

4

REALIST SYNTHESIS: NEW PROTOCOLS FOR SYSTEMATIC REVIEW

This chapter sets out the method of realist synthesis. It is the pivot of book. In the first chapter I voiced the rationale for attempting to synthesize volumes of previous research and gave a glimpse of the daunting challenge involved in doing so. In the second, I introduced realism as a general logic of social science, as a tool for understanding how social programmes work, and as a framework for understanding their complexity. In the third, I followed the arduous route march of systematic review and found fault with the conventional way of making the journey from research evidence to policy recommendations. Now I want to apply the general principles of realist explanation and the lessons about programme complexity to the business of research synthesis.

The new model for systematic review is constructed in two ways. First, I pay attention to domain principles, establishing a totally revised model in which theory-building takes pride of place as the prime activity in research synthesis. Secondly, I provide a blueprint for realist synthesis in way that captures the step-by-step process of searching, tracing, filtering, eviscerating and analyzing all of the primary materials. All of the ‘counterpoints’ raised in the previous chapter are activated. Overall, this chapter aims to establish a new protocol for conducting a review in a manner that can be followed by anyone intent on conducting a realist synthesis. Much changes except, that is, for the hard slog.

First principles: synthesis as theory-building

Let us go back to square one. What are we actually doing in the protracted process of gathering together voluminous piles of primary research and rendering them down into policy advice? What is the essence of the activity? What does it mean to synthesize research? One can take a cue on this matter from other applications of the term synthesis, which in its various uses in chemistry, philosophy and linguistics suggests something rather more active and creative than combination or aggregation. In these formulations, the product emerging from synthesis is reckoned to be more than the sum of its parts. Some knowledge gain, some novel compound, some added value is produced in the process of synthesis. Such a notion is also vital to evidence-based policy. There is a need for systematic review to go beyond reportage and summary of an existing state of affairs. The point, after all, is to support fresh thinking to revise policy and launch it in new circumstances.

In the realist model, the primary ambition of research synthesis is explanation-building. The purpose is to articulate underlying programme theories and then to interrogate the existing evidence to find out whether and where these theories are pertinent and productive. Primary research is examined for its contribution to the developing theory. The overall intention is to create an abstract model of how and why programmes work, which then can be used to provide advice on the

implementation and targeting of any novel incarnation of the intervention. The process is illustrated schematically in Figure 4.1.

FIGURE 4.1 ABOUT HERE

Realist synthesis begins by identifying its subject matter – normally taken to be the class of programmes about to be reviewed – but the real work begins with the construction an embryonic theory of how they may work. This may be articulated in a variety of ways but (as illustrated in the top section of the figure) will contain some of the familiar feature of realist explanation, namely conjectures on the generative mechanisms that change behaviour, ideas on the contexts which influence its operation, and hunches about the different outcome patterns that ensue. In other words, the preliminary hypotheses will raise speculative ideas about ‘what works for whom in what circumstances and in what respects’.

The synthesis then moves on to the primary data. These are regarded as case studies, whose purpose is to test, revise and refine the preliminary theory. It is expected that they will reveal a mix of methods, a mix of information and, above all, mixed messages about the success of the intervention. The process is illustrated schematically in terms of the six primary inquiries in the middle section of the figure. The initial theory provides a lens through which to view the studies, and one anticipates a spectrum of different refractions. The research may offer direct support or outright contradiction to the preliminary theory, although a subtle modification in understanding is more likely. The evidence may improve or weaken or modify or ameliorate or revise or supplement or refocus the underlying programme ideas, and the purpose of synthesis is to make sense of these quite different challenges.

Let us now consider some of the different ways in which primary evidence might confront the review hypothesis, in order to see how the various empirical challenges might be taken on board. Some studies will be relatively revealing about underlying mechanisms, some will concentrate on outcomes, others may describe context in depth (illustrated by the presences and absences of these features in cases 1 and 2). Here, the review is concerned with *juxtaposing* the evidence as, for instance, when one study provides the process data to make sense of the outcome pattern noted in another.

Contradiction between two primary studies (illustrated by the positive outcome in study 3 and the failed (hence dashed) outcome in study 4) is also a common occurrence. In this case the analysis may attempt to *reconcile* them by unearthing contextual or implementation differences in the original programmes and showing how these led to opposing outcomes. An alternative in this situation is for the reviewer to *adjudicate* between the inquiries on the basis of a methodological excavation that reveals the relative virtues or shortcomings of the original conclusions.

In another familiar story, studies may arrive at different accounts of programme impact (illustrated by two successful outcomes in cases 5 and 6, which are explained as the product of different mechanisms and circumstances). In this case the reviewer might attempt to *consolidate* the results into a multifaceted explanation for programme success, and the synthesis might make the case for a combined approach.

Alternatively, the reviewer may attempt to *situate* the rival explanations, arguing that the programme operates successfully through mechanism 5 in context 5 but that mechanism 6 might come into its own in contexts like 6. Both explanations survive as a result of understanding and reinterpreting their scope conditions.

The review continues in this way through many more primary inquires, each one leading to an adaptation of the original review hypothesis (compare Layder, 1998 on adaptive theory). The finale is not an arithmetic verdict on a family of programmes but a refinement of its underlying theory. Figure 4.1 thus ends schematically by drawing together the exploration of the successes and failures of a programme in an abstract model explaining how efficacy (O_N) varies, depending on the particular *configuration* of its constituent mechanisms (M_N) and contexts (C_N).

The vital point to emphasize is that the analytic processes described above as juxtaposing, adjudicating, reconciling, consolidating and situating the evidence are none other than the elusive idea of synthesis. It is these reasoning processes that embody the logic of realist synthesis. It is this ability to think through studies and to make sense of their uniformities and discrepancies that embodies the real craft of systematic review.

The notion of synthesis as explanation-building has another key facet (and potential advantage), embodied in the role that *abstraction* plays in policy formation and evaluation. In devising and evaluating interventions, simple distinctions between a ‘programme’ its ‘implementation’ its ‘outputs’ and its ‘outcomes’ are made routinely. Similarly, the targets for intervention are often differentiated according to whether they are focused on the ‘individual’ or the ‘area’. Likewise, interventions are described habitually in terms of the exchange between ‘stakeholders’, who comprise ‘policy-makers’, ‘practitioners’ and ‘programme subjects’. These simple concepts can be applied to all programmes and not just the concrete one sitting before the evaluator, the point being that we perpetually carve up and define, navigate and make sense of interventions by thinking of them in terms of such abstractions.

In much systematic review, this basic vocabulary of intervention terms is taken for granted; the terms are treated as the given parameters of the investigation rather than as objects of investigation. It is simply assumed that the review will track this type of ‘programme’ aimed at this type of ‘subject’ seeking that type of ‘outcome’. However, the precise coinage of such abstractions is of immense significance in determining the appropriate evidence base. Consider, for instance, the importance of defining the ‘subjects’ in a cognitive-skills programme designed to assist in offender rehabilitation. The term ‘offender’ is itself an abstraction and the perceived problem – a lack of thinking skills – might well, on deeper consideration, be thought to differ according to whether the inmate is in for criminal damage or embezzlement or aggravated vehicle taking or drug dealing or rape or fraud. The problem, to put it glibly, is that some of them might be just too damn clever to begin with. The programme theory (and thus the programme) will succeed or fail according to how accurately the notion of the ‘cognitively deficient offender’ is drawn. In general, policy wisdom depends on hitting the right level of abstraction and paying meticulous attention to the boundaries of explanation.

