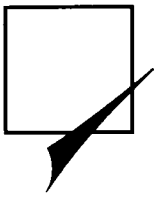
Evaluation

SAGE Publications
London, Thousand Oaks
and New Delhi.
Vol 2(1): 35–50



Demonstration, Exemplification, Duplication and Replication in Evaluation Research

NICK TILLEY
The Nottingham Trent University

The nature and meaning of replication is crucial not only for those undertaking studies of efforts at replication, but also for those conducting, using and evaluating demonstration projects, since their rationale lies in lessons learned for use elsewhere. The issue of replication is discussed using as a case study the highly influential British Burglary Prevention Project in Kirkholt, Rochdale and efforts to emulate it. Three ways of construing replication are presented: 'strict', 'relativist' and 'scientific realist'. Serious weaknesses are identified in the first two, and the third is advocated. Major lessons for practitioners and evaluators are drawn, in order that most benefit can be obtained for practice and policy from demonstration projects and their successors.

1. Introduction

'Demonstration projects' are used to find out whether a proposed policy or practice can effectively address an identified problem. They form part of what Karl Popper termed 'piecemeal social engineering' (Popper, 1945, 1957). The idea is that before embarking on widespread and expensive changes, organizations—in Popper's case governments—should try out innovations and examine whether they work. Popper also hoped that piecemeal social engineering could be a useful tool for social scientists to test their hypotheses (Popper, 1957).

The issue of replication is crucial for demonstration projects. A demonstration project can only usefully inform policy and practice if its results are replicable. To the social scientist the replication of findings can be an important indicator as to their generalizability. Moreover, where there are scientific disputes about issues the route to resolution is often seen to lie in replications of contentious studies.

This article begins by highlighting a number of theoretical and practical problems in relation to the nature, conduct and evaluation of replications, and then goes on to suggest how they might be resolved. The argument is developed through detailed discussion of a highly influential British burglary prevention project, and efforts to replicate it within a national crime prevention program. The demonstration project in

question is called 'The Kirkholt Burglary Prevention Project', and the national program, 'The Safer Cities Programme'. The first phase of the Safer Cities Programme was run by the Home Office between 1988 and 1995. It operated in 20 cities, each of which had £250,000 per annum to spend on efforts to reduce local crime problems.

Sections 2 and 3 of the article discuss Kirkholt and some candidate replications of it within Safer Cities. Referring to the examples outlined, Section 4 considers critically various ways of construing replication, and advocates a scientific realist approach. Section 5 concludes with general comments on the conduct of evaluations of demonstration projects and of their replications.

2. A Case Study of a Demonstration Project

The recorded crime rates for the years before and after the implementation of the Kirkholt Burglary Prevention Project indicated dramatic and sustained falls in burglary. Using March to February figures, in 1986/7—the year before the project—there were 526 burglaries. The corresponding figures for succeeding years are given in the 1990 report of the project. There were 223 burglaries in 1987/8, 167 in 1988/9 and 132 in 1989/90 (Forrester et al., 1990). These represent a fall from approximately 25 to 6 percent per annum of the 2280 households on the estate. Moreover, the probability of reburglary amongst those already victimized was reduced from four times the expected rate to zero in the first 7 months of the project. Clearly practice and policy interest in replication follows in this case from the apparent success of Kirkholt.

Controversy, however, has surrounded interpretation of the 'success' of the scheme. It has been argued that other estate improvements undertaken by the local authority could have been responsible (Safe Neighbourhoods Unit, 1993). Farrington (1992) has attempted independently to referee between the two interpretations of the burglary reduction by reanalysing data from Kirkholt itself. Replications clearly offer another and classic way of arbitration. Moreover, unlike the Safe Neighbourhoods Unit and Farrington discussions, which are concerned with *internal validity* (the nature of the causal relationships within the project), replications attend also to the issue of *external validity* (the reproducibility of these causal relationships elsewhere). It is the latter which is crucial for practitioners, policy-makers and social scientists.

It was decided, therefore, to look at replications of the Kirkholt project within the Safer Cities Programme. The first task was clearly to identify a sample of replications. Difficulties in doing so turned out to be not so much technical as conceptual—they begin to highlight methodological issues.

Attributes of the Kirkholt project

1. 'Kirkholt' was conceived and undertaken as a well-resourced *demonstration project*, whose evaluation was funded by the British Home Office Crime Prevention Unit and handled by its Research and Development Section. This Home Office connection may have brought credibility. It certainly made the project relatively well provided for in terms of expertise, and may have helped draw in other resources. £298,398 was spent between 1985–6 and 1989–90.
2. 'Kirkholt' was about developing crime prevention measures in *high crime*

areas. By the prevailing national standards Kirkholt had an extraordinarily high burglary rate. As already indicated, the recorded incidence rate in the year preceding the Burglary Prevention Project stood at 25 percent of households. This compares with a national rate for recorded burglary of 2 percent in 1988, a British Crime Survey (BCS) national rate of 5 percent for 1988, and a maximum BCS rate of 13 percent for the sorts of areas most vulnerable to burglary. In Kirkholt the rate was especially dramatic 'given that 90% of units on the estate are of types associated in the British Crime Survey with only a medium rate of burglary victimization' (Forrester et al., 1988: 2).

