovations often run in waves. Sometimes it is a matter of considering whether to follow a current programming fashion. Sometimes, the initial trials and pilots of a new initiative emerge with really promising results. At this point local practitioners and funders begin to ponder, 'would it work on our patch?'

Methodologically speaking, there isn't a more difficult question around. Repeating success is not just a question of imitating slavishly the day to day workings of a programme. What has to be considered is the entire CMO configuration of the intervention. Gomm (2000), writing in the realist tradition, has produced a useful checklist that policy makers and practitioners might use in thinking though (that phrase again!) whether to mount their own version of a thriving programme. It is adapted in the following table, with column A referring the established and successful programme, column B referring to the new potential location, and the third column asking some pertinent questions about the differences.

	System A	**System B**	**Desirability and/or feasibility of changing practice, procedures and context of system B to match those of system A**
The innovation	What are the salient features of the innovation as it is currently used in system A?	What are the salient features of the innovation as it is intended to be used in system B?	Where there is a mismatch, could and should the system B adopt the same innovation as is used by system A?
The resources	What resources were used in producing the outcomes (staff time, money, equipment, space, etc) in system A?	What resources are available to system B?	Has system B got the resources to emulate the practice of system A? If not, would it be feasible or desirable for system B to enhance or redeploy resources?
The people	What are the salient characteristics of the key actors in system A in terms of expertise, experience, commitment and so on?	What are the salient characteristics of the key actors in system B?	Insofar as there is a mismatch, would it be desirable or feasible to recruit different staff, invest in training, go through teambuilding activities etc?
Institutional factors	How far were the outcomes dependent on (for example) organisational / departmental structure, organisational culture, etc	How far does the organisational structure and/or culture of system B determine practice?	Insofar as there are differences, would it be feasible or desirable to change the institutional structures and/or cultures in system B?
Environmental factors	How far were the outcomes dependent on particular environmental factors (e.g. political, legislative, etc)?	How far is the external environment of system B comparable?	Insofar as there is a difference, would it be feasible or desirable to change the external environment of system B?
Measures	What baseline, process, outcome and other measures were used to evaluate success?	Does system B (or could it) use the same measures:	Would it be desirable or feasible for system B to change the way it measures and records practice?

Procedures	What exactly was done in system A that led to the outcomes reported?	Does system B do exactly the same (or could it)?	Insofar as there are differences, would it be desirable or feasible for system B to change what it does?
Outcomes	What were the key outcomes, for whom, at what cost, and what are they attributable to (see previous rows)? What was the cost per successful outcome?	What key outcomes are measured in system B? Are they achieved for the same actors as in system 1? What outcomes does system B achieve that system A does not? To what are these outcomes attributable? What is the cost per successful outcome in system B?	Insofar as the outcomes are different, to what are the differences attributable? Are there outcomes that system B is not achieving that it would be desirable for it to? Could system B achieve the same outcomes at a lower cost? Would system B have to forgo some current outcomes in order to achieve the same outcomes as system A?

References

Ackroyd S and Fleetwood S (2000) *Realist Perspectives on Management and Organisations* London: Routledge

Archer M (1995) *Realist Social Theory*, Cambridge: Cambridge University Press.

Bemelmans-Videc M-L, Rist R, and Vedung E (1997) *Carrots, Sticks, and Sermons: Policy Instruments and their Evaluation* Brunswick, NJ: Transaction

Campbell C and MacPhail C (2002) 'Peer Education, Gender and the Development of Critical Consciousness' *Social Science and Medicine* 55(2) pp. 331-345

Clarke, R. and Goldstein, H. (2002) 'Reducing theft at construction sites.' In N. Tilley (2002) Analysis for Crime Prevention. Monsey, NY: Criminal Justice Press.

Collier A. (1994) *Critical Realism* London: Verso.

Duguid S (2000) *Can Prisons Work?* Toronto: University of Toronto Press

Gill, M. (2003) *CCTV* Leicester: Perpetuity

Gomm R (2000) 'Would it work here?' in R Gomm (ed.) *Using Evidence in Health and Social Care* London: Sage

Greenwood, J. (1994) *Realism, Identity and Emotion* London: Sage.

Henry G., M. Mark and G. Julnes (1998) 'Realist Evaluation: An Emerging Theory in Support of Practice', *New Directions for Evaluation, 78*. San Francisco: Jossey-Bass.

Lawson, T. (1997) Economics and Reality, London: Routledge

Laycock, G. (1997) 'Operation Identification, or the Power of Publicity?' in R. Clarke (ed) *Situational Crime Prevention: Successful Case Studies* Guilderland NY: Harrow and Heston

Layder D (1998) *Sociological Practice* London: Sage

Mark M, Henry G and Julnes, G (2002) *Evaluation* San Francisco: Jossey-Bass

Merton R (1968, 3rd edition) *Social Theory and Social Structure*, New York: Free Press.

Mills C (1959) *The Sociological Imagination,* New York: OUP

Mitchell K (1997) 'Encouraging young women to exercise: can teenage magazines play a role?' *Health Education Journal* 56(2) pp. 264-273

Moore, B. (1966) *Social Origins of Dictatorship and Democracy*. New York: Peregrine.

Norrie A (1993) *Crime, Reason and History* London: Butterworths

Norris, C. and Armstrong, G. (1999) *The Maximum Surveillance Society: The Rise of CCTV* Oxford: Berg.

Painter, K. and Tilley, N. (1999) *Surveillance of Public Space: CCTV, Street Lighting and Crime Prevention* Monsey NY: Criminal Justice Press

Pawson R (2000) 'Middle-Range Realism' *Archives Européenes de Sociologie*, Vol. XLI pp. 283-325

Pawson R (2002) 'Evidence and Policy and Naming and Shaming' *Policy Studies* 23(3/4) pp. 211-230

Pawson R (2004 forthcoming) 'Simple Principles for The Evaluation of Complex Programmes' in M Kelly, A Kanaris, A Morgan, B Naidoo, E Barnett-Page, C Swann, G Powell, C Bannon, A Killoran and T Greenhalgh (eds) *An Evidence-Based Approach To Public Health and Tackling Health Inequalities: Practical Steps And Methodological Challenges* London: Sage

Pawson, R. and Tilley, N (1997) Realistic Evaluation. London: Sage.

Sayer A, (2000) *Realism and Social Science* London: Sage

Tilley, N. (1993) *Understanding Car Parks, Crime and CCTV* Crime Prevention Unit Paper 42, London: Home Office.
Tilley, N. (2000) 'Doing Realistic Evaluation of Criminal Justice', in V. Jupp, P. Davies and P. Francis (eds.) *Criminology in the Field: the Practice of Criminological Research* London, Sage.
Weiss C and Bucuvalas M (1980) *Social Science Research and Decision-making* Sage: Newbury Park
Weiss C (1987) 'The Circuitry of Enlightenment' *Knowledge: Creation, Diffusion, Utilization* 8 pp. 274-81