The overall importance of careful abstraction for realist enquiry is thus established:

Social systems are always open and usually complex and messy. Unlike some of the natural sciences, we cannot isolate out these components and examine them under controlled conditions. We therefore have to rely on abstraction and careful conceptualisation, on attempting to abstract out the various components or influences in our heads, and only when we have done this and considered how they combine and interact can we expect to return to the concrete, many-sided object and make sense of it. Sayer (2000: 19)

Sense-making, our ultimate objective, relies heavily on vigilant conceptualization and watchful abstraction, and these tasks are a central objective of realist synthesis. Instead of assuming that the review is about a given programme X, acting on given subjects Y, with given objectives Z, the central intention is to question the reach and scope of each of these components. If one begins research synthesis from this standpoint, there are quite different opportunities to draw collective lessons about why programmes work. Let me pursue the most radical departure that can ensue by starting with programme concepts rather than programmes per se. A curious point about programme theories is their shared agenda, a feature noted decades ago by Salaman in urging the following change of emphasis:

Rather than focusing on individual programs, as is now done, or even collections of programs grouped according to major 'purpose' as is frequently proposed, the suggestion here is that we should concentrate on the *generic tools of government action* that come to be used, in varying combinations in particular public programs. (1981: 256)

We tend to think of social and public policy as comprising a million different initiatives. We have welfare-to-work programmes and dietary improvement programmes and head-start programmes and offender rehabilitation programmes and quit smoking programmes and peer mentoring programmes, and so on ad infinitum. Instead, Salaman argues that we should appreciate that these initiatives share common programme theories. Examined closely, it becomes apparent that interventions in quite different policy domains are expected to operate through the same or very similar programme mechanisms. The bravest rendition of this brave idea is the book by Bemelmans-Videc and colleagues (2003), which argues that, if one scrapes away programmes to their elemental bones, there are only three types of mechanisms on offer, namely 'carrots, sticks and sermons'.

Take, for instance, 'carrot theory' or the idea of using incentives. These have a time-honoured place in public policy and come in a whole raft of different grants, bonuses, subsidies, payments, loans, give-aways, premiums and so forth. Subjects have been offered incentives by way of subsidized nicotine replacement patches to quit smoking; free distribution of smoke alarms to increase household safety; transitional payments to keep released prisoners from reoffending; cash payments to donate blood; bike loan schemes to reduce inner city congestion; price reductions to induce early payment of service charges; grants to improve insulation to save energy; subsidies for low-income families to reduce the cost of attending university, and so on. Since incentives are at the root of so many programmes, policy-makers on the point of introducing a fresh one may have much to learn about its chances by trawling through the evidence on these many and varied schemes.

📖 Supplementary reading. Readers interested in further explorations of carrot theory should consult: R. Pawson *Evidence based policy: the promise of realist synthesis* and R. Pawson and N. Tilley *Realist evaluation bloodlines*, which are included in the web page support materials. 🌟

The key methodological principle, however, lies beyond this specific example. The crucial point is that as soon as one begins to explore policy-making at different levels of abstraction, and as soon one allows elements of programme theory to be the unit of analysis, then the cases and comparisons that are useful in the review change markedly. No longer is it a matter of chasing just the primary research that bears the title of a particular brand of intervention. Rather, the significant evidence is found by locating research that explores the precise set of assumptions held in the programme theory. Later in the chapter, we shall see how realist synthesis operates with a mixture of purposive samples of the primary data in order to pry open the evidence in relation to particular features of programme theory. For now, and for the purposes of establishing first principles, we arrive at the central claim of the book that:

Research synthesis operates through processes of policy abstraction and theory building rather than data extraction and number crunching.

Rethinking the standard template: practical steps in realist review

How can these markedly different principles be transformed into a work plan for conducting a review? This section develops the realist template by way of a step-by-step contrast with the standard review strategy that was criticized in the previous chapter. For easy reference, the standard review sequence laid out in Chapter 3 (Box 3.1) is reproduced again in miniature as Table 4.1. Realist synthesis also moves from bygone evidence to present-day policy advice and thus has a comparable structure. However, there are many subtle differences in emphasis, in duration, in running order and, above all, in the methodological content of each stage. This section will absorb the critiques, develop the realist alternatives, and put an alternative prospectus in place. This is summarized in Figure 4.3. It might be useful for readers to locate this diagram at the very end of the chapter; a quick preview will reveal the calm that follows the storm.

TABLE 4.1 ABOUT HERE

Despite the differences, there is one area of methodological consensus between the traditional systematic review and the realist review, and that is the need for transparency or auditability in the review process. The purpose of the basic template is to set out a clear design for new reviews and novice reviewers to follow. But it also allows others – researchers, decision-makers and other stakeholders – to ‘look behind’ an existing review, to assure themselves of its rigour and of the validity, reliability and verifiability of its findings and conclusions. Realist synthesis involves different kinds of hypotheses and methodological decisions but it is equally important that the thread of explanation building is kept visible as possible so that others can understand (and question) the roots of the policy recommendations. This perturbing principle is revisited through the following sequence and again in Chapter 8.

Stage one: identifying the review question

All reviews commence with an exercise in conceptual sharpening, attempting to refine and define precisely the question to be pursued in the forthcoming synthesis. In meta-analytic mode this involves paying close attention to defining the treatment under consideration, identifying the relevant population to which it is applied, and stipulating the outcome of interest. These decisions are enshrined in operational definitions within the review protocol, and used in the subsequent extraction of data from primary studies. The previous chapter showed how such operational closure leads to explanatory impoverishment, so a new approach to agenda-setting is needed.

The realist approach, too, must start with a sharpening of the question to be posed but this task goes well beyond the clarification of key terms. The divergence stems from the nature of the interventions studied (complex processes inserted into complex structures) and the very purpose of the review (explanatory rather than summative). These differences bite enormously hard at stage one of a realist review, and effectively break it into several sub-stages. Both reviewers and commissioners should anticipate that ‘focusing the question’ will be a time consuming and ongoing task, often continuing to the half way mark and even beyond in a rapid review. One author (Greenhalgh et al., 2004) has referred to this stage of the synthesis of evidence on complex programmes as ‘the swamp’. Anyone anticipating the quick fix of a watertight review question in an explanatory review should beware that they will very soon find themselves neck-deep in alternatives.

A realist review cannot comply with a commission to discover ‘whether’ an intervention works, but trades instead on its ability to discover ‘why’, ‘when’ and ‘how’ it might succeed. However, this explanatory orientation is not a single point of reference and will automatically generate a whole range of potential sub-questions. The best way to grasp this point is to refer back to the anatomy of programmes provided in Chapter 2 and summarized in Figure 2.7. There, in seven separate items, an agenda for understanding programmes was constructed and this also constitutes a potential explanatory landscape of realist synthesis, which is reconfigured as Table 4.2.

TABLE 4.2 ABOUT HERE

Even the standard realist shorthand for all this, namely ‘what works for whom in what circumstances and in what respects’, is itself an explanatory mire. But, in one way or another, diving into this explanatory swamp marks the first task in realist synthesis. The reviewer has to reconnoitre conceptually and decide on a chosen path, and this process can be usefully broken down into three steps, as follows.

Mapping the territory

First, the reviewer pitches in, using a checklist such as the one in Table 4.2, to surface the impending questions for review. The fate of an intervention depends on all of these processes and they must have at least the potential to surface in the review. The first task is thus to give them some substance in relation to the intervention under review: to establish which programme theories seem important; to get a feel for the circumstances in which they tend to get tried; to gather hunches on implementation difficulties; to tease out ways in which the programme might misfire; and so on.

Experienced reviewers might well be able to anticipate some key issues, but the prospective explanatory agenda should be mapped in conjunction with an initial, exploratory reading of the primary materials.

Prioritizing review questions

The second task is to select the chosen questions for review from the preliminary set just identified. No review can ever be completely comprehensive and unless the reviewer has a year or two – or more – to spare, it is quite impossible to probe in depth all of the issues in Table 4.2. Prioritization has to be the rule but there is no golden rule for selecting the key explanatory issues. These may be settled pragmatically, on the basis of the prior interests of the commissioner, reviewer or practitioner. Somewhat more strategically, a particularly characteristic trait, a feature of palpable novelty, a point of potential fragility, or an area of dispute within a programme may be singled out as the burning issue for review.

By way of illustration, I introduce a couple of the syntheses pursued in later chapters. Chapter 5 reviews Megan’s Law, the US scheme notifying neighbourhoods of the impending release of a former sex offender into their community. One of its distinguishing features is its particularly lengthy implementation chain, which runs back and forth from prison and police authorities to the public and probation services and, of course, to the released offender. Accordingly, the crucial issue determining whether a ‘long sequence’ intervention like Megan’s Law works is the integrity of the implementation chain, and this was chosen as the focus of this review. The question for a second review on ‘naming and shaming’ (Chapter 7), was inspired by a quite different issue. The basic idea of taming underperformance by public disclosure of performance data through ‘league tables’, ‘star ratings’ and other such devices, is used right across the policy waterfront. It might be thought of as a ‘broad brush’ scheme and so this review heads for the contextual boundaries, trying to find out the sectors in which shaming might be successful.

