Evidence-based Policy

A Realist Perspective

Ray Pawson

SAGE

Evidence-Based Policy

A Realist Perspective

Ray Pawson

SAGE Publications

London ● Thousand Oaks ● New Delhi

Realist Methodology: The Building Blocks of Evidence

This chapter introduces the platform on which I wish to build. The basic argument is that, of the rival perspectives claiming to act as the foundation stone of social science, 'realism' provides the most comprehensive account of principles and practice, theory and method, promise and limitations. Given this pedigree, realism is solidly placed to supply a durable understanding of the process of cumulation of social scientific knowledge. Evidence-based policy seeks to stockpile the collective wisdom of thousands of pieces of applied research and can do no better than to look to realism for a methodology of synthesizing the available evidence.

Terminological prologue

Realist social science has a long and complex lineage and, with advancing age, has evolved significant family differences of its own. Before I embark on the task of identifying the key realist principles that should inform evidence-based policy, it is appropriate to locate my own interpretation within the wider spectrum of realist thought. This preamble offers a very brief glimpse of the family tree and points the reader to the precise bloodline.

Realism now figures strongly in the litany of competing perspectives and paradigms in modern social science, and examples of realist inquiry can be found in almost every sub-discipline, for example: law (Norrie, 1993), psychology (Greenwood, 1994), economics (Lawson, 1997), sociology (Layder, 1998), management studies (Ackroyd and Fleetwood, 2000), geography (Sayer, 2000), nursing (McEvoy and Richards, 2003), comparative historical studies (Steinmetz, 1998) and evaluative inquiry (Pawson and Tilley, 1997; Henry et al., 1998; Mark et al., 2000).

Bookmark

Realism is a methodological orientation, or a broad logic of inquiry that is grounded in the philosophy of science and social science (Bhaskar, 1978, 1979; Harré, 1978; Putnam and Conant, 1990; Collier, 1994). In these circles, realism is regarded as the principal post-positivist perspective, whose place is at the centre of things where it steers a path between empiricist and constructivist accounts of scientific explanation. It perceives social change to be neither linear nor haphazard but transformational (Archer, 1995). In terms of the practice of research, it favours neither the qualitative nor the quantitative (Sayer, 1992). It is 'neither nomothetic (that is law-seeking) nor ideographic (concerned with documenting the unique)' (Sayer, 2000). And some say that because it engages in neither abstracted empiricism nor grand theory, it is Mertonian in its preference for the middle range (Pawson, 2000).

At the core of all this far flung scholarship lies agreement on the basic apparatus of social scientific explanation. What makes this body of work realist is a common understanding of some very basic building blocks of social science, such as the nature of causation, the constitution of the social world, the stratification of social reality, the emergent nature of social change, and so on. All of these features will be explained in due course. It is sufficient here to emphasize that it is this explanatory apparatus I want to celebrate and hone for the purpose of conducting research synthesis.

There is, however, one schism on which realism itself divides, namely on the 'open system' nature of social explanation. Put simply, this says that social systems are the product of literally endless components and forces. When social science tries to focus on what seems a uniform pattern of behaviour it soon discovers that it is shaped by historical forces, with the result that it may occur in one culture but not the next. Secondly, institutional forces also play an inevitable part. These render behavioural patterns susceptible to change under different organizational arrangements and political structures. Thirdly, behavioural regularities are, of course, also influenced by the volition and choices of the people who act them out. A key aspect of these decisions is the human capacity to modify the uniformities in which behaviour is channelled. Collectively, our actions are always prone to change the conditions that prompt

them. On top of all this, even the research act itself is transformative; social research always has the tendency to disturb what it is trying to describe. A ceaselessly changing complexity is the norm in social life, and this is the open system predicament.

Grim news apparently follows. Such a complex and messy social reality appears to render extremely unlikely the opportunity for experimentally isolating and manipulating all the contributory explanatory elements. However one looks at it, the implication is that we can never exercise control over all the historical and contemporaneous, macro-and micro-conditions that have influenced the situation we wish to explain. All roads appear to lead away from 'closed systems' science and its key ability to isolate systems physically (as in laboratory experiments and machines).

The open systems dilemma has produced a variety of responses in social science. One has been to ignore it and attempt to approximate to experimental manipulation using randomized controlled trials as the lead investigatory strategy. Findings from artificially closed comparisons are then used to proffer advice on how to organize the incessantly shifting complexity of policy systems. I will return to the futility of this approach in the next chapter.

The consequence I wish to pursue here is that social science's struggle to imitate closed system investigation has created a dividing of the ways in realism itself, on the matter of whether social science should primarily be a critical exercise or an empirical science. On the one hand, there is the push towards 'critical realism'. The guiding assumptions here are that there will always be an overabundance of explanatory possibilities, that some of these will be mistaken, and that the primary task of social science is to be critical of the lay thought and actions that lie behind the false explanations (Archer et al., 1998; Bhaskar, 2002). This requires the social scientist to find a privileged standpoint from which to commence investigation, and ultimately draws realism into finding some moral high ground from which to sustain the critical edge (Edgley, 1998; Bhaskar, 2000). Much more could be said about the political and religious vantage points that have been chosen. All I am able do here is shake my head about how, in this normative

turn, the exacting understanding of the mechanics of scientific explanation produced by the pioneers of realism has evaporated into doctrinaire idealism.

Fortunately, there is another realist pathway in social science. This is the route taken by a somewhat looser amalgam of researchers who have tried to develop realism as an empirical method (Pawson, 1989; Hedström and Swedberg, 1998; Williams, 2000; Carter and New, 2004). To date, this second sect lacks a distinctive nomenclature, although labels like scientific realism, empirical realism, emergent realism, analytic realism, 'realismo pane e burro' and middle-range realism have been suggested (and found wanting). The guiding impulse is that it is still worth trying to adjudicate between alternative explanations even in the knowledge that further explanatory possibilities remain untapped in the unrelentingly open systems in which we live. What is more, it is assumed that much of the classic apparatus of empirical science – such as clear conceptualization and hypothesis-making, the usage of critical comparisons, the discovery of empirical patterns and the monitoring of their scope and extent – are of considerable use in this explanatory quest.