3. 'Kirkholt' was about tackling high crime areas, which are clearly *circumscribed* and can thus be treated as *identifiable communities*. Kirkholt comprises 2280 dwellings. It is a relatively self-contained estate, clearly bounded by a motorway and major roads. Anyone entering or leaving the estate would know that they were doing so. It is also culturally fairly homogeneous, comprising almost exclusively members of the white working class.
4. 'Kirkholt' was about the *removal of highly attractive targets* (coin meters), which had rendered the area a popular one with burglars in which such 'money boxes' could confidently be expected. In the year leading up to the project 49 percent of the burglaries included theft of or from coin meters for electricity or gas, and 27 percent involved loss of meter cash only. It is not clear what percentage of households had these meters. Part of the scheme, however, included their removal. Coin meter removal will have disabled 'own goal' meter theft as well as external theft.
5. 'Kirkholt' was about carefully *diagnosing a particular crime problem* (burglary in an estate) and tailoring responses to these. Prior to the project the nature of the crime problems of Kirkholt were examined through interviews with 76 burglars, 237 victims and 136 neighbours of victims. An analysis of police crime report forms was also made. This work then informed the planned suite of interventions. Data collection from victims and neighbours continued in part to inform developments in the project in the light of changing burglary patterns.
6. 'Kirkholt' was about developing an *effective interagency response* to crime. Manchester University, the local police, the probation service and the local authority worked together closely during the project. They also drew in many other agencies. The geographical location of the project offices alongside the local housing offices facilitated a close working relationship. The police and probation services led the project in succession and, when not actively leading the project, members of each of these agencies had time set aside to work on it.
7. 'Kirkholt' was about *harnessing the community* to protect itself from crime (through cocooning). Kirkholt is not the sort of area which has traditionally provided fertile ground for the establishment of Neighbourhood Watch schemes (Husain, 1988; Mayhew et al., 1989; Laycock and Tilley, 1995). The formation of mini Neighbourhood Watch schemes (termed 'cocoons') was very actively encouraged and much work put into their maintenance. To begin with they comprised only six or so dwellings but grew to 20–25.
8. 'Kirkholt' was about focusing on *multiple victimization* and reducing it. Crime

prevention resources were allocated to victims, who were found to be four times as likely to be revictimized as those on the estate who had not been victimized. Target hardening, adapted to the specific risks revealed by research and judgements of trained police officers, was allocated on this basis. Cocoons were fostered by bringing together the victimized and their neighbours.

9. 'Kirkholt' was about *clarity of initial research, clarity of crime prevention method tailored to research findings, and clarity of leadership* in implementing measures. If the post hoc descriptions of Kirkholt are accurate (and there is no reason to believe they are not), the scheme was characterized by unusually clear planning, implementation, monitoring and evaluation. There was a strong academic input from Manchester University.
10. 'Kirkholt' was about burglary prevention and was *offence specific*. Kirkholt was not an all-purpose crime prevention project. It turned on specific analysis of the burglary problem and the development of particular measures to reduce it.

These attributes of Kirkholt are taken almost exclusively from published accounts of the project. Critics have latched on to other features. Of course, both the authors of the account of Kirkholt (who were also directly involved in the project) and its critics are inevitably selective. Theirs are reconstructions from the flux of what went on, abstracting what are presumably deemed the essential features of the initiatives and how they were developed. The above list is, of course, a further abstraction. No description of any demonstration project can fully reflect what went on in it.

3. Replicating the Demonstration Project

In view even of the range of attributes which are mentioned in the participants' texts, deciding for evaluation purposes what constitutes a real replication is clearly problematic. Are all these features necessary? Are others which are omitted crucial? Are some elements more important than others? Can individual aspects be abstracted? In this case, is it the particular collection of specific crime prevention measures which must be adopted, or is it the procedures for identifying those that are appropriate in the circumstances? The same kind of difficulty attends the practitioner trying to take advantage of lessons to be learnt from the project. Must it be adopted as a package? What can be abstracted and adapted? What supplementation still makes sense? In short, which attributes are incidental and which are essential?

Kirkholt has been and can be read in various ways. Using the 10-point grid as a yardstick, we can attempt an initial mapping of some similarities and differences in the following examples of candidate replications.

Candidate Replication 1 (?R1)

This was conceived in mid-1989, self-consciously using Kirkholt as a model.

1. *The project was resourced as a standard Safer Cities scheme.* It spent just over £95,000 over 2½ years, including £30,000 non-Safer Cities government funds. Its expenditure was thus much less than Kirkholt's.
2. *The project was not located in a very high crime area.* At the start of the scheme

the burglary rate was about a quarter of that found when the Kirkholt project was established.