Formalizing the model

The third and final step in the first stage of a realist review is to formalize the model to be tested. These hypotheses are not in the form of yes/no, pass/fail questions about programme efficacy; they are explanatory and so exactly like the conjectures of substantive social science. They are also realist and so will speculate on configurations of mechanisms and contexts that give rise to varied programme outcomes. For example, the review of youth mentoring in Chapter 6 starts with a simple four-stage model of the requisite mechanisms. It postulates that in order to successfully re-engage disaffected youth mentoring must work as befriending (making mentees feel differently about themselves), as confidence-building (offering a guiding hand though the difficult choices confronting the mentee), as coaching (coaxing mentees into practical gains, skills and qualifications) and as advocacy (sponsoring mentees by using contacts and networks). The model anticipates that this sequence will be difficult to achieve and sets the reviewer the task of finding the mentoring partnerships and circumstances that will trigger this sequence.

Note that I refer to the realist review hypothesis as a model. This is because, as in the mentoring example, it should refer to a *linked subset* of questions. Compared to a standard ‘does X cause Y?’ hypothesis, such a model will be policy and substantively rich. It locates the decisive aspects of the programme theory, allowing the reviewer to

pass into the inner workings of a programme and establish programme re-engineering as the potential goal. So, when it comes to examining any particular youth mentoring programme, the review is primed to focus inquiry on whether it delivers on the four stages and why. I stress again the parallel with conventional empirical inquiry. The realist evaluator might ask that question of a particular scheme; the realist reviewer seeks the evidence across the totality of existing studies.

Note, again, that the process of abstraction is a key aspect of theory-specification. The explanatory terms in the mentoring hypotheses – befriending, confidence-building, coaching and advocacy – operate at a middle level of abstraction. They are concrete enough to be identified in particular primary studies and yet abstract enough that material on them can be uncovered across a medley of case studies. This intermediacy allows the reviewer to test and develop the theories under review and is the key to producing transferable policy lessons.

The first counterpoint to meta-analysis is thus sketched out within the first column of Figure 4.3 as a three-step process of mapping, prioritizing and formalizing a model to be explored in the review. This stage ends with a set of hypotheses that are explanatory, middle-range and interlinked.

Before moving on to the second stage of the realist review, an important aside is in order. I have emphasized the non-linearity of realist synthesis, and the connections with other stages can already be foreshadowed. First of all note that, with these initial decisions on the focus of the review, and with this initial analytic purchase, synthesis has already commenced. Note further, that it is only possible to carve out such a review question *after* a period of hard reading. One can only fix on a model with real potential for learning transferable lessons about policy if one already knows the territory quite well. Accordingly, a good deal of preliminary searching and extraction of materials will have already occurred in shaping up the preliminary model. Note finally, that commissioners have an important role to play here. Simply plumping for an ‘explanatory review’ is not a sensible alternative to meta-analysis. Rather than commissioners merely handing over an unspecified bundle of questions, and rather than reviewers picking up those sticks with which they feel most comfortable, both parties should work together on this pre-review stage to negotiate and clarify the focus.

Stage two: searching for primary studies

The second stage in a standard systematic review is to seek out studies that will throw light on the question established in stage one. Conventionally, this task has involved hunting down all the primary research examining the relationship between a tightly specified treatment and a narrowly defined outcome in a closely delineated population. As a result, searching is often seen as relatively mechanical exercise whose sole purpose is to be meticulous and as comprehensive as possible in identifying every single research paper that corresponds to the given specification.

Realist review starts with a more complex question or, as just argued, with a series of interlinked hypotheses. Search procedures are correspondingly more intricate, and locating the apposite evidence is a prolonged task that stretches from spring (before the review question is fully framed) to autumn (when the synthesis is well underway).

Put another way, realist synthesis, like any good empirical enquiry, feeds on fresh evidence as it unfolds. Accordingly, it is useful to think of the search strategy for realist review as having four separate components, although even this implies a neatness and linearity not achieved in the hurly-burly of a real paper chase.

A background search

This sizing up of the forthcoming task is no more than a preliminary range-finding exercise, and there are no specific technical tricks or procedural rules to be followed. Thus my profound existential advice is ‘just do it’. Rather more helpfully, I might add that it is a good idea to get an information professional involved in order to pick the brains of someone who has knowledge of the spread of potentially useful databases and experience of the nuances of search terms and search techniques. The review of mentoring programmes in Chapter 6 was preceded by such a preliminary scout around in which I learned, to my surprise, that not only were mentoring programmes installed in every walk of life from accountancy to zookeeping but also that there was an industrial-sized literature on each. Hence came the abrupt lesson for the reviewer – draw in horns. There no correct or incorrect procedure for making such discoveries. The measure of the success of such a scoping search is to enable the reviewer to make an initial judgement on whether the right volume of materials of the right substance is out there to answer the questions the review is likely to pose.

A search to track the programme theories

This search focuses on locating the sources of administrative thinking, policy history, legislative background, and key points of contention that lie behind the intervention: to repeat for emphasis, it is *not* a search for evidence about the programmes per se. The purpose of such a reconnaissance is to initialize theory-building, to help surface the ideas and aspirations and plans and guidelines that underpin the intervention. In other words, it is a search whose purpose is to help formulate the review question, and thus occurs in parallel with stage one of the review.

As an example, let me return to the vexed question of publicizing league tables of the performance of public bodies such as hospitals and schools. The search for this review uncovered a variety of sources, making different claims about how the measure might work. Amongst other theories, there were those that saw league tables as a market measure (informed consumer choice drives up standards) and others that saw them as an internal regulator (ammunition to pressure poor performers into pulling up socks). Again, there is no formal method for searching out appropriate sources for key ideas, and what most characterizes this phase is the type of material under scrutiny. Programme theories are to be discovered in primary evaluative studies, especially in their ‘summary’ and ‘discussion’ phases, but they also feature in the ‘no go’ areas for traditional reviews – namely, the administrative, legislative and planning documents that accompany interventions and also in the ‘think pieces’ and critical work that surround them.

The goal is to find solid evidence about the main programme theories that guided the construction of an intervention, and to ensure that that they are ‘in play’ in accounts of its implementation. This is the rather pragmatic test of whether this phase of the search is adequate. Programme theories, of course, stretch right down into the minutiae of, say, a manager’s capacity to lead project staff, but the aim here is to uncover core theories, the investigation of which may yield transferable lessons.

A search for empirical evidence to test the theories

This is in some senses the search proper in which the reviewer moves on from browsing to shape up ideas, to delving in order to track down apposite evidence on the selected theories from a range of primary studies using a variety of research strategies. The methodological stakes are raised at this point, the goal being to scan the literature as thoroughly as possible.

The purpose of such a search is to find primary studies that will help interrogate the explanatory model about how the programme works. Take the Megan's Law example, which is described in detail in Chapter 5. I have characterized the intervention as having a long implementation chain, and the search thus needs to capture empirical material on each of the key linkages. The model in question (somewhat simplified) says that in order to work properly, the programme has to have: a secure means of identifying high-risk sex offenders; a method of producing and updating accurate registers; a clear channel to notify a community about the release of an offender; a scrupulous surveillance system mounted by the community; and stigmatized offenders with an awareness of heightened risk and decreased opportunities. Finding studies pertinent to each linkage requires subtle changes in conceptual focus and corresponding adjustments in keyword searches. The material sought here, however, is much more like the orthodox empirical evaluations that are the bread and butter of standard reviews.

A final search to fine-tune the synthesis

Once the synthesis is almost complete the reviewer seeks out additional studies to test those further and revised programme theories that often emerge in the course of the review. Realist synthesis is selective in terms of its focus on particular mechanisms and contexts within an intervention, but its understanding of these features is bound to develop in the course of the review. For example, in the review of mentoring programmes (Chapter 6), the spotlight was always on the relative status of mentors and mentees. However, it became apparent only relatively late in the review process that such an angle could be further teased out by examining the unusual configuration of patient support groups for serious illness, in which the mentor is embedded rather more deeply in the (low-status) out-group than is the mentee. Such decisions are clearly not premeditated and, once again, the exercise of judgement is paramount in the extent and the manner in which searches are expanded.

I stress again that this is a very orderly presentation of the prolonged and repetitive agony of locating appropriate primary materials. The most obvious simplification is that a study identified earlier in the sequence (for example, in the theory-building stage) will, in some cases, also pass muster as a useful member of the later samples. The key point is that each of the searches, and all of the samples uncovered, are defined in terms of the use to which they are put. As such, they might be thought of as relatively pragmatic and unpremeditated tasks. However, no criticism is intended in the use of these labels. The logic is that of purposive sampling, aiming to retrieve materials purposively to answer specific questions or test particular theories.