Remember that I am tracing genealogy here and so make only scant attempt to justify this preference at this point, although I will return to the practical ramifications of the open systems problem in Chapters 4 and 8. The background information on the two tribes of realism is proffered for two reasons. For newcomers it might lessen confusion, should they attempt to read across from one clan to the other. The second and crucial motive, however, is to assert that critical realism is not one jot of use to us here because its leap into the arms of the normative (Sayer, 2000) is precisely the political embrace from which evidence-based policy is trying to escape. Accordingly, the realism pursued here is the other sort, the one without the adjective. Fortunately, the lack of label presents no problem because this movement has had its clearest impact in the field of evaluation methodology in social policy (Pawson and Tilley, 1997; Henry et al., 1998; Mark et al., 2000). It is from this literature that the signature arguments of this book are drawn, which form the basis of its reformulation of systematic review and evidence-

based policy.

Before leaving the matter of realist terminology, here is a final, delicious twist. It may already be clear that I shall be casting some doubt on the approach to research synthesis of the Campbell Collaboration. This assemblage of scholars constitutes the orthodox paradigm in systematic review and meta-analysis and is named in honour of Donald T. Campbell, whose work was a blend of empirical inquiry, methodology of policy research and philosophy of science. And the name he gave to the principles that underlie this admixture? It was 'post-positivist critical realism' (Campbell, 1984).

To explain this fully would require a festschrift but, to summarize, Campbell's thinking was as follows. The 'post-positivist' element is not a problem. Campbell's adherence to realism was prompted by his recognition of the failure of logical positivist accounts of scientific discovery. He appreciated that scientific explanation went far beyond the painstaking measurement of facts and the steady accumulation of empirical generalizations. In particular, he emphasized the importance of theory, both in manufacturing the data and in explaining the observed regularities. Campbell thus sits between positivism (and the assumption that facts speak for themselves) and relativism (and the belief that we see what we want to see). He was a realist in that he accepted the existence of a reality which is independent of our senses, but which we can only discover through our senses.

It is the 'critical' element that causes the confusion. Fittingly, it turns out that Campbell is a critical realist in a quite, quite different sense from Bhaskar and his emancipatory colleagues. For Bhaskarians criticism is warranted on the basis of the analyst's privileged understanding of the oppressive aspects of the social condition and those responsible for it. For Campbell, criticism is something that scientists apply to each other, and this 'competitive cross-validation' is the means by which they get closer to the truth. His vision was of the community of scientists in constant, focused disputation, attending to each other's arguments and illustrations, mutually monitoring and 'keeping each other honest' until some working consensus emerged. The contrast in critical intent with the Bhaskarians could not be starker. It is, in short, that between

world of evidence-based policy.

The signature argument

Evidence-based policy is dominated by one question. Attend a conference, read a textbook, peruse a proposal, buy a tee-shirt on the said topic and somewhere in headline font appears the phrase – 'what works?'. This is a causal question. It is a challenge to bring together all the research on the effects that follow social interventions. The very purpose of interventions is to produce change and the mission of systematic review is to find out whether they do so. However, before diving headlong into the data it is worth pausing a moment to consider precisely what 'what works?' means. How do social programmes bring about their effects? How do interventions intervene? What is the nature of causality in the world of policies and programmes?

Realists have a ready-made answer to these questions. Indeed it is the signature argument. From its youthful first steps as a philosophy of science, through all the assorted disciplinary applications, and despite the bouts of infighting, realism stands foursquare behind the *generative model of causation*.

Its distinctive feature is to look for causal powers within the objects or agents or structures under investigation. If one asks why gunpowder has the capacity to explode, one would seek the causal explanation in terms of its chemical composition. If one asked why I have the power to examine PhD theses, one would look (hopefully) to my experience, qualifications and stock of knowledge. If one asked why a huge number of organizations took a bureaucratic turn in the mid-twentieth century, one would look to the advantages that flow from a division of labour, hierarchical structures, written rules, powers of surveillance and so on.

In each case there is a regularity involved (ignite powder and stand clear, expect viva with old timer, traditional authority gives way to bureaucratic regulation), but it is not the empirical uniformity that convinces us of the causal link. Indeed, our understanding of the causal linkage will survive even in the face of

some irregularity. So, for instance, gunpowder does not always ignite in the presence of a flame. The powder barrel may have become damp or the powder trail may be insufficiently compacted. Since PhDs have become ten-a-penny and the pay rate for examining them rather similar, you may now see a fresh-faced youngster conducting the oral. Bureaucracies are also known for their tendency to stifle innovation and to overcomplicate transactions, and so the iron cage turns out to have plastic bars as new organizational structures are sought to gain the competitive edge.

Although I have put the case in a rather merry way, a profound conclusion lies within the examples, namely this:

> Consequently, for realists, causation is *not* understood on the model of the regular success of events, and hence does not depend on finding them or searching for putative social laws. The conventional impulse to prove causation by gathering data on regularities, repeated occurrences, is therefore misguided: at best these might suggest where to look for candidates for causal mechanisms. What causes something to happen has nothing to do with the number of times we observe it happening. (Sayer, 2000:14)

Here, note well, is a preliminary shot across the bows of those who have sought to discover 'what works' on the basis of pooling observations in search of programmes with consistently powerful net effects. This remark is left as a censorious aside here because the immediate task is to complete the model of generative causal explanation. The basic trio of components is depicted in Figure 2.1 (Pawson, 1989; Pawson and Tilley, 1997).

Outcome patterns

Let us begin with O, the outcome pattern. Whilst it is not a sufficient base for establishing causality, there is little doubt that the sighting of regularities, uniformities and constants is what makes science perk up and take notice. Indeed, when it comes to the applied sciences like engineering the ultimate objective is to achieve

control over such regularities. As applied social science, evidence-based policy's mission is to choose an intervention on the basis that it has a reasonable chance of repeating successful outcomes achieved elsewhere. However, this is not achieved by the simple repetition of a winning formula. We know that there are no universal panaceas and no magic bullets in the world of social and public programmes. Everyone understands that what works in Dulwich might not go down so well in Darlington, still less in Detroit. The consequence is quite simple; in order to identify causal connections, we need to understand outcome *patterns* rather than seek outcome *regularities*. It is the totality of outcomes – successful, unsuccessful, bit of both – that may act as an initial empirical guide for future optimal locations.