3. *The area comprised some 8000 households. Thus it was close to four times the size of Kirkholt.* It had few clear boundaries separating it from adjacent areas and no clear centre.
4. *There were few electricity and coin meters.* They did not figure as targets of burglary, as they had in Kirkholt.
5. *Problems were diagnosed but using different data sources from those used in Kirkholt.* Police records of burglary in the area were analysed in the early months of the project. Victims were interviewed twice, first as soon as practicable after the offence and second some 6 weeks later, over most of the project's duration, and these responses were periodically analysed. Fourteen burglars were interviewed over the course of the project, and their characteristics examined, though not till within 6 months of the close of the project. No neighbours were interviewed. The slogan adopted in ?R1 was 'Act on facts' though the facts collected were not identical to those assembled in Kirkholt nor were they collected at the same stage of the project.
6. *?R1 was a multi-agency project, but structured differently from Kirkholt.* In ?R1 the police and probation officers were the central players. There was a full-time secondment from both. Neither was at any point clearly and explicitly in the lead, as had been the case in Kirkholt. The local university had limited involvement.
7. *Efforts were made to establish 'Neighbourhood Concern Groups', to mobilize community activity in crime hot spots.* These approximate Neighbourhood Watch schemes, though there is not the same direct and necessary contact with the police. Area coverage was patchy. At 25 percent, nothing like Kirkholt's 90 percent coverage was achieved.
8. *Analysis of the recorded burglary data on ?R1 revealed that council tenants had the highest rate of revictimization, and they had first call on security upgrading work.* Target hardening council tenant victims was thus the first priority. Housing association tenants who had been burgled also benefited from offers of free target hardening. Others judged vulnerable were also included, though owner-occupier victims, unless in receipt of state benefits, were not included. As in Kirkholt there was thus a focus on multiple victimization, though it was expressed in slightly different ways.
9. *There was a fairly clear conception of the project at the start of it, but this became less so over time.* There was continuing academic input as in Kirkholt.
10. *Like Kirkholt, ?R1 was clearly and explicitly about burglary.*

Outcomes. In the year before ?R1 571 burglaries were recorded. This rose to 694 during the first year of the project and 991 over the second year. The rate of increase was marginally less than that in the surrounding area.

Candidate Replication 2 (?R2)

This comprised a suite of initiatives aiming at responding to various crime problems on a housing estate, one element of which was concerned with household burglary.

1. *?R2 was a standard, relatively low budget Safer Cities scheme. It had far fewer resources than Kirkholt.* ?R2 cost £55,894, including £24,000 for lighting improvement, £10,280 for a BMX/skateboard track and £21,614 for lock fitting. The lock fitting element was specifically directed at burglary prevention.
2. *At the start of the project the burglary rate was 9 percent, about a third of the rate on Kirkholt when work began there.*
3. *?R2 is clearly circumscribed, with one major entry point.* It comprises 835 dwellings, just over a third of the size of Kirkholt. Asians make up 5 percent of the population.
4. *There are no precise figures for burglaries involving gas and electricity coin meters, but suppliers' numbers in the region dropped dramatically.*
5. *There was very limited problem diagnosis at the planning stage.* Forty-nine residents of ?R2 were interviewed by special constables in 1989, prior to the project. There were no interviews with victims or neighbours of victims. No burglars were interviewed.
6. *No full-time staff were seconded to ?R2. The project was led by Safer Cities.*
7. *No special efforts were made to establish Neighbourhood Watch in any form.*
8. *Security upgrades were offered to all on the estate.* Security work, including fitment of window locks, door locks and chains, was undertaken at 81 percent of the properties on the estate, some already having adequate security and others not providing access.
9. *There was no academic input into the ?R2 project.*
10. *Burglary was a major focus for the target hardening on ?R2.*

Outcomes. In the 20 months prior to the point at which 80 percent of the target hardening undertaken was complete (it took 4 months for this) there were 111 domestic burglaries on the estate. In the following 20 months there were only 38.

Candidate Replication 3 (?R3)

The burglary reduction project in ?R3 formed one element of a wider ranging approach to community safety in the area, which was overseen by a multi-agency task group. Other elements included a secondary school project, an alcohol awareness project, a girls and young women's project and a women's project.

1. *?R3 was a standard, relatively low budget Safer Cities scheme.* ?R3 cost a total of £51,150 over a 2-year period from 1991 to 1993.
2. *The recorded domestic burglary rate prior to the scheme was about 5 percent per annum, a fifth of Kirkholt's.*
3. *?R3 has 3936 households, getting on for twice the number in Kirkholt.* It is quite well circumscribed, and is culturally heterogeneous, with varying housing and tenure types. There is a large student population.
4. *Meters did not figure as a significant issue in domestic burglary in the area.*
5. *There was limited problem diagnosis at the time the project began, using different data sources from Kirkholt's.* ?R3's work was informed by a 1990 survey in which among other things it was found that 72 percent of residents were either very or fairly worried about being broken into. It was this that drove

the decision to develop a burglary reduction project. There was neither interviewing of victims and neighbours nor of offenders, as there had been in Kirkholt. There was no measurement of rates of revictimization.

6. *?R3 was not a multi-agency project.* A carpenter was employed to undertake necessary target hardening for victims of burglary from April 1991. He was supervised by the police and based at the local police station.
7. *No special efforts were made to establish Neighbourhood Watch in the area as part of the scheme.*
8. *?R3 focused on those already victimized as well as others deemed to be at risk.*
9. *One of the local universities undertook some crime mapping at the initial stages of the project.* This was not used significantly in implementation or monitoring.
10. *Burglary prevention was the major focus of ?R3.*

Outcomes. Between 1989 and 1992 there was an increase of 9 percent in the number of domestic burglaries recorded in ?R3, whilst the local subdivision as a whole experienced an increase of 139 percent. The proportion of repeat burglaries went down from 22.8 to 20.1 to 13.6 percent between 1989 and 1991. It is clear that ?R1, ?R2 and ?R3 include a wide range of similarities to and differences from both each other and the Kirkholt project. We now turn to various ways of understanding what is involved in replication, using Kirkholt and its candidate replications as examples.