It may be useful to rehearse some of the time-honoured characteristics of such an approach. Purposive sampling is iterative in that it is usually repeated as theoretical understanding develops. In a qualitative case study a researcher may discover the

importance of a key actor or group of actors relatively late in the day. They can be approached and given a voice as and when this occurs. In the same way, research synthesis needs some flexibility to allow for supplementary searches that fine-tune a developing line of explanation (Long et al., 2002).

Purposive samples have a progressive focus as understanding unfolds. Ethnographers learn to navigate their way around communities and come to understand the significance of certain events over time. In doing so, they learn that one useful encounter often begets another and that key informants are often a key source of advice on where to proceed. There are parallels with realist synthesis. Search strategies and terms used are likely to evolve as understanding grows. Because useful studies will often make reference to companion pieces that have explored the same ideas, purposive searching makes as much use of ‘snowballing’ (pursuing references or authors by hand or by means of citation-tracking databases) as it does of conventional database searching using subject keywords. In a recent systematic review (Greenhalgh et al., 2004) conducted along realist lines, the authors took the trouble to chart this balance. They found that 52 per cent of all empirical studies referenced in the final report were identified through snowballing, compared with only 35 per cent through database searching and 6 per cent through hand-searching.

Purposive samples achieve closure through ‘theoretical saturation’ (Glaser and Strauss, 1967). In qualitative case studies the sampling of actors, events or actions ceases at the point when it seems that it is unlikely to add new knowledge. Realist synthesis is an exercise in identifying, testing out, and refining programme theories, and an almost infinite set of studies could be relevant to such a task. Consequently, a decision has to be made, not just about which studies are useful in developing explanations but also about when to stop looking. The same rule of thumb – cease when sufficient evidence has been assembled to satisfy the theoretical need or to answer the question – applies. For instance in the mentoring review (Chapter 6), scores of enquiries demonstrated the facility of one of the four mechanisms identified, that is to say ‘befriending’ appears to come relatively easily. The sheer weight of numbers on this score was not particularly useful (and the search for further instances curtailed) for the main point of the review was to examine the balance across all four mechanisms.

Note, however, that the principle of theoretical saturation may offer a different and more salutary lesson to the reviewer. In practice, it is rare to find an over-abundance of useable primary studies. As soon as an intervention is subdivided into its different processes and components, the search becomes targeted at quite specific theories and it is often difficult to find material that meets the precise specification. At certain points, the process often feels more like scavenging for information than picking the plums from an embarrassment of riches.

As far as the mechanics of searching goes, realist reviewers use bibliographic databases, libraries, the internet and personal contacts as sources of information, as in conventional systematic review. They also employ the same techniques to extract useful data, including keyword and citation searching of databases, search engine explorations of the internet, hand-searching of key journals and so forth. There are some different points of emphasis, however, which can be summarized as follows:

- Because it deals with the inner workings of interventions, realist review is much more likely to make use of the administrative and other ‘grey literature’ rather than relying solely on formal research in the academic journals (and on the databases that cover this material).
- Because it takes on programme theories rather than programmes per se, a much wider breadth of empirical studies may be deemed relevant, and these will sometimes be drawn from different bodies of literature in different settings. For example, studies on the public disclosure of performance data by schools will have important lessons for health care organizations and vice versa. Hence, a tight restriction on the sources to be searched, and the keywords to be used in searching, is often inappropriate.
- Because it looks beyond treatments and outcomes to the processes occurring in the ‘black box’, the keywords chosen to instigate a search are significantly more difficult to fix. As a rough approximation one can say that in terms of their ability to score definite and useful hits on a bibliographic database, proper nouns (such as Big Brother) outstrip common nouns (such as mentor), which in turn outdo abstract nouns (such as relationship). However, theory-building and testing utilizes these terms in the opposite proportion. Accordingly, if one is trying to locate material on, say, what makes for a good ‘match’ between ‘mentors’ and ‘mentees’ of different ‘status’, snowballing is likely to be significantly more fruitful than plugging these often imprecise terms into a search screen.

Stage three: quality appraisal

The next, and time-consuming, stage in traditional systematic review involves the quality appraisal of each candidate primary study. The methodological stakes are upped at this point, and the traditional call-to-arms is as follows: if evidence is to have its say, then it should be based on primary studies that have been carried out to the highest methodological standards. Whilst I applaud the underlying sentiment – the commitment to quality – I have spelled out already some of the profound difficulties in applying it. And I argue unambiguously that the hierarchy of evidence descending from biomedical interventions, with RCTs sitting imperiously atop, has to be abandoned.

The evaluation of interventions of interest in realist synthesis (complex systems thrust amidst complex systems) has to draw on complex bodies of evidence, interrogating programmes both in terms of process and outcomes, and delving at their micro-, meso- and macro-levels. Such an evidence base will inevitably draw on the whole repertoire of social science research and the approach to quality appraisal has to be sensitive to the entire range. The ostensible alternative, namely to manufacture quality checklists for every conceivable brand of social research and apply them collectively prior to the synthesis, is also fatally flawed. Thus far it has resulted in quality frameworks that are gargantuan, abstract and permissive (Spencer et al., 2003). They involve a wholesale grilling of the primary inquiries and still fetch up with no more than a considered opinion on quality. The dream of cross-matching hundreds of primary studies with dozens of indulgent appraisal tools, often drawing on more than one yet-to-be-devised checklist per study, will only end in unadulterated nightmare.

The realist solution is to maintain the commitment to the methodological scrutiny of a wide ranging evidence base but to cut much more directly to the judgement on quality. The guiding principle is the one used throughout, namely, that the appraisal criteria should be subordinate to the usage to which the primary study is put. The reviewer should not attempt to line up and appraise every candidate study on its own terms and as a whole but, rather, appraise the contribution that each one makes to the developing synthesis. That contribution is unlikely to stem from the entirety of a study.

The realist approach to synthesis has at its heart a model of programme theory, but primary studies are very unlikely to have been constructed with an exploration of that particular theory as their *raison d'être*. More probably, they will have been conducted across the multiplicity of banners under which evaluation research and policy analysis are organized. However, in so far as they have a common commitment to understanding an intervention, few of these investigations will have absolutely nothing to say about why programmes work. In the case of qualitative research, for instance, there is a reasonable expectation that key programme theories will get an airing alongside descriptions of everything from stakeholders to implementation hitches. In a quantitative study, there is likely to be emphasis on net outcomes but there may also be discussion of relative success across subgroups and, perhaps, a judgement on the integrity of implementation. Some distillation of this information may provide useful clues on a particular programme theory.

The focus on programme theory raises a completely revised expectation about the nature of research synthesis, namely that evidential fragments or partial lines of inquiry rather than entire studies should be the unit of analysis. In terms of research quality there is a parallel transformation down to the level of the specific proposition. Quality assessment should be review-specific; research appraisal should be theory-driven. Because realist synthesis takes a specific analytic cut through the evidence, it is not a sensible requirement that every clause and every paragraph of every one of the many-sided claims in a cross-disciplinary, multi-method evidence base has to be defensible. What must be warranted, however, are those elements that are eventually put to use in the overall synthesis.

How might this revised expectation be put into practice in the course of realist synthesis? As with the search for primary studies, it is useful to think of quality appraisal as occurring by stages. The reviewer asks, 'is this study good enough to provide *some evidence* that will contribute to the synthesis?' and there are two grounds upon which to deliver an answer, and two crucial moments to do so.

Assessment of relevance

The first decision is to resolve whether a primary study has the appropriate content to add to the review. Is it relevant in the first place? Is it in the right ballpark? Does it connect at all? The preceding discussion of realist search procedures stressed that they, too, are driven by similar questions. The purposive task is to hunt down studies that are relevant to the programme model under scrutiny, and searches typically involve the progressive focusing of keywords and more pronounced usage of snowballing. These techniques alone are insufficient, of course, to ensure that the study identified is indeed fit for the purpose of interrogating the theory under review.

One simply cannot take on trust titles, abstracts and secondary discussion. It is necessary to read the material!