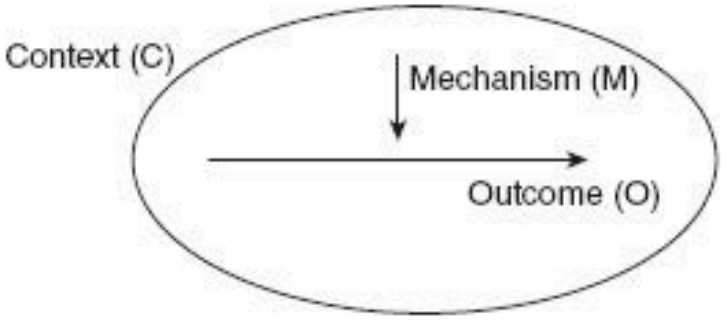

FIGURE 2.1 *Basic components of realist causal explanation*

Lawson (1997) provides a useful realist tag to describe these badly behaved uniformities, namely 'demi-regularities' (or demi-regs for short), which feature prominently in natural science explanation and subsequent innovation. Let us return to our plot about gunpowder for an example. The firing of early flintlock pistols was so haphazard it might be considered a demi-semi-reg. Reliability (regularity) was achieved only after considerable theoretical and empirical effort involving the idea of better encasing the powder to make it more stable, and controlling the spark to focus the ignition.

Even the so-called laws of physics are demi-regs in raw empirical investigation. If we take the gas law, it is certainly not the case that there is a linear increase in pressure every time the

temperature of a gas is increased. Try introducing hot air in a seminar room. Will its atmospheric pressure increase? Metaphorically maybe but not necessarily in a physical sense, because the room is not sealed and the law demands that it applies only to a fixed mass. Even under experimental conditions the law is not perfect. For instance, gases near liquefaction point do not obey the linear law. Natural science, however, is content to go along with the approximation, even to refer to it as a law, because it has a beautiful explanation awaiting (to be described in a moment) of why increasing temperature should cause pressure to rise.

The moral of the tale for evidence-based policy is that we should not be discouraged by demi-regs. On the contrary, the empirical evidence that welfare-to-work regimes work better for young men than young women, the data showing that many criminal justice interventions work well at first before their effects tail off, the huge disparity between the results of pay-for-performance schemes in the public and private sectors are all manna from heaven. They show that interventions work selectively. They are the beginnings of causal explanation.

Generative mechanisms

Mechanisms (M) are the engines of explanation in realist analysis. We can make rough sense of the world through its demi-regularities. The rhythms and associations of natural and even social systems are constant enough that we can navigate our way through them, although, as just argued, we are never particularly surprised when things don't work out as expected. We rely on mechanisms to tell us why interconnections should occur. A sequence of events or a pattern of behaviour are explained as being part of a system and the mechanism tells us what it is about that system that generates the uniformity. Mechanisms explain causal relations by describing the 'powers' inherent in a system, be those systems substances (like gases and gunpowder) or agents (like examiners or policy-makers) or structures (like bureaucracies or social programmes). In all cases it is something about the 'propensity' of the system that explains the causal regularities. The

mechanism explains what it is about the system that makes things happen.

In order to reinforce the point that this is a general model of explanation, let us continue with the same examples. Bureaucracies make things happen because of what they are. Because the workforce is organized in a hierarchy and because agreements are struck on responsibilities, work gets done in routinized ways. The structure generates the work pattern. Gunpowder has the tendency to go off with a bang because of what it is. Chemicals react and combine in different ways. Some combinations give off a large amount of energy under the application of heat (exothermic reactions) and it is this capacity that makes the mix of potassium nitrate, charcoal and sulphur so excitable. The chemical composition generates the capacity to explode.

Scientific knowledge begins to accumulate when the same generative mechanisms are used generically. The generic, generative mechanism that is used to explain the behaviour of gases is kinetic theory. From Bernoulli's time physicists have come to understand the properties of gases 'internally', as part of a system governed by the motion of a mass of microscopic particles in a confined space. Pressure is understood as the molecular force created on the walls of the container, and temperature is related to the extent of molecular motion (or kinetic energy). Using this mechanical model it is possible to calculate the precise linear relationship between temperature and pressure for a fixed mass and volume of gas. The maths are omitted here, but even the lay person can appreciate that if the application of heat makes the molecules go faster, the greater is the bombardment and thus the pressure on the container. It is this theory that has made the gas laws persuasive. The explanation has become even more compelling as those irritating demi-regularities have been brought into the fold. Near liquefaction, gas molecules are more tightly packed. Accordingly, kinetic theory makes an adjustment for a greater degree of intermolecular collisions and a lessening of impact on the container, and thus accounts for the observed departure from the linear law at low temperatures. Gases behave as they do because of what they are.

Moving to the example that concerns us here, we arrive at the premise that social programmes make things happen because of what they are. But what are they? What is it about them that makes a difference? What goes on within them to influence people to change? What are the underlying generative mechanisms? The precise causal levers will be explained in detail throughout the book, but in abstract terms one can say that programmes work only if people choose to make them work. At the broadest level of generality, one can say that programmes offer resources and whether they work depends on the reasoning of the subjects. The nature of the carrot of inducement may be different (material, social, cognitive) and the offer may include resource withdrawal (the stick). But whatever the intervention, it can only work as intended if the subjects go along with the programme theory and choose to use the resources as intended.

For the realist, then, causal explanation cannot begin without the identification of generative mechanisms. As far as evidence-based policy goes it means breaking with the lazy linguistic habit of supposing that it is programmes that work, and resting content with counting how often they work. The prerequisite is to look beneath the surface in order to inspect how they work. The development of cumulative knowledge about 'what works' requires sustained investigation of the generic mechanism, namely the operation of choices under the inducement of programme resources.

Contextual conditions

Context (C) is mechanism's partner concept in the realist understanding of causality. Causal relationships only occur when a generative mechanism comes into operation. Discovering the explanatory mechanism in action, however, is only half the battle because the association between its operation and the occurrence of the expected outcome is not fixed. Rather, outcome patterns are also contingent on context.