4. The Meaning of Replication

Three ways of construing replication: the strict, the relativist and the scientific realist.

The Strict Conception: Replication as Duplication

The strictest view insists that a real replication must duplicate exactly. Since clearly the repeat cannot occupy the same space and time this is, as it stands, quite impossible for experiments either in the natural or social worlds. Replications of experiments can be the same only in particular respects. As Popper puts it,

> All the repetitions we experience are *approximate repetitions*; and by saying that a repetition is approximate I mean that the repetition B of an event A is not identical with A, or indistinguishable from A, but only *more or less similar* to A. But if repetition is thus based upon mere similarity, it must share one of the main characteristics of similarity; that is, its relativity. Two things which are similar are always similar *in certain respects*. (1959: 420–1)

In the social world, there are some specific problems in repeats of experiments. Different individuals will not have identical backgrounds or characteristics. The same individuals cannot be used as subjects in replication studies because their participation in the first event will have altered them. The same experimenter cannot be used for the much the same reason. The problems are greater still in open-air demonstration projects involving social interventions, since the local economic and political conditions in

which they take place are in continuous flux. Given that strict replication is in principle not possible, as Popper suggests, decisions have to be made about what constitute the essential features of the original project. Without some *theory* of how the original worked, the choice of essential features is liable to be arbitrary. Baseball batting furnishes a useful example. Those hitting well wish to replicate their successes. They are, evidently, led into seemingly bizarre replications of success conditions. They will not have their kit washed. They will put it on in exactly the same order. They will eat identical meals before the match. And so on. Their replication is, as the title of Gmelch's article (Gmelch, 1981) describing it suggests, magical.

In relation to Kirkholt ?R1 is closest to a strict replication in the sense that there, there was a self-conscious effort to replicate as much as possible, and, indeed, many features are reproduced. It did, however, differ from Kirkholt in various respects. The architects of the scheme were unable exactly to duplicate what went on, because the area was differently constituted (for example, it was bigger, multiracial and had an initial recorded burglary rate which was much lower), the resources available were different (for example, there were concurrent full-time police and probation secondees, no employment trainees, few special constables and no Home Office backup), the personnel differed, and so on. Whilst again deliberately drawing on Kirkholt, ?R2 and ?R3 had attributes even less like it than ?R1 and thus are even less strict replications of the original.

The much lower level of success in ?R1 could quite plausibly be said by defenders of the earlier project not so much to cast doubt on the success of Kirkholt, as to reflect the fact that it was not a strict replication. They would have much ammunition. The strict replication model can be very useful for those who have invested in the success of an initial project. Where others have subsequently failed, they have done so because they have missed some element, and they inevitably always will have done so given the nature of strict replication. In Popperian terms the requirement for strict replications would, thus, immunize Kirkholt from falsification. It thereby ceases to be empirically testable.

The Relativist Conception of Replication

The relativist line on replication accepts that strict replication is not possible. For the relativist there is no real replication. Instead there are socially located interpretations of projects fuelling decisions about what to reproduce and how to do so. There can, moreover, be no full description of the original to be replicated. In regard to the natural sciences relativist sociologists of scientific knowledge argue that understanding an original experiment to be replicated involves deployment of tacit knowledge. Often, personal contact will be required for the transmission of this to enable the person undertaking the replication to reproduce the experimental result (Travis, 1981; Collins, 1985). Reports of experiments or projects always involve a reconstruction in terms of what the author or authors choose to pick out because they deem it to be most important. The original account is inevitably selective. In being so, it will often serve the interests of those writing it. Debates in science over experimental results frequently turn on what exactly is to count as a competent replication of an experiment (Travis, 1981; Collins, 1985). Eventually negotiation results in more or less consensus

conclusions within the scientific community about the meaning of the phenomena being examined or manipulated. The relativist argument is that there is no 'true' replication. There are only tacit agreements about what is to count as a replication.

The relativist account of replication is distinctly unsettling. It denies any external standard against which what is to count as a replication can be judged. It leaves both the academic or evaluator, who might wish to examine a replication, and the practitioner, who might wish to undertake one, in a quandary. In a slogan adopted by Paul Feyerabend (1975), 'Anything goes'. Views on what is deemed to be or not to be a replication can quite reasonably vary very widely, and are largely matters of inclination. Any consensus that emerges represents only an acquired common taste. No text on its own can be sufficient to make a sensible stab at replication.

Returning to Kirkholt, the slipperiness of replication in the relativist account of it makes any or all of ?R1, ?R2 and ?R3 replication a matter of preference. ?R1, we have already seen, can be viewed in some respects as a replication, though there are good grounds to reject it if so desired. The same goes for ?R2 and ?R3.