This initial scrutiny marks the turn from search to assessment, and the first point of reference is to judge whether the primary research is relevant *at all* to the particular line of inquiry being pursued. Can it even begin to deliver inferences that are useful to the review hypothesis? Once again, there is no exact formula for making such a judgement. Some of these decisions will be relatively broad-brush. Searches on ‘youth’ and ‘mentoring’ will fetch up studies in which young people are both providers and recipients of advice and, having done some initial reading to decide which study belongs to which category, the axe will fall according to the slant of the review.

As the synthesis develops, such decisions become much finer-grained. Realist synthesis builds models about how programmes work and the assessment of the relevance of a primary inquiry has to be made against the specific propositions involved. For example, part of the model investigated in reviewing Megan’s Law (Chapter 5) was about whether media influence distorts the intended mode and official means of disclosing the identity of the ex-offender to the community. Does ‘the news’ go beyond the intended message, which stresses the need for surveillance and vigilance but warns against intimidation? The only studies I could find that came near to addressing this issue were about the media’s treatment of released sex offenders in general and not specifically about those released under the law. Here, one faces a much tougher call on the relevance of available material.

Appraisal on the grounds of relevance has to proceed case by case and, in general terms, it is likely that there will be a large (and auditable) wastage of primary investigations that have nothing pertinent to say about the review hypotheses. However, the matter of relevance is far from the last hurrah of quality assessment.

Assessment of rigour

There always comes a point in any research synthesis when a proper methodological appraisal is required. The first eligibility hurdle for a primary research study is relevance, but it is also vital that it is trustworthy. The realist approach takes the same line on both relevance and rigour, that is to say, both are subordinate to the overall strategy of the synthesis. Judgements about rigour are made not on the basis of pre-formulated checklists, but in relation to the precise usage of each fragment of evidence within the review.

The notion of explanation-sensitive standards will ring alarm bells in the homogenized world of meta-analysis, but there is nothing alien to scientific inquiry in such a notion. The iterative relationship between theory and data is a feature of all good inquiry. All investigation starts with understanding E_1 and moves on to more nuanced explanations E_2, E_3, \dots, E_N absorbing and rejecting many different kinds of evidence in the process. There is gradual change in what is sought by way of evidence, and the data have to meet different challenges. Applying this model to research synthesis introduces an alternative primary question for quality appraisal, namely, is it of sufficient quality to help in clarifying the particular explanatory challenge that the synthesis has reached?

The best way to explain this is to picture the synthesis and any particular primary study as two parallel inquiries. Both are investigating the complex systems that are the subject of the review. They will share some common objectives but by no means all. The primary study will bring together a particular suite of evidence to support a particular set of inferences. The synthetic study will not pursue an identical explanatory agenda, so will only use a subset of that initial bloc of evidence to support its own set of inferences. The test of research quality to support such a manoeuvre can only be this: does the primary subset of evidence support the secondary inference?

The synthesis builds by accommodating further primary studies, and the same procedure is followed. Primary study number two will bring together a somewhat different concatenation of evidence to support its unique set of inferences. The synthetic study will not pursue an identical explanatory agenda, so again it will only use a subset of this second tranche of evidence to support its developing set of inferences. And the test of research quality to support such a manoeuvre remains the same: does the new subset of primary evidence support the emerging secondary inference?

For illustration, take an example from the review of youth mentoring (Chapter 6). Two qualitative studies were deemed particularly relevant in testing the ‘phasing model’, which postulated that engaging disaffected youth requires a mentoring role that progresses from befriending to confidence-building to coaching to advocacy. Study one provided evidence of a programme in which the mentor is able to fulfil all of these functions. Study two provided evidence in which mentoring fails to proceed beyond the initial, affective domain. Both studies, however, commit a rather common methodological sin in qualitative research in that they over-generalize, implying that theirs is the common fate of all mentoring programmes.

The quality assessment of these two studies concluded that the grand inferences were unwarranted, but that the primary data on the respective success and failure of the two schemes provided a valuable contribution to an overall model seeking out the conditions for successful engagement mentoring. In this example lies a significant and decisively different tenet of research assessment, compared with the standard approach to systematic review. Studies that are technically deficient in some overall sense may, if inspected closely, still provide trustworthy nuggets of information to contribute to the overall synthesis. The attempt to remove bias procedurally (by admitting only one mode of research) is replaced by the goal of safeguarding inferences (by inspecting the precise usage of evidence).

 Supplementary reading. Further analysis of this principle, and of the illustration, may be found in: R. Pawson *Digging for nuggets: how ‘bad’ research may yield ‘good’ evidence*, which is included in the web page support materials. 

Thinking ahead to the review template (Figure 4.3), there is a further timely lesson here about the rhythm of the appraisal exercise. I have argued that there is an inevitable difference in the usage of each primary study in the synthesis. It follows that the test of research quality must be tailored to each specific usage. And from this it follows that the methodological assessment can only be made in and alongside the synthesis. This is a rather radical departure from the running order in the traditional model. Meta-analysis likes to get its quality appraisal over and done as a single exercise. I am suggesting that primary studies be assessed in two different ways at

different points in time. There should be an assessment of relevance, following in the tailwind of the search process, but the assessment of research quality has to be part and parcel of the synthesis itself. It should occur there in the text, right alongside the substantive analysis. The worth of a study is determined in the synthesis, and judgements on this score will occur as each study, from first to last, is written into the synthesis. This is how quality assessment is timetabled into Figure 4.3.

Stage four: extracting the data

The next stage in systematic review is its most characteristic. Given that it is generally an uphill, time-consuming slog, it is also the aspect that reviewers most dread. Conventional reviews proceed at this point by lining up primary studies that have made it through the quality filter; fine-tuning the set of characteristics on which to compare them; combing through the studies to extract precisely the same items of information from each; and, finally, recording these data onto a standard extraction form. In the simplest meta-analysis, the information retrieved is relatively sparse, namely information on the modalities of the treatment, its effect size and spread of impact. The extraction form becomes, so to speak, a data matrix from which the overall conclusion on net effects is calculated. In ‘mediator and moderator’ reviews, a range of additional information is collected from the primary studies about further attributes of participants, settings and interventions. This information is mined in the form of variables, since a data matrix remains the intended product of the exercise. Perhaps more surprisingly, qualitative reviews often conform to this expectation about uniformity of data extraction. A crucial difference, however, is that grid entries take the form of free text, consisting of short verbal descriptions of key features of interventions and studies.

The good news, for anyone who has performed this thankless task, is that the data extraction exercise as described above has no exact equivalent in realist review. The bad news is that the hard labour in question gets transferred to other points in the process. It is perhaps worth re-emphasizing that all research synthesis involves a thorough reading and detailed processing of information from primary studies, with the consequence that after carrying out a couple of reviews many researchers yearn for retirement or a new career.

There are two reasons, quite familiar by now, why realist synthesis has little use for standardized data extraction forms: first, the original sources are used for *quite different* purposes in the course of the review and, second, each source is expected to contribute *quite different* information to the synthesis. Accordingly, realist synthesis has a more elongated model of extracting (or, more accurately, processing) the information from primary sources. The reviewer is not loading uniform bits of information onto the review conveyor belt, but erecting an explanation, and so the amount, detail and form of materials excavated varies through the process. It is useful to think of this close encounter with the primary texts as having three stages and an extended timetable, which is noted in Figure 4.3.

Annotation

Firstly, as discussed above, some primary sources are asked to do no more than provide potentially relevant concepts and theories. If the reviewer is in this theory-tracking mode, documents are scoured for ideas on how an intervention is supposed to

work. Here, the extraction of primary materials begins with nothing more profound than marking the relevant passages with a highlighter pen. These are then noted and given an approximate label. Further documents may reveal neighbouring or rival ideas. These are mentally bracketed together until an initial model is constructed of key aspects of the intervention theories. Take, for example, the process of developing an understanding of how mentors and mentees interact, discussed in detail in Chapter 6. This is a long-term, relatively unplanned exchange and described, moreover, in hundreds of different ways in the literature. Nevertheless, there are similarities and difference between these pen pictures, and what emerged from immersion in the primary sources was a basic typology differentiating affective, cognitive, aptitudinal and positional mentoring. In summary, when in theory-tracking mode the reviewer is reading primary materials in order to marshal ideas. Ideas are not lodged uniformly through the literature and so realist reviews begin the processing of primary information by note-taking, annotation, conceptualisation and abstraction.