To take our trio of examples, gunpowder has the chemical composition to create exothermic reactions under an initial application of heat, but whether it does so depends on other

conditions such as the absence of damp and the presence of oxygen. Gas molecules have kinetic energy so that their movement creates uniform relations between properties such as pressure, volume and temperature. Irregularities and non-linearities begin to occur if the mass of gas is not fixed or the molecules are sufficiently compressed so that intermolecular forces become significant. I have the academic capacity to do a reasonable amount of external examining but this is not turned into a regularity in the absence of an adequate system of rewards, norms of collegiality, a supply of candidates with appealing topics and so on. Bureaucracies organize work routines in certain ways and provide tightly specified employment functions for their workforces, but whether any of this happens depends on the availability of work and, ultimately, on the overall economic health of a nation. The efficacy of bureaucratic management is also contingent on the type of work carried out. In sectors that thrive on innovation and entrepreneurial activity, the application of fixed duties to fixed roles is likely to flop.

The success of social programmes is, likewise, limited by contextual constraints. Interventions, by definition, are always inserted into pre-existing conditions. A mass of different contextual constraints lurks in wait for every programme and the interrelationships, institutions and structures in which it is embedded all shape its fortune. Despite the differences in such circumstances, it is possible to provide a general picture of how context works. It operates by constraining the choices of stakeholders in a programme. Programme subjects are always faced with a choice, but it is both a limited and a loaded one. They have different pre-given characteristics that leave some well disposed, and some badly disposed, to the programme theory. They enjoy different pre-existing relationships that leave some well placed and some ill placed to take up the opportunities provided by the intervention. They come to programmes with power, or a lack of it, which enables some to resist and some to embrace the ideas of the programme. There is always choice but it is never a matter of free will. Programmes are met with constrained choices, located in pre-existing conditions, and these, as well as the processes internal to the intervention, determine the balance of winners and losers.

Thanks to context, there will always be a footprint of programme success and failure, and this brings us back full circle to 'demi-regs'.

What this little realist tutorial on causality is designed to show is that understanding causal powers is an explanatory quest. Knowing how social programmes work involves tracing the limits on when and where they work, and this in turn conditions how, when and where to look for evidence. Interventions offer resources which trigger choice mechanisms (M), which are taken up selectively according to the characteristics and circumstances of subjects (C), resulting in a varied pattern of impact (O). These three locations are the key sources of evidence. In realist jargon the causal connections are established via 'context, mechanism, outcome configurations' (CMOCs). Although this is a clumsy term it does present a stark contrast with the successionist view, which prioritizes the search for outcome regularities. In the realist view, all three elements must be considered in order to address the master question, 'what works?' Put rather more concretely, it introduces a new bottom line. Evaluative research only really begins if it tackles the question of 'what works for whom in what circumstances?'.

Systematic review widens the evidential canvas in considering the impact of whole families of programmes in many applications, but it does not change the causal question. The crucial evidence is still to be found in terms of outcomes *and* mechanisms *and* contexts. This maxim provides the broad agenda for building a model of realist synthesis.

The basic agenda for research synthesis The nature of causality in social programmes is such that any synthesis of evidence on whether they work will need to investigate how they work. This requires unearthing information on mechanisms, contexts and outcomes. The central quest is to understand the conditions of programme efficacy and this will involve the synthesist in investigating for whom, in what circumstances, and in what respects a family of programmes works.

The basic anatomy of a social programme

This section of the chapter takes a much closer look at the workings of social programmes and interventions, following an obvious but profound working principle of good science, namely, that it should utilize methods that are suitable for and compatible with the subject matter under investigation. This simple rule applies with just as much force to secondary analysis and research synthesis as it does to undertaking primary inquiries. Having established an overall framework for understanding the causal powers of programmes, it is necessary to examine much more minutely how they are implemented, for this too will teach us where to look for evidence.

The explanation, once again, is that insufficient attention has been paid to the way that social programmes roll out before the evidence on them has been obliged to roll in to the machinery of a review. If there is a culprit, it may perhaps be careless thinking about the word 'intervention'. This is a useful catch-all term in that it captures a totality of activities subsumed across social and public policy but, in doing so, it conflates initiatives that are, ontologically speaking, quite separate. Take, for instance, health interventions. It is as clear as day that a clinical 'treatment' is not the same thing as a health care 'programme', which is not to be confused with health 'service delivery', which is a different animal from health 'policy'. There are also endless subdivisions within these categories, as when the focus of attention on, say, service delivery switches from 'innovation' to 'management' to 'regulation'. If the focus is broadened to include interventions in education, welfare, criminal justice and all the rest, one sees that the nature of intercession, the very subject matter of inquiry, is constantly and subtly in transformation.

The key methodological point is that the conventional techniques of systematic review and meta-analysis are much better developed for pooling research results from the clinical treatment end of this spectrum, and there is grave danger in assuming that they will have utility elsewhere. I delay pursuit of this critique until the next chapter in order to capture some of the essential features of interventions that fall outside the clinical treatment category. Seven

elements are identified, which arguably figure in most 'mainstream' social and public programmes.

Interventions are theories

This is the most fundamental realist claim about interventions. A more conventional perspective sees interventions in more tangible terms such as collections of resources, equipment and personnel but, for the realist, such resources are theories incarnate. Interventions are always based on a hypothesis that postulates 'If we deliver a programme in this way or we manage services like so, then it will bring about some improved outcome'. Such conjectures are grounded on assumptions about what gives rise to poor performance, inappropriate behaviour and so on, and then move to speculate 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. Improvements in patterns of behaviour, events or conditions are then generated, it is supposed, by bringing fresh inputs to that system in the hope of changing and re-balancing it.

Let us begin with a particularly entertaining example of an intervention hypothesis. Some health education theories explain the unhealthy life styles of adolescents by the undue influence of popular culture and the 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 (for example, sports stars) into the pages and onto the airwaves of the teen media. Such a conjecture, known amongst UK denizens of health education as the 'Dishy David Beckham theory', runs risks in both diagnosis and remedy. Suffice to say that the evidence confirms the popularity of poring over pictures of Beckham and friends in the teen magazines, but shows that as an activity it continues to exercise girls' minds rather than their bodies (Mitchell, 1997).

Agenda item 1 Broadly speaking, we should expect as a core task that reviews pick up, track and evaluate the theories that underlie

families of interventions.