Whilst strict replication is too demanding, since it rules out everything, relativist replication is too tolerant, since so much can count or not count according to whim, and no guidance can be given. Because these positions are so unhelpful to the practitioner, academic and evaluator, there are good grounds for trying to find some alternative position on replication which whilst remaining coherent can offer guidance.[1] It is at this point that we turn to scientific realism.

The Scientific Realist Conception of Replication

Scientific realism describes a philosophy of science (e.g. Hesse, 1974; Harré, 1986) which has informed some methodological writing in the social sciences (e.g. Pawson, 1989; Sayer, 1984), and which has been applied specifically to evaluation research (Pawson and Tilley, 1992, 1993, 1994). It is distinctive in using a generative, as against constant conjunction, account of causality. Thus, according to scientific realism it is the job of science to unpick the real causal mechanisms giving rise to regularities. It is not simply to observe and record recurrent patterns of events and thereby to impute causal relationships between variables. Moreover, scientific realism notes that real causal powers or causal potentials of phenomena are only released if the context is conducive to their activation.

According to scientific realism, experiments constitute efforts to construct conditions in which conjectured causal powers will be released. Gunpowder furnishes a simple example. It embodies certain causal powers: it has the causal potential to explode. It does not do so all the time, of course. However, if the right conditions are created, this potential causal power can be activated.

It has been argued elsewhere that quasi-experimental evaluation methods are rooted in classic constant conjunction constructions of causality and that they exhibit rather particular weaknesses on this account (Pawson and Tilley, 1992, 1994). This issue and the details of debates surrounding scientific realist philosophy of science lie beyond the scope of this article. It is hoped, though, that sufficient has been said to indicate the broad thrust of scientific realism, and the interested reader is directed to the works cited above for further details.

The scientific realist construction of replication attempts to redefine the issue in ways which avoid the unhelpful cul-de-sac into which it is propelled by the relativist and the implausible strictures of the strict reproducer. It does so by emphasizing three linked features of experiments and projects: *mechanisms*, *contexts* and *outcome patterns*.

A given measure, introduced for instance in a crime prevention project or experiment, is deemed to have its effect by triggering one or more underlying causal mechanisms. *'Mechanism' thus describes what it is about an intervention aimed, say, at crime prevention which leads it to have its outcome.* But mechanisms are only triggered in given contexts conducive to their activation. *'Context' comprises the conditions necessary for a causal mechanism to be triggered to produce a particular outcome pattern.* The context in which a given intervention measure in a project is introduced may or may not permit the activation of causal mechanisms which lead to particular outcome patterns. *'Outcome patterns' describe the results following from a project intervention effectively triggering causal mechanisms in a context conducive to their operation.*

Figure 1 represents graphically the relationship between intervention measure, context, mechanism and outcome pattern. According to scientific realism a theory is needed to specify conditions and causal mechanisms which will be fired by an intervention measure. This will explain or predict particular outcome patterns.

For the scientific realist replication is not about mimicking a large number of 'attributes', 'variables', 'conditions' and so forth from one trial to the next. Rather, the trick is to recognize and to reproduce those salient features of the *context* which are needed for the *mechanism/s* to be activated. Identification of these should help experimenter and practitioner alike to create specifiable *outcome patterns* of the original which is to be replicated. A sequence of studies in which *variation* is introduced may well be needed to check empirically conjectures concerning which contextual features are conducive to the effective operation of a given mechanism.

To illustrate the scientific realist view of replication, we return to the original Kirkholt project. Some of the attributes of Kirkholt which were described clearly relate to its context, whilst others refer to what was done which might trigger burglary

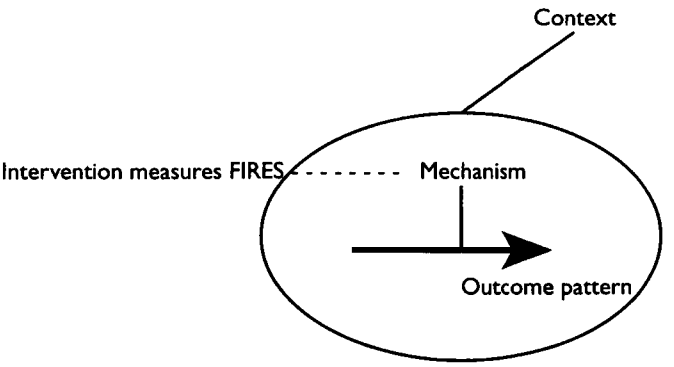


Figure 1. Context, Mechanism and Outcome Pattern

prevention mechanisms. Thus, that Kirkholt was a high crime area, that it was clearly circumscribed and that it was culturally homogeneous are features of the *context* in which the project was established.

The removal of meters, the establishment of co[illegible]s and the target hardening of already victimized dwellings were *measures* taken, which in turn triggered causal *mechanisms*. These might include respectively reducing rewards, increasing the offenders' perceived risk of recognition and apprehension, and rendering more difficult and more risky entry to properties hitherto found to be most vulnerable. These mechanisms (classic in the crime prevention literature, see Clarke, 1980, 1992) were not included at all in the project descriptions above, though they are crucial in generating the project's *outcome patterns*.