Collation

When the review turns to theory testing mode, the processing of evidence becomes more complex but still operates in purposive mode. Here the reviewer is dealing with empirical studies that have passed the test of relevance. Note is first made of which review hypothesis or hypotheses they address, what claims are made with respect to which theories, and how the apposite evidence is marshalled. Again, processing begins with highlighting and annotating the passages containing the key evidence. These extracts are then pigeon-holed together as being appropriate to the testing of a particular feature of the model under review. Once again, there is no uniform technical trick for creating the appropriate classification. In the Megan's Law example (Chapter 5), the synthesis is conceived in terms of testing out the strength of a long and linked implementation chain. Accordingly, the processing of evidence consists of identifying which bits of data speak to which linkages: which passages of which studies test the 'accuracy of the initial registration of high risk cases' hypotheses; which segments of which research interrogate the 'accurate maintenance of registers' hypothesis; and so on though the half dozen key links in the chain. The retrieval of data is about creating a tracking system to apportion appropriate evidence to the different parts of the overall model, not a task that can be undertaken using the standardized extraction form of the conventional systematic review. In general in explanatory reviews, different bodies of evidence using different modes of inquiry, drawn from different fragments of a primary study are required to test different parts of the model under review.

Reportage

There is one further feature of data retrieval in realist synthesis that distinguishes it from the standard procedures, and which also takes it on from simple annotation, classification and note-taking. Significant portions of the primary evidence are propelled into the synthesis itself. Synthesis, by realist lights, is all about trying to draw careful inferences from evidence and this requires that vital (and often lengthy) extracts of that evidence have to find their way into the synthesis. Many, many examples follow in later chapters. The general point made here is that if, say, intervening variables are deemed vital to understanding the working of a programme then the path analysis in which they are uncovered has to be presented in some detail. Alternatively, if the testimony of particular subjects is deemed significant, then summaries of their reasoning should be extracted. The objective is to extract and

present enough of the original data to ensure that the reader knows the basis on which inferences are made. The strategy mirrors and supports a previous principle about the value of a study being determined in the synthesis.

Two further aspirations of the realist reading of evidence are worth noting. The first is the attempt to remain true to the difficult principle of transparency. As with any mode of synthesis, there is a tendency to end up with piles of paper on the floor as the reviewer tries to recall which bit of which study speaks to which bit of the developing theory. These heaps will have already survived another ordering in which the reviewer tries to chart whether a source is useful for theory-tracking or theory-testing or both. Just as a conventional review will append a list of studies consulted and then give an indication of which contributed to the statistical analysis, so too a realist review should trace the usage and non-usage of primary materials. However, the archaeology of decision-making is more complex and thus harder to track in an explanatory synthesis. One is inspecting multiple theories, and specific studies may interrogate none, one, more, or all of them. Nevertheless, as realist synthesis moves beyond a developmental stage, reviewers should expect to develop a record of the different ways in which studies have been used (and omitted).

A second point is to recall that the steps involved in realist review are not in fact linear; studies are returned to time and again and processing occurs all the way down the line. Note in particular, that there always comes an unspoken point in the sifting and sorting of primary materials where one changes from model-building to model-testing and from theory-construction to theory-refinement. The reviewer experiences a shift from divergent to convergent thinking as ideas begin to take shape and the theories underpinning the intervention gain clarity. Accounts of systematic review which insist on its reproducible and thus mechanical nature are being economical with the truth in not recognizing this ineffable point of transformation and defining feature of good scientific inquiry.

Stage five: synthesizing the evidence

The defining moment, the point of it all, has now been reached, namely, the act of synthesis. I have already signalled a major difference in underlying philosophy. Realist review eschews the notion of delivering summative verdicts and, instead, perceives the task of synthesis as one of refining theory. Interventions are theories employing highly complex, non-linear sets of ideas that envisage different roles for individuals, teams, institutions and structures. Realist synthesis inhabits the real world of policy formation in which decision-makers know that, in attempting to prompt change, programmes operate through highly elaborate implementation processes, passing through many hands and unfolding over time. Realist review starts with a preliminary understanding of such processes, which it seeks to refine by bringing empirical evidence to bear on the various highways and byways of the initial theory map. It begins with theory and ends with – hopefully – more refined theory. What is achieved in this signature stage of the review is fine-tuning of the reviewer's understanding of an intervention and so 'synthesis', by these lights, refers to making progress in explanation.

Such an explanatory quest is inevitably complex. Gains may be sought on a number of fronts and the act of synthesis may be understood in terms of (and directed at) a

range of objectives. The initial task in this section is to spell out some of these different purposes. Recall, first of all, that the diversification of potential goals stems from the realist understanding of causality. Realism eschews the mechanical model of cause and effect, replacing it with a configurational puzzle. The traditional ‘does it work?’ question transforms instantly into at least five different lines of inquiry: *what* is it about this kind of intervention that works, for *whom*, in what *circumstances*, in what *respects* and *why*? The *why*? question remains paramount in any realist synthesis, of course, and there is no expectation that every aspect of every other theme can be covered. For instance, a ‘for whom’ synthesis may lean towards questions about the optimal targeting of an intervention. The review would consist of a close inspection of the evidence on the inner workings of an intervention and this would be married to the available data on the pattern of beneficiaries. Sustainability might be the trigger issue for an ‘in what respects’ review, in which evidence would be marshalled on implementation and the long term pattern of outputs and outcomes.

A further glimpse of the scope of realist synthesis may be seen in Figure 4.3, labelled a compendium of explanatory tasks, and which was itself drawn from Figure 2.7, a diagram in Chapter 2 that reveals the complex anatomy of social interventions. These figures sketch out the broad landscape of realist explanation, the point being that synthesis cannot capture everything about everything. It must make a choice on where to focus its efforts. Freedom from the unanswerable ‘does it work?’ question can lead realist synthesis to some quite unusual and yet perfectly salient issues from the compendium. There is no reason, for example, why realist synthesis cannot address the ubiquitous question about the need for ‘joined-up thinking’ on modern interventions. Many, many services are delivered by multiple agencies and it should be possible to produce a theory about the optimal arrangements by reviewing material on joint working arrangements. And at the outer limits, what about this one? Since no family of interventions has ever failed to generate perverse consequences, would it not be possible to review the collected mishaps and seek to discover whether unintended effects are, to some degree, predictable!

This clamour of potential focal points is daunting and will, no doubt, be anathema to the ‘one question, one answer’ school of systematic review. At the risk of imposing a straitjacket on a research strategy in its development stage, these doubts may be assuaged by identifying some of the more characteristic issues on which realist synthesis settles.

Synthesis to question programme theory integrity

This approach to synthesis aims to discover the typical weak points and major stumbling blocks in the implementation of the interventions under review. It addresses the issues represented in Figures 2.2 and 2.3 in Chapter 2, namely that implementation chains are often long, thickly populated and reliant on the actions of many different stakeholders. In such cases, as the theories-of-change literature suggests, programmes are only as strong as their weakest link. Using primary studies to identify data on each linkage is thus a powerful analytic strategy, and the nature of synthesis bears precisely upon the breaking strength of the overall apparatus. The review of Megan’s Law in the next chapter follows this strategy and discovers some vital points of strain.

Synthesis to adjudicate between rival programme theories

Although realist review stresses that interventions are theories incarnate, there is often dispute about precisely how they work. This state of affairs should come as no surprise because programmes are often based on simple, and thus contestable, theories about human nature. To achieve longevity, programmes need broad appeal and thus usually have the capacity to accommodate different ideologies. In such circumstances, a realist synthesis interrogates the evidence in order to adjudicate between rival theories of how interventions work. Marshall et al.'s (2000) review of hospital rankings and report cards is a good example. This focused on whether such public disclosure impacted on hospitals through consumer choice, purchasing decisions, enhanced regulation or practitioner shaming. The notion of synthesis as arbitration can clearly be seen in one of its telling conclusions: 'currently available report cards are rarely read by individual consumers or purchasers of care and, even if accessed, have little influence on purchasing decisions'.

Synthesis to consider the same theory in comparative settings

This approach to synthesis assumes that particular programme theories work in some settings and not others, and aims to make sense of the patterns of winners and losers. In many ways it is the quintessential realist strategy in that it focuses on contextual constraints on the action of programme mechanisms. It also allows for some out-of-the-box policy learning, in that such hypotheses are likely to lead the reviewer across policy domains and responsibilities. Policy ideas are frequently borrowed in the corridors of power and, as new problems crop up, old ideas are dusted down in devising potential solutions. A worked example explored in Chapter 7 looks at one such ubiquitous policy in the form of public disclosure, or 'naming and shaming'. To emphasize the point again, the intention is not to reinstate the verdict business, pronouncing that public disclosure programmes work in sector X but not in agency Y. The goal is to produce a general theory of the conditions that support and hinder the programme theory.