Interventions are active

This proposition considers how interventions bring about change. The triggers of change in most interventions are ultimately located in the reasoning of those touched by the initiative, so that effects are generally produced by, and require the active engagement of, individuals. Take two dental health programmes: the fluoridation of water and publicity on the wisdom of brushing twice a day. The former is an example of a passive programme. It works whenever water is swallowed and happens to whole populations who are not required actively to engage with it. In the second example, however, 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. Equally, 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, management strategies, service delivery changes and so on. The inescapable fact that policy is delivered through active interventions to active subjects has profound implications for research methodology. In clinical trials, human volition is seen as a contaminator. The experimental propositions under test are about whether the treatment (and the treatment alone) is effective, and researchers will go to considerable lengths to protect this causal inference, including random allocation of subjects, the use of placebos and double-blinding. The aim is to remove any shred of human intentionality from the investigation of whether treatment brings about cure. Active programmes, by contrast, only work through the stakeholders' reasoning, and knowledge of that reasoning is integral to understanding their outcomes.

Agenda item 2 Broadly speaking, we should expect that, in tracking the successes and failures of interventions, reviews will find elements of the explanation in the reasoning and reactions of different stakeholders.

Intervention chains are long and thickly populated

Intervention theories have a long journey. They begin in the heads of policy architects, pass into the hands of practitioners and, sometimes, into the hearts and minds of subjects. According to the make-up of the initiative, different groups will be crucial to implementation. Sometimes the flow from management to staff (and through its different levels) will be the vital link; sometimes the participation of the general public will be the key interchange; almost always the reception of the theory by programme subjects will be of the essence. The critical upshot is that interventions carry not one, but several, theories. The success of an intervention thus depends on the cumulative success of the entire sequence of theories.

An example is the chain of reasoning that supports the registration and community notification programme for released sex offenders in the USA (Megan's Law). Decisions have to be made on who are the high risk cases, what information should be registered on them and over what periods, how and to whom their identities should be released, how to monitor and regulate their movement, how to control community reactions and encourage surveillance, and so on.

In each of these instances those responsible for the programme have ideas about 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). One official even reports that his county draws the line on the basis of 'looking at how far the offender has to travel to buy cigarettes'. I hope that a rather jovial observation on 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 on the combined efficacy of such theories.

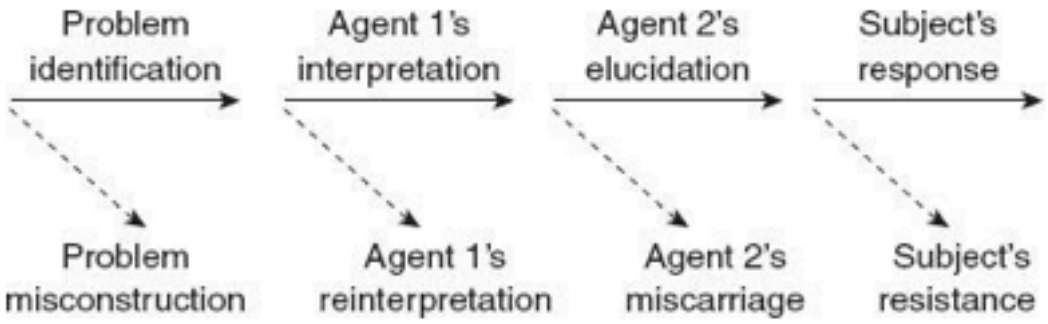

FIGURE 2.2 *Theory chains (with intended and unintended outcomes)*

All such theories are potentially fallible. The initial identification of the problem may be on the button or wide of the mark. As successive groups of stakeholders are summoned to forward the programme, they may sustain or undermine earlier implementation decisions, and their own decisions will induce support or opposition down the line. And, as we have seen, the final recipients (programme subjects) always have a measure of choice over whether to accept the concoction finally served up. In short, the intended sequence may misfire at any point, leading to unintended outcomes, as depicted in Figure 2.2.

Agenda Item 3 Broadly speaking, we should expect reviews to inspect the integrity of the implementation chain, examining which intermediate outputs need to be in place for successful outcomes to occur, as well as noting flows and blockages and points of contention.

Intervention chains are non-linear and sometimes go into reverse

So far, I have presented interventions as a series of decision points or programme theories being passed down an intervention chain, and Agenda Item 2 reminds us that each of these stages is active in that it depends for its effect on the recipients' response. This places a further onus on the evaluator or reviewer, namely to appreciate that such responses themselves have the power to shape and reshape the intervention, meaning that most intervention chains are non-linear.

There are several modes whereby a top-down intervention becomes, in some respects, bottom-up. The most obvious is the negotiation between stakeholders at every transaction within a scheme. Consider the application of performance measurement regimes such as school 'league tables' or hospital 'star ratings'. Quite customarily, one sees a struggle between professional associations and management authorities about the fairness of the indicators (on the need for risk-adjusted and value-added indicators, and so on). The indicators that finally come to be used take shape according to the punching power of the respective parties. This is depicted in Figure 2.3 by applying dotted, double heads to some of the arrows in a typical implementation chain.

FIGURE 2.3 *Negotiation and feedback in interventions*

A more marked inversion of an implementation scheme occurs if there is commitment to 'user involvement'. This is a popular notion in, for instance, urban regeneration initiatives in which much of the wisdom about renewal is considered to lie in the hands of members of the community (Stewart and Taylor, 1995). Many 'independent living' programmes in social care trade on a similar philosophy (Kestenbaum, 1996). What this produces is a feedback loop in implementation. Members of a community are consulted on the

optimal shape of the intervention. These theories are then thrust back up the chain of stakeholders so that they can amass the appropriate resources to put them into place. Once again, the actuality and viability of such adjustments depend on the respective powers of the agents and agencies involved. The feedback notion is also depicted in the dashed reverse arrow in Figure 2.3. Note that in reality there will probably be multiple stakeholder groups vying for influence at more than one point in the implementation chain.

Agenda Item 4 Broadly speaking, we should expect the review to be able to take into consideration how the relative positioning and influence of different decision-makers are able to direct and redirect programme implementation.

Interventions are embedded in multiple social systems

Thus far, interventions have been depicted as if they were populated only by individuals, and activated only through individual reasoning and behaviour. However, a critical feature of all programmes is that, as they are delivered, they 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. Interventions are fragile creatures. Rarely, if ever, is the 'same' programme equally effective in all circumstances. The standard requirement of realist inquiry, therefore, is to take heed of context and in the case of social programmes this means unravelling the different layers of social reality that make up and surround them.