Some of the other Kirkholt attributes which were listed describe ways in which, implicitly at any rate, the salience of context in determining which measures to introduce to fire which mechanisms entered into the very conduct of the project. Through its problem-diagnostic, research-based approach the project in effect matched crime prevention measures to the details of the particular context in which it was hoped to impact on burglary. The remaining features of Kirkholt describe the necessary conditions for the delivery of this method of crime prevention: an adequate resource base, strong and clear leadership and an effective interagency partnership.

There is, thus, an implementation context which is necessary for the prosecution of a project-aligning context, measures and mechanisms, as well as a context in which the measures introduced by the project can trigger causal mechanisms generating particular outcome patterns.

To the scientific realist the interdependence of context and mechanism is crucial for an intervention measure to produce particular outcome patterns. For instance, it might be conjectured that the relatively small, economically disadvantaged, culturally homogeneous and well-bounded estate with little through traffic, which contained many of the offenders as well as victims, furnished conducive contextual conditions for the establishment of a cocoon home watch which could fire various crime-reducing mechanisms such as increasing perceived risk for prospective offenders.

The suite of initiatives introduced on the Kirkholt estate were tailored to its context and to those financial and other resources could be made available for their implementation. This allows a mechanism/context model to be developed, from which specific outcome patterns could be deduced. This was done in part in the examination of changing patterns of revictimization of those who had been burgled. The multiple methods mean that some effects will be 'overdetermined', making a predictive model very difficult. What could be produced, however, is a *theory* of the project-linking constituent measures, mechanisms, contexts and broad outcome patterns. Following our discussion, it might include the following:

1. *Security upgrading.*
   Context: A high rate of crime together with a high rate of reburglary.
   Measure: Target hardening of previously burgled premises.
   Mechanism: For the prospective burglar, an increase in the difficulty and risks of apprehension in obtaining entry to otherwise attractive properties.

Outcome: Lower rates of revictimization together with a reduction in the burglary rate overall.

2. *Target removal.*
   Context: A high proportion of burglaries involving cash from meters, alongside high numbers of cash meters.
   Measure: Removal of cash meters.
   Mechanism: Removal of incentive to burgle, through reduced actual or perceived rewards.
   Outcome: Reduction in the percentage of burglaries involving loss of cash through meter breakage, reduced risk of burglary at dwellings where meters are removed, and a reduced burglary rate overall.
3. *Cocoon home watch.*
   Context: A medium-sized, homogeneous, clearly defined estate with little through traffic.
   Measure: Stimulus and maintenance of near universal cocoon home watch.
   Mechanisms: Increased perceived risks of recognition of offenders, plus heightened levels of informal social control.
   Outcome: A reduced burglary rate overall and a general reduction in crime and incivilities.

Of course, these mechanism/context/outcome pattern relationships are conjectural. They are thought to be consistent with the Kirkholt reports. There are, however, other possibilities. Certainly other measures and mechanisms fired in the Kirkholt context may have produced outcome patterns which include a reduction in overall rates of burglary. As intimated above, a report prepared by the Safe Neighbourhoods Unit (1993) suggested that local authority work, including some security upgrading, some fencing to create 'defensible space' and some window replacement may have produced particular outcome patterns including burglary reduction. Properly formulated, these could be constituted as alternative/additional conjectures explaining particular outcome patterns, including overall reductions in the rate of burglary, though rather more detail of its outworkings would be needed. It could also, of course, be part of the 'overdetermination' of reduced rates. These are matters which require detailed theory building and research.

Those undertaking projects may well, of course, have a good idea of what happened within them and how any effects were produced. That said, views may differ. Moreover, first views of what mechanisms were triggered and what outcome patterns followed may not always be accurate. Hempel (1966) describes a series of plausible accounts to explain the differing rates of fatality amongst babies in two wards of the Vienna General Hospital before Semmelweiss hit on the ways in which infection had been passed on from doctors moving to the high fatality ward directly from performing autopsies. What was important was that Semmelweiss worked with conjectures and put them to the test. What we have from Kirkholt are plausible conjectures, referring to measures, mechanisms and contexts, making sense of patterned outcomes. They cannot be the last word.

The scientific realist construction of replication sidesteps the issue of duplication, which is the unrealistic aspiration of strict replication, and whose implausibility gave

rise to the unhelpful relativist position. Unfortunately, vestiges of relativism do remain. The context for implementing interventions is infinitely variable, and is also essentially open: it cannot be fully controlled and changes are in principle unpredictable. The natural scientist's laboratory can approximate empirical closure, reflecting theoretical specification. In 'open-air' situations, even though theoretical specifications of closure may be approximated, its accomplishment is in practice not possible. There is thereby an inescapable element of uncertainty. Put more simply in the example examined here, Kirkholt is not nor can it be insulated from internally and externally generated changes both firing new mechanisms which may impinge on outcome patterns, and altering the context for scheme-fired mechanisms. A series of studies, construed in scientific realist terms, can reduce but not eliminate this problem.