Synthesis to compare official expectations with actual practice

This approach to synthesis aims to compare the 'official' intervention theory and what goes on in practice. It is a particularly useful framework for analysis if the intervention under review has clear a legislative or regulatory foundation. Criminal justice programmes often carry such a birthright, and this analytic strategy is used to supplement the 'implementation chain' synthesis carried out in the Megan's Law example in Chapter 5. The theory to be tested is drawn quite directly from the legislative documents and mission statements underpinning the law. Note, incidentally, that this 'formal theory versus informal practice' motif is a characteristic of many primary studies, for example, investigations of service delivery in health and care settings (Shaw et al., 2004). Thus raw materials for this style of review are often plentiful.

This elaboration of the choice of analytic framework returns us to the well-worn theme about the elongation of the traditional stages in a review. The selection of one or other of the above explanatory motifs occurs in the foundation stages of a review: synthesis is not just a wrapping-up operation and its major theme has to infuse the whole process. Whilst reviewers are still considering priorities, when they are busy tracking down and annotating programme theories, they will have at the back of their minds some of these major analytic frameworks (a point reinforced in Figure 4.3).

So much for goals – what of the deed itself? The dominant activity at this stage must have at its core a practical means of handling each piece of evidence and bringing them together for the purpose of theory-testing and refinement. The technical process of realist synthesis is illustrated in Figure 4.2. The raw materials of this process are explanatory propositions and explanatory inferences, and the preliminary theory to be tested is depicted at the top of the figure. This initial model (the great circle) is itself a set of proposition explaining in abstract terms how certain key aspects of an intervention work. Each primary study is then inspected for evidence, according to how it supports, weakens, modifies, supplements, reinterprets or refocuses the preliminary theory. This process is depicted in the second stage of the figure, as the gathering of smaller circles around the main explanation. These are spread around the explanatory whole in order to show that primary inquiries are quite likely to impact on only a portion of the working model. The smaller and large circles intersect as a simple representation of the idea that only a portion of the evidence marshalled in a primary study is likely to be relevant to the synthetic explanation. Three of the primary studies overlap in order to illustrate that the synthesis may also involve squaring one instance with another. Synthesis, in other words, is not a simple, one-study-at-a-time process. The absorption of primary materials will also require attention to their sometimes contradictory and sometime complementary evidence.

FIGURE 4.2 ABOUT HERE

The dynamics of synthesis itself are represented in the third section of the figure. Adaptive theory is the product of realist synthesis and this gradual transformation in understanding the intervention is depicted as a movement from an initial to a revised model. The shift in explanatory power is mimicked by the changes in the shaded silhouette from perfect to battle-worn circle. There is no uniform mode of adaptation on each encounter with empirical material, and it is not a case of thumbs up or thumbs down for the explanatory model. In order to describe the practice of synthesis, it is useful to return to the domain principles announced at the start of this chapter. Evidence speaks to theory in contradistinctive tones and in moving from case to case, reviewers are involved in juxtaposing, adjudicating, reconciling, consolidating and situating the evidence.

The reviews presented in the next three chapters will resonate with such reasoning, but let us follow a couple of illustrations schematically, as in figure 4.2. Cases 1 and 2 represent primary studies in the mentoring review that is described in detail in Chapter 6. This begins with a model anticipating severe difficulties for an individual mentor in trying to provide support in the affective, cognitive, aptitudinal and positional domains. Case 1 says otherwise and provides qualitative case history materials of very successful growth through these stages. The review does not collapse in a heap at this news but absorbs the evidence by paying special attention to the contextual conditions pertaining in the case. In other words, it *situates* the data. It uses information on the characters and communities involved to better specify the model in respect of for whom and in what circumstances the programme works. This growth in understanding is represented in the lower portion of Figure 4.2, which bulges at this point in recognition of a significant addendum to the model.

Case 2 in this particular review also involves qualitative case histories and presents evidence showing that mentoring fails to get beyond the affective domain. Indeed, the author claims that success in befriending actually negates progress in orthodox education and career terms. Once again, the findings are absorbed into the review by noting the personal and institutional conditions pertaining in this particular programme, which in terms of the emerging synthesis take a bite out of the overall understanding of contextual contingencies.

Also evident in these two cases is another key synthetic manoeuvre, namely that of adjudicating between studies. Alongside synthesis comes quality appraisal and, as noted earlier, such assessment is part of the synthesis. Both studies were considered to fall foul of a common error in qualitative research, namely that of reaching a gross verdict on youth mentoring on the basis of highly specific samples. Their overall glosses on youth mentoring (enthusiastically positive and utterly negative) are brought into balance (via the bites and bulges) in the developing model.

Cases 3, 4 and 5 in the figure represent another sequence of synthetic reasoning. This illustration will be treated in depth in Chapter 5 on Megan's Law but, for the present, we commence with study 4. This trial shows that matched samples of offenders, released before and after the enactment of the law, reoffend at the same rate. This disappointing evidence creates a sizeable but unexplained bite out of the programme theory about how the law will provide increased protection for communities. However, the same study also shows that arrest rates quicken after the introduction of the law, hinting at the paradoxical outcome that the programme improves the detection rather than the deterrence of repeat offences. This revised explanation is hardened by study 3 (which shows that the opportunity for the preventative surveillance of a sex crime is remarkably low) and by study 5 (which maps out the upgrading of probation and police record-keeping as a result of the law). After consideration of three studies, there is a developing, adaptive theory that what actually happens with the enactment of the law is improvement in the classic apparatus of locating and hunting down suspects. Methodologically speaking, whole tranches of evidence are being juxtaposed, reconciled and consolidated – and this is synthesis laid bare.

It remains to locate this revised understanding of the act of synthesis into the overall template in Figure 4.3. Looking down the fifth column, one sees that synthesis begins early by establishing an overall theme, by devising a particular explanatory model of the programme in question. Thereafter, synthesis is the act of spelling out the current state-of-play of the model upon each encounter with a primary study, and this is reflected in the way the synthesis is presented. Studies are written up as a consideration of the model in the light of each empirical interrogation. The lessons learned are not singular and discrete, as illustrated by the combined usage of evidence above from cases 3, 4 and 5. Accordingly, by the time the mid-synthesis stage is reached, the presentation must also convey how successive suites of primary material have been absorbed into the evolving model, and this developmental analysis is evoked in the figure.

Finally, the synthesis reaches a resting point and this will consist of an abstract summary of how the model has been revised across the whole empirical piece. The Megan's Law review concludes with a summary of the entire implementation chain

and an assessment of the breaking strength of the overall apparatus. The mentoring review builds a model of the totality of relationships that must be present for long-term, engagement mentoring to succeed. The public disclosure review constructs a theory of the key contextual conditions that need to be present in order to attempt to name and shame.

Throughout the fine-tuning of the model, every adjustment in reasoning must be evident within the presentation of the review. Realist synthesis heeds the old advice of the maths teacher to always ‘show your working’. This little motto is reinforced if one reads across the template in figure 4.3. The key act of synthesis is described as ‘absorbing primary materials into developing synthesis’. In order to do this transparently, reviewers have to justify the inferential shifts to the reader and, to do so, they must provide an intelligible account of the original empirical material and a justification of why it can be relied upon to warrant the inference made. In other words, key aspects of data extraction and quality appraisal are carried out in harness with synthesis.

Stage six: dissemination

The act of synthesis may be the coup de grâce in research terms, but the ultimate finishing blow has to be delivered in the policy arena. Judgement on the big issue of whether the findings of systematic review can deliver decisive policy thrusts, or whether they are bound to remain pin-pricks, is delayed until the final chapter. Here, the focus is on the mechanics of dissemination. Contemporary accounts (Walter et al., 2003) stress that, for research to be properly utilized, this concluding stage should go well beyond the submission of a final report to the commissioner. The situation in which the systematic review jury retired for several months and appeared with a verdict many steps removed from the real world of policy is becoming less common, and two important changes for the better are in hand.