Take, for example, a prisoner education programme, introduced with the Herculean goal of reducing reoffending. Such a proposal will carry a theory about how the intervention is assumed to work; for example, that adult education provides a second chance to regain a place in the job market, or that the cognitive skills acquired may allow for second thoughts in the face of opportunities for crime.

Of course, these theories about education and recidivism may be fundamentally flawed, as illustrated in the previous sections, but even if they are not the success of the educational policy will also depend critically on the setting in which it is introduced. The 'same' prisoner education package will unfold very differently in a young offenders' institute, maximum security penitentiary, local jail, vulnerable offenders' wing, open prison, day release scheme in a local further education college, and so on. And these variants, moreover, just begin to scratch the surface of contextual variation.

In general, realist analysis admits to the shaping influence of at least four contextual layers (the four Is):

- The *individual* capacities of the key actors: In the above example, do the educators have the appropriate motivations, capabilities and credibility to take the intervention forward? Do the prisoners have the corresponding characteristics and motives?
- The *interpersonal* relationships supporting the intervention: How intensively can a learning environment be created? Are the lines of communication between prison management, administration and custody staff supportive or damaging to the delivery of education by the teaching staff?
- The *institutional* setting: Does the culture, character and ethos of the prison support a rehabilitative thrust or is this overwhelmed by concerns with punishment, containment, warehousing, control, safety?
- The wider *infra-structural* system: Does the intervention have the political backing to drive it into the heart of the prison service? Are there welfare resources to underpin it? Is there public support for offering such opportunities and a second chance to released offenders? Will the criminal community override second thoughts?

All interventions are conditioned by the action of layer upon layer of contextual influences, and this state of play is depicted in Figure 2.4. Such contingencies represent the greatest challenge to

evidence-based policy. Generating transferable lessons about interventions will always be difficult because they are never embedded in the same structures and contexts.

Agenda Item 5 Broadly speaking, we should expect the 'same' intervention to meet with both success and failure (and all points in between), when applied in different contexts and settings. Whilst it is impossible to cover every angle (the open system predicament), a review presents a crucial opportunity to analyze the contextual differences operating across the primary systems investigated.

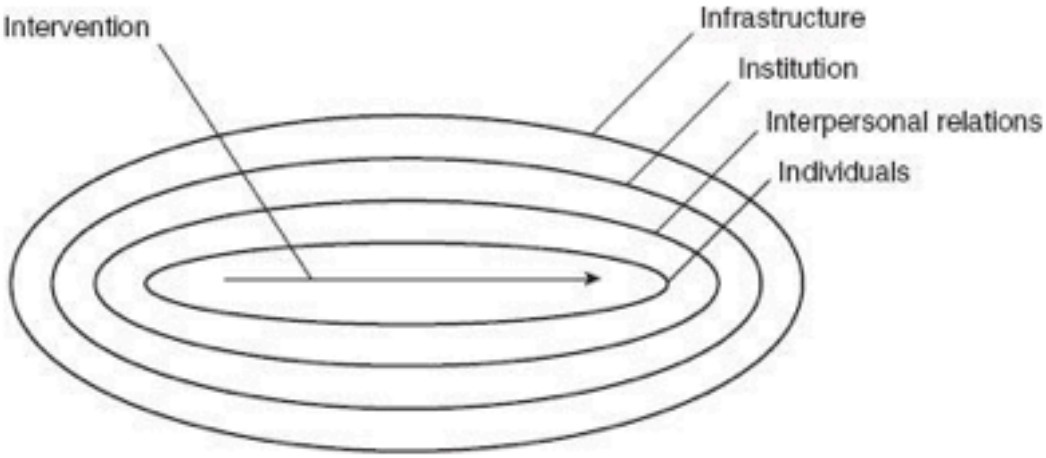

FIGURE 2.4 *The intervention as the product of its context*

Interventions are leaky and prone to be borrowed

One of the greatest bugbears of evaluation occurs when researchers perceive that the programme under inspection is changing within the grasp of their inquiry. The account thus far has stressed that interventions are sequences of theories, and that viewpoint makes plain why interventions always evolve in the course of research. A programme or service delivery reform might begin with something like an 'official' intervention theory and an 'expected' implementation chain. It will be put into practice in many different

locations and by many different hands, and in the course of implementation further programme theories will enter from outside the officially sanctioned process.

The reason for this is all too obvious. Practitioners and managers implement change and in the process of doing so they talk to each other. When a multi-site scheme is put in place there is always cross-fertilization and borrowing of ideas from other participants. When it comes to putting flesh on the bones of an intervention mission statement, practitioners will consult with colleagues. Especially when it comes to ironing out snags, there will be a considerable amount of rubbernecking from scheme to scheme as stakeholders compare notes on solutions. I have already given a rather dramatic account of some unfortunate consequences of practitioner information exchange in the first chapter with my New Deal for Communities practitioner suffering extreme cynicism brought on by a terrible case of 'interventionitis'.

In general, such a diffusion of ideas may be perfectly benign, with most programmes possessing a rummage bin of ideas that are drawn upon and adapted by a range of stakeholders. Such sideways chains of communication are actively encouraged in all programme-producing bureaucracies. In modern health services, for instance, large-scale innovations will generally be supported in national progress meetings, which encourage the sharing of tricks-of-the-trade as, for example, in quality improvement collaboratives, learning sets and quality circles (Øvretveit et al., 2002). The hallmark of professionalism is always said to be induction into secret knowledge (Eraut, 1994), and the professionalization of the public services quickens this circulation of tacit knowledge.