How then do our putative replications of Kirkholt fit into this scientific realist conception? ?R1, the replication most self-consciously following Kirkholt, partially followed one of its measures. That is, it targeted hardened victimized properties, though the focus was on tenants rather than all residents. Efforts were made to establish Neighbourhood Concern Groups, though these were never near universal and did not necessarily have a police focus. Meter removal did not figure in the scheme. As indicated, the context for ?R1 was different from Kirkholt—with a larger area, lower crime rate, more social heterogeneity, a less identifiable geographical area, and more through traffic. This would make implementation of the same measures firing the same mechanisms an inappropriate aim and an impossible aspiration. ?R1 is a replication of Kirkholt only in the sense that it used a similar repertoire of methods which were tailored to the local situation, which had been subject to systematic examination. The differing outcome pattern is to be expected. Whatever this had been it could neither confirm nor disconfirm the findings in Kirkholt.

In the case or ?R2, though the outcome pattern indicated success similar to Kirkholt's, the context, measures and mechanisms differed radically. For example, the estate was smaller, rather less culturally homogeneous, and had a lower initial burglary rate; the measures introduced included lighting improvements and improved recreational facilities not forming part of the Kirkholt project, and did not focus on repeat victimization or introduce cocoon neighbourhood watch which were included in Kirkholt; and the mechanisms triggered may have included diversion of prospective offenders, not brought about in the Kirkholt project as described here. ?R2 cannot be considered a replication in scientific realist terms of any or all of Kirkholt, and thus its success is of no confirmatory value.

?R3 looks at first sight as different as ?R2 from Kirkholt. Yet part of it, like Kirkholt, comprised an offer of security upgrading to all those in the area already victimized. Though the crime rate was not as high as in Kirkholt, there was a similar decrease in burglary risk amongst those already victimized following target hardening. There may have been some similar linked elements of context, measure, mechanism and outcome pattern, and thus partial replication of Kirkholt in scientific realist terms.

Kirkholt aligned context and measures to trigger mechanisms to achieve preferred and specifiable outcomes. That method of implementing crime prevention could be replicated. Given the openness of the social world, however, the same constellation of salient contextual circumstances will only occasionally and fortuitously be found, if

ever. However, though all features of the context may not be found, elements will, and here Kirkholt points to some measures which can successfully trigger mechanisms to produce specific outcome patterns. These might be replicated. Later work on repeat victimization patterns and measures to reduce them follows this path (Farrell and Pease, 1993; Bridgeman and Sampson, 1994).

5. Conclusion

If there is merit in the scientific realist approach to evaluation in general and its application to replication in particular, several lessons follow for those involved in conducting and evaluating demonstration projects as well as for those undertaking, planning and evaluating attempts at replication.

1. Those planning demonstration projects, which are intended to yield transferable lessons for policy-makers and practitioners, need to plan their work so that their conjectures concerning measures, mechanisms, contexts and outcome patterns are made explicit. The demonstration project needs to exemplify this theory. The conjectures may be mistaken, of course. Moreover, during the conduct of a demonstration project changes may occur in measure and context, and so too may practitioner conjectures about change mechanisms triggered. This does not, of course, assume that those running development projects will have the last word on what may have produced any effects observed. What is important is that the conjectures are made clear so that the evidence for them can be collected. In practice scope for reasonable alternative interpretation is likely to remain.
2. To enhance their value, evaluations need not only to indicate that a change has occurred but also what brought it about. 'What brought it about' turns out to describe the context and the mechanism fired in it. If this is fully understood then both internal and external validity standards will have been achieved. Evaluations of demonstration projects need, therefore, to be designed in terms of a theory of measure, context and mechanism which will predict detailed outcome patterns. Data need to be collected which relate directly to these. Typically several measures are introduced, each of which may trigger a number of mechanisms in the context in which the intervention is introduced. Here, a fairly wide range of data may be needed to arbitrate between several possibilities.
3. Practitioners and policy-makers reading evaluation reports and considering adoption of interventions described therein need to reflect on the contextual conditions conducive to the successful firing of change mechanisms activated by measures adopted. If these conditions are not present or measures are not implemented in ways which trigger the mechanism successfully then the same measure cannot be expected to yield the same outcome. What might be termed 'replication failure' may thus explain the apparently disappointing results sometimes found.
4. If demonstration projects are construed in the way suggested here they could inform decisions about where and how evaluated replications of successful elements of the project would be helpful to refine theory or reduce uncertainties. In this way series of projects could produce cumulative understanding relevant

> both to theory and to policy and practice. Evaluators of replication projects, when comparing practices, must in any case be clear that they are matching like with like. This does not entail duplication—we saw that this is an impossible and inappropriate aspiration. It does involve looking at salient contextual similarities for the triggering of mechanisms to be fired by the measures adopted. Replication evaluation, as construed here, should thus be of paramount interest to social scientists. It offers a way of checking conjectures, and corroborating theories.

These scientific realist principles obviously go for any area in which project evaluation is undertaken. The very least they do is bring home the 'negative' lesson that the same program can have quite different 'outcomes' when 'replicated' in different contexts. Had this realist truism been appreciated and followed earlier there would have been less chance of the destructive 'nothing works' conclusions for corrections in the 1970s and 1980s being reached nor would the doctrine have held sway for such a long time. Moreover, even if 'nothing works' no longer works as a slogan for the criminal justice system, the story of much intervention and evaluation remains one of uncertainty and inconsistency. There has not been cumulative development in theory or practice. This follows, it is argued, at least in part from weaknesses in methodology.