The first is for commissioners of reviews to be much more closely involved in the production of the research synthesis, a state of play that Lomas (2000) has called linkage. Researchers can only address themselves to a question, and decision-makers can only find pertinence in the answer, if that question has been adequately honed and left without major ambiguity. The second form of redemption is for reviewers to bring their technical expertise closer to the policy issue in question. Research synthesis needs to be able to locate recommendations in relation to the policy options on the table, and this objective is supported if the research takes cognizance of the practical needs of a range of stakeholders in the shaping of an intervention. Both requirements place a premium on avoiding overly technical language in dissemination, cutting instead to the quick and using the parlance of decision-making.

Realist synthesis is comfortable with both of these innovations. Indeed, it raises the status of linkage from a recommendation to a methodological requirement. Realist synthesis is theory-driven. The tasks of identifying the review question, articulating key theories to be explored, and choosing an overall analytic theme cannot occur meaningfully in the absence of input from practitioners and policy-makers, because it is *their* questions and *their* assumptions about how interventions work that form the focus of analysis.

Similarly, and by its very nature, realist synthesis is well placed to meet the second desideratum. It eschews the notion that reviews deal in political arithmetic and it operates automatically with the cautious and contextualized grammar of policy discourse. Realist reviews deliver models, which in policy terms are not the end but the beginning of the story. They initiate a process of thinking through the tortuous pathways along which a successful programme has to travel, and their conclusions take the form of reflections on how to navigate some significant highways and byways. Accordingly, what the recommendations describe are a series of key decision points through which an initiative has proceeded, and the findings are put to use in alerting the policy community to the caveats and considerations that should inform those decisions. For each decision point, a realist synthesis 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, L and M have also been considered’; ‘N’s effect tends to be short lived’; ‘O really has quite different components – P, Q and R’; and ‘S works perfectly well in T but poorly for U. The review, inevitably, will also reflect that, ‘little is known about V, W, X, Y and Z’.

Only time (and perhaps Chapter 8) will tell whether such a strategy may find favour in the policy community. There is, however, a strong rationale for why this middle level of abstraction might prove the most useful. Chapter 3’s critique of the standard model of systematic review warns of the perils of over-confidence. Trying to offer decision-makers concrete verdicts on whole families of interventions can only produce simplistic advice, which is usually inaccurate to boot. Chapter 2 warns of the untold levels of complexity that lurk in any policy question, but there is little utility in describing to decision-makers the thousandfold reasons why they cannot make a decision. The solution described in this chapter shows how realist synthesis can take some strategic cuts through the implementation swamp. It does not take on the full A-to-Z of programme complexity but concentrates on a subset of the lexicon. The four priority themes described in the previous section (analysis of weakest links, rival explanations, rival targets and departure from official expectations) provide feasible and modest goals for evidence-based policy. Prioritizing specific programme theories in research synthesis delivers a further dividend, moreover, in that it allows policy-makers to insert a priority of their own into the basic designs – namely, that realist synthesis should concentrate on the policy levers that can actually be pulled.

It remains to place these ideas on dissemination and utilization onto the realist review timetable. When should the liaison between reviewers and decision-makers occur? The popular and growing recommendation is that they should hold hands throughout the review. However, this prospect is usually somewhat unrealistic and the tryst is surely best located at the beginning and end of the process. In practice, this means the commissioner coming to the reviewer with a broad list of questions about an intervention. The reviewer questions the questions, and suggests further angles that have resonated through the existing literature. Then there is more negotiation and, eventually, a firm agreement about which particular lines of inquiry to follow. This process is thus assigned to the template in Figure 4.3 alongside other preliminary steps involved in navigating through the initial conceptual swamp.

As well as this initial meeting of minds, realist synthesis also anticipates that the review itself will partly reorder expectations about what is important. Realist thinking

on unintended consequences must also be applied reflexively. This means that room for further rounds of negotiation must be left open about whether, say, an unforeseen chink in the implementation chain deserves closer inspection. When and how often these adjustments should take place is itself a matter for negotiation, but for illustrative purposes one such reorientation period is depicted on the template. Note, however, that at several intermediate points there are long periods when reviewers should be left to their own devices. They should, for example, be able to apply their expertise on matters such as the methodological rigour and relevance of the primary research materials.

The final, and by definition still most crucial, act of dissemination occurs with the propagation of the findings of the review and this is marked at the terminus of the time and task dimensions in Figure 4.3. Realist synthesis has the traditional role of providing an independent and dispassionate overview of an intervention through the existing research. Conclusions and recommendations have to reflect this objective and this standpoint. However, the end product is a more refined theory rather than a final theory. The progress made in a realist review is not one from ignorance to answer, but from some knowledge to some more knowledge. As a result, extraordinary care must be taken at the point where findings are transformed into recommendations, and close liaison with decision-makers is once again required in thrashing these out.

The intended outcome of the dissemination process, as with all systematic review, is that those on the ground take note of the findings and implement them. In the case of meta-analysis, such changes might be monitored in terms of simple behaviour changes in the direction of particular recommendations (for example, are clinicians prescribing therapy X for condition Y?). However, implementation of the findings of a realist review is a complex process involving multiple actors, multiple processes and multiple levels of analysis. Furthermore, implementation is not a question of everyone stopping doing A and starting to do B. Rather, it may involve subtle shifts of emphasis in a programme in one setting, expansion of that programme as it stands in another setting, and the withdrawal of exactly the 'same' programme in a third setting. Quite different decisions may flow as more informed judgements are made about what works, for whom, how, in what circumstances and in what respects. The ultimate goal is that individuals, teams and organizations take account of all the complex and inter-related elements of the programme theory that have been exposed by the review and apply these to their particular local contexts and implementation practices.

Conclusion

The conclusion to the chapter is, of course, the long-awaited Figure 4.3, which sets out an agenda and a timetable for realist synthesis. All systematic reviews start with a question and end with answer, but the sequence and tempo of the intermediate steps – as well as the nature of the answer – are different in realist synthesis.

FIGURE 4.3 ABOUT HERE

The figure attempts to summarize the entire apparatus of a synthesis but it is important, however, to clarify some expectations about the template. In trying to set down research designs, the methodologist always confronts the twin perils of over-simplification and over-prescription.

Hectic as it is, Figure 4.3, is still an over-simplification. I have stressed all along that realist synthesis is an iterative process, and that throughout the mission the reviewer has to make judgements on where next to turn. This means that the 'time' and 'task' locations on the template are approximate. Realist synthesis has to be able to respond to new data and to new ideas, and no one can say precisely when a new lead will surface and where it will take the review. In short, a maze of feedback loops and repetitions will be encountered in bringing a real realist synthesis to completion. By and large these are not depicted on the figure. The willingness to acknowledge such thinking on one's feet as an integral part of the design will be disconcerting to many in the review community who regard procedural uniformity as the be all and end all. My view is that such pauses for thought occur in all reviews and their suppression in the classic protocols is 'brutishly destructive of some of the most important aspects of research and scholarship' (MacLure, 2005).

Because it is so hectic Figure 4.3 is also over-prescriptive. I have this little nightmare of the tyro researcher trying to conducting a realist review by working through this chapter, navigating across the template and fretting, 'well, it must be time to be getting on with step four, phase three'. Such ritualism would be a mistake, and a sign that this book has failed, for what I am trying to describe is not some esoteric ceremony in which the routines must be exact for the magic to work. The core is quite simple. Programmes are theories about how to change behaviour. Primary research provides evidence on the utility of those theories. Systematic review draws together that evidence in order to refine the theories. If the reviewer keeps the spirit of this little sequence of postulates in mind, then many of the practical steps on the way will follow automatically.

This is not to say, of course, that there is no point formalizing the method. Quite the contrary; it is vital to do so because the ability to 'rationally reconstruct' method lies at the heart of any claim to be scientific (Popper, 1972). All of the steps in this chapter describe a 'logic of discovery' and to defend the conclusions of a synthesis is, in the last analysis, to be able to defend that logic. The strategy described in Figure 4.3 is thus not only a research design but a repository of all the previous arguments about the nature of social interventions and realist explanation. Doing realist synthesis is not just a question of following the above logic but of fashioning the very text of the review in terms of that logic. This returns us to the venerable principle of transparency, although in the case of realist synthesis one is transparent about explanatory processes that other methods cannot reach. This thought provides the last word of advice on how to conduct and how to write a realist review. Above all else, what must remain clear is the process of theory-building and refinement.