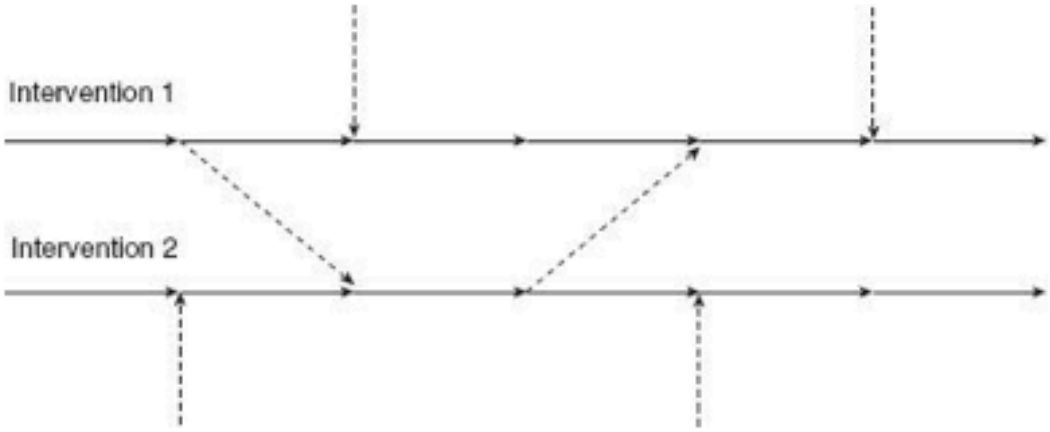

FIGURE 2.5 *Intervention facsimile and overlay*

The continual makeover of schemes under borrowing and imitation from others is illustrated by in Figure 2.5. The horizontal arrows show the intended unfolding of an intervention in two settings. The shaping, and sometimes distorting, forces of other schemes and services are illustrated by the vertical arrows (in the case of distant, external programmes) and by the diagonal arrows (in the case of cross-fertilization between local, companion sites). The result is that the overlay of formal and informal programme theories can become massively convoluted, especially if the service change in question is itself about promoting communication and collaboration!

The key point here is that tacit knowledge about a scheme may sometimes standardize it and may sometimes fragment it, but will always change it. The result is that reviewers must always beware what Øvretveit and Gustafson (2002) call 'label naiveté'. The intervention to be reviewed will carry a title and that title will speak to a general and abstract programme theory, but that conjecture will not be quite the one evaluators have encountered, or that reviewers will disinter.

Agenda Item 6 Broadly speaking, we should expect the implementation of programmes to be enwrapped in established expectations about how to deliver such schemes. Interventions

always have a history and a place within a wider range of policy decisions, and another potential agenda for a review is to consider how their success is shaped by previous/co-existing service delivery.

Interventions are open systems and change the conditions that make them work in the first place

As we have seen, interventions are active and absorptive, permeable and plastic. The result is that they are never implemented in quite the same way and never interpreted in quite 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 (Archer, 1995). Successful interventions may well change the conditions that made them work in the first place. Such 'morphogenesis' may be self-defeating or self-affirming.

The so-called arms race in crime reduction programmes is a prime example. Having suffered a setback from the introduction of a crime reduction scheme, lawbreakers are often able to figure out the intervention's *modus operandi* and adapt their own criminal *modus operandi* accordingly. A rather vivid example is provided by the changing impact of town centre CCTV cameras. On installation, these were regarded with some foreboding by marauding youth. However, once their positioning and range was understood, and when it was grasped that their impact depended on the deciphering of hazy images by a distant operator, and as soon as it was figured out that the police response could not be instant, a different set of options opened up. Norris et al. (1998: Part 4) noted the bizarre chicanery that followed. Familiarity bred contempt and, after the initial honeymoon period, the authors observed youths staging mock fights in front of city centre cameras. The combatants were plausible enough to prompt operator action and smart enough to become solid citizens at the sound of the sirens. The result here and across the crime prevention field is that schemes often become self-defeating, and a constant stream of fresh initiatives is required to keep pace.

Rarely are programme theories decoded and resisted to such

dramatic effect. There is, however, a more modest self-defeating effect in many interventions. On their first introduction, performance targets and progress reviews may lead to a significant period of self-reflection on the activities in question. If such monitoring becomes routinized, various short-cuts and tricks-of-the-trade may also follow, and the desired introspection can become perfunctory (see Evans, 2003 on this theme in relation to the National Health Service appraisal scheme for senior clinicians).

There are other conditions that lead interventions to become self-fulfilling, at least in the short and medium term. Familiarity may breed contempt *or* content. Management innovations tend to work if they curry favour with existing staff, and can be greatly assisted if recruitment and promotion of programme-friendly staff are also part of the package. Such a condition remains self-affirming only in so far as staff restructuring can keep pace with innovation in ideas. Otherwise, managers are faced with the self-defeating task of teaching new tricks to old dogs.

The post-conditioning of initial outputs by subsequent decisions is illustrated in Figure 2.6. The line of solid arrows represents the initial reception of a programme theory. The dashed arrows represent subsequent reinterpretation of the primary programme measure, with the untoward consequence pushing the intervention off course.

Agenda Item 7 Broadly speaking, in reviewing the collective impact of programmes, a further task is to chart the significance of familiarization and habituation, including the self-defeating and self-affirming effects that appear as programmes mature.

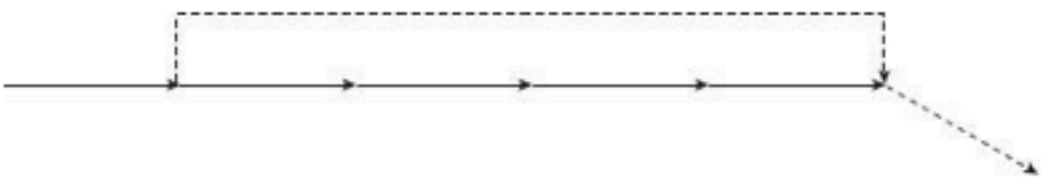

FIGURE 2.6 *Self-affirming and self-defeating change*

From the agenda and into action

The case for systematic review has been put most famously in Lipsey's compelling metaphor, 'what can you build with thousands of bricks?' (Lipsey, 1997). His answer was that it is high time to put aside solitary evaluations, which tend to come up with answers that range from the quick and dirty to the overdue and ambivalent. These can and should be replaced, he goes on to say, with the considered appraisal of the collective findings of dozens, hundreds and, just occasionally, thousands of primary studies to construct a solid citadel of evidence.

This chapter has attempted to give a sketch of the building blocks. What I have tried to show is that in order to be true to the nature of causal explanation and to be faithful to the character of social interventions, the evidence base must attempt to get to grips with social processes of extraordinary complexity. Thanks to the creativity of language (or put more unkindly, management-speak), innovations can seemingly be captured in a few words and may appear quite singular. In the realist view, social interventions are always *complex systems thrust amidst complex systems*.