Though this article has sought to show that replication is complex and difficult, it should also be clear that replication studies are crucial if we are to increase our understanding of what works in what circumstances, and thereby improve the (cost)-effectiveness of interventions. Demonstrations cannot finally settle issues. In order sensibly to capitalize on their findings in the longer term, properly evaluated replications, which are well thought out and carefully planned, are important.

The example of Kirkholt shows that efforts to replicate it strictly would not be sensible. Moreover, efforts to take the measures off the peg and apply them expecting an automatic reduction in burglary would equally be unwise. Contextual variation is crucial. What this discussion of Kirkholt does indicate, however, is what can be achieved when contexts, measures and mechanisms are aligned effectively to produce particular outcome patterns. Kirkholt suggests a suite of conjectured burglary reduction alignments of this kind. It is these which may potentially be reproduced in appropriate circumstances. If subjected to empirical test, replications in this scientific realist sense promise more refined/corrected conjectures about what can work in what conditions. And this is what policy-makers and practitioners need to understand.

Notes

This article was based on a Police Research Group Paper, 'After Kirkholt—Theory, Method and Results of Replication Evaluations', to which Crown Copyright applies.

1. Among others Gellner (1974) and Tilley (1993) have attempted critiques of relativist positions, but the details of this debate lie beyond the scope of this article.

References

Bridgeman, C. and A. Sampson (1994) *Wise After the Event: Tackling Repeat Victimization*. London: Home Office.

Clarke, R. (1980) '"Situational" Crime Prevention: Theory and Practice', *British Journal of Criminology* 20: 136–47.

Clarke, R. (1992) *Situation Crime Prevention: Successful Case Studies*. New York: Harrow and Heston.

Collins, H. (1985) *Changing Order: Replication and Induction in Scientific Practice*. London: Sage.

Farrell, G. and K. Pease (1993) *Once Bitten, Twice Bitten: Repeat Victimization and its Implications for Crime Prevention*, Crime Prevention Unit Paper 46. London: Home Office.

Farrington, D. (1992) 'Was the Kirkholt Burglary Prevention Project Effective?', unpublished.

Feyerabend, P. (1975) *Against Method*. London: New Left Books.

Forrester, D., M. Chatterton and K. Pease (1988) *The Kirkholt Burglary Prevention Project, Rochdale*, Crime Prevention Unit Paper 13. London: Home Office.

Forrester, D., S. Frenz, M. O'Connell and K. Pease (1990) *The Kirkholt Burglary Prevention Project: Phase II*, Crime Prevention Unit Paper 23. London: Home Office.

Gellner, E. (1974) *Legitimation of Belief*. Cambridge: Cambridge University Press.

Gmelch, G. (1981) 'Baseball Magic', in I. Robertson (ed.) *The Social World*, pp. 286–9. New York: Worth.

Harré, R. (1986) *Varieties of Realism*. Oxford: Blackwell.

Hempel, C. (1966) *Philosophy of Natural Science*. Englewood Cliffs, NJ: Prentice-Hall.

Hesse, M. (1974) *The Structure of Scientific Inference*. London: Macmillan.

Husain, S. (1988) *Neighbourhood Watch in England and Wales: A Locational Analysis*, Crime Prevention Unit Paper 12. London: Home Office.

Laycock, G. and N. Tilley (1995) *Policing and Neighbourhood Watch: Strategic Issues*, Crime Detection and Prevention Series Paper 60. London: Home Office.

Mayhew, P., D. Elliott and L. Dowds (1989) *The 1988 British Crime Survey*, Home Office Research Study 111. London: HMSO.

Pawson, R. (1989) *A Measure for Measures: A Manifesto for an Empirical Sociology*. London: Routledge.

Pawson, R. and N. Tilley (1992) 'Re-evaluation: Rethinking Research on Corrections and Crime', *Yearbook of Correctional Education 1992*, pp. 19–49. Vancouver: Simon Fraser.

Pawson, R. and N. Tilley (1993) 'Oxo, Tide, Brand X and New Improved Evaluation', presented at the BSA Annual Conference, Essex.

Pawson, R. and N. Tilley (1994) 'What Works in Evaluation Research?', *British Journal of Criminology* 34: 291–306.

Popper, K. (1945) *The Open Society and its Enemies*. London: Routledge.

Popper, K. (1957) *The Poverty of Historicism*. London: Routledge.

Popper, K. (1959) *The Logic of Scientific Discovery*. London: Hutchinson.

Safe Neighbourhoods Unit (1993) *Crime Prevention on Council Estates*, prepared for the Department of the Environment. London: HMSO.

Sayer, A. (1984) *Method in Social Science: A Realist Approach*. London: Hutchinson.

Tilley, N. (1993) 'Popper and Prescriptive Methodology', *Metaphilosophy* 24: 155–66.

Travis, G. (1981) 'Replicating Replication? Aspects of the Social Construction of Learning in Planarian Worms', *Social Studies of Science* 11: 11–32.

NICK TILLEY is Professor of Sociology at the Nottingham Trent University (Burton Street, Nottingham NG1 4BU, UK). He is seconded part-time as a research consultant to the Home Office Police Research Group in London, where he has undertaken a number of scientific realist evaluation studies.