This message about the intricacy and convolution of programme delivery will come as no surprise to key stakeholders. I am confident that there will be a sense of recognition on the part of policy-makers, managers and practitioners that the processes and structures described above are routine features of most interventions. If anything, those on the streets and in the hospital wards and classrooms are likely to perceive an even more messy and animated process. Negotiation, leakage, borrowing, resistance, mismatch, adjustment, bloom and fade and so on are part and parcel of everyday programme implementation. This core condition of programmes is summarized in Figure 2.7, which brings together all the propositions and diagrams in this chapter. It depicts the passage of four interventions, which are nominally the same, but as a result of variation in contextual locations and policy-maker, practitioner and subject interpretations they begin to bend. Thanks to user feedback and negotiation, the exchange of know-how, familiarization and over-familiarization with due process, they

begin to twist over time, generating further shifts and unconsidered effects.

This make-up of programmes places profound limitations on what can be expected from evidence synthesis. The reviewer will always be confronted by an array of programme theories played out in varying contexts and implemented in different ways. It goes without saying that reviewing the effectiveness of such systems-within-systems will be a battle with complexity, and this chapter concludes by laying out the broad parameters of that challenge. One thing that is evident from the start of the endeavour is that the evidential bricks will not be uniform. Data on all the constituent mechanisms, contexts and outcomes will not come in standardized slabs, reducible to a net effect score. Potentially, the entire apparatus of social science will be needed to examine processes at the individual, interpersonal, institutional and infrastructural levels. Three significant limitations can thus be anticipated from the outset.

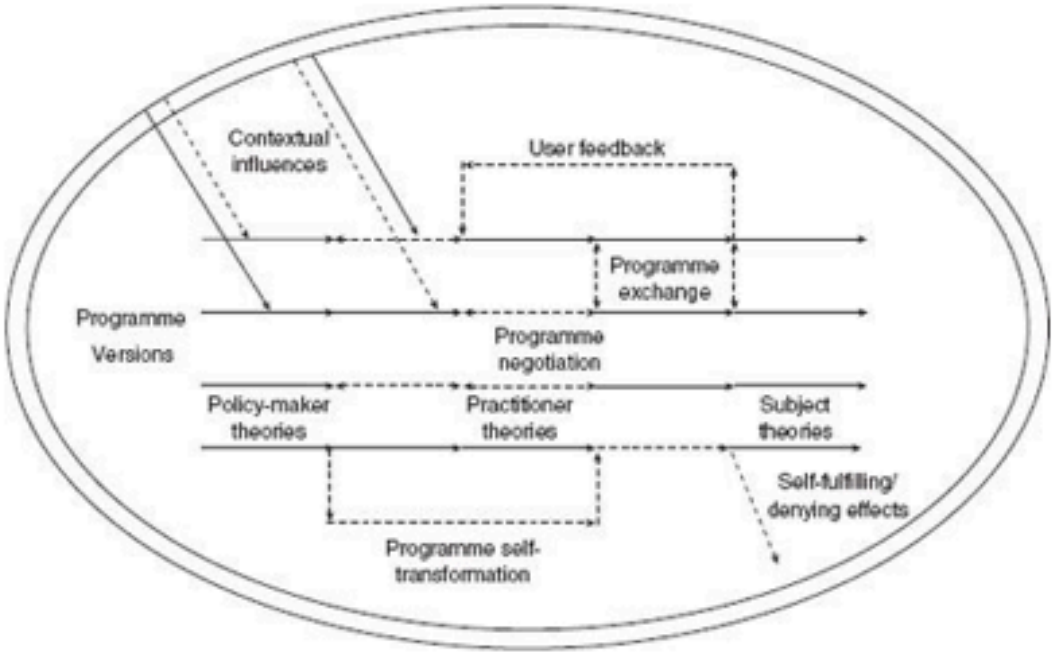

FIGURE 2.7 *Programme complexity*

First, there is a limit on what a review can cover. The realist

approach shifts the unit of analysis from programmes to programme theories. Quite properly, this gets us away from the vision of social programmes as treatments or dosages, which themselves have effects that can be pooled and averaged. Recognizing programme theories as the engine of change, however, does have a drawback because these working ideas are endless. At the limit, there will be mundane links in a particular implementation chain about 'employing Eric to do the IT because he's a reliable sort'. Although the programme may well suffer if this conjecture is wrong, the moral of the story is that realist synthesis will need to prioritize the investigation of particular processes and theories.

Secondly, there is a limit on what information can be retrieved. Primary research studies will probably have focused on formal documentation (such as policies, guidance, minutes of meetings), tangible processes (such as the activities of steering groups), and easily measured outcomes (such as attendance figures or reported opinions). Information about the informal (and sometimes overtly off the record) exchange of knowledge, the interpersonal relationships and power struggles, and the subtle contextual conditions that can make interventions float or sink in an organization will be much harder to come by, and is often frustratingly absent from reports. Even if the review is committed to tracking some back-stage mechanism or unintended process, it may not always be possible to do so.

Thirdly, there is a limit on what a review can deliver. The reviewer can never grasp the totality of the constraints on the effectiveness of interventions, and certainly cannot anticipate the circumstances in which subsequent schemes might be implemented. This places ineluctable limitations on the recommendations that follow a review and the certainty with which they can be put forward. A necessarily selective and prioritized review will generate qualified and provisional findings, and thus modest and cautious recommendations. Realist reviewers eschew the goal of discovering best buys and delivering verdicts. Rather, they attempt to place on the table an account of the workings of complex interventions and a hopefully better understanding of how

theory may be improved.

As an overall conclusion I return to my pet irritation about the use of the term 'intervention' to describe the lumbering pachyderms of modern social and public policy. Policy masterplans always end up as elephantine in their complexity. In response, I have argued that research synthesis needs to work with a more powerful understanding of causality and an appreciation of at least seven different ways in which programmes may be said to 'work'. Already, it is clear that it will not be possible to review them all, and since selection and prioritization will be part of the analysis, there must always be caution and diffidence in deriving any conclusions and policy recommendations. If evidence-based policy were human, he would be a modest soul with much to be modest about.