What is a mechanism? What is a programme mechanism?

The RAMESES II Project

The concept of a mechanism lies at the heart of realist explanations of how and why change occurs. Realism is based on a generative, rather than a successionist view of causality. Pawson & Tilley (1994: 293) explained how successionists try to understand cause/effect externally by describing constant associations between events (e.g. in billiard balls, when one ball collides with another and forces it to move). On the contrary, realism philosophy understands causation 'internally' and "cause describes the transformative potential" of the observed phenomena.(e.g. in gunpowder's potential to explode describes powers inhered in its chemical composition).

Successionists measure the social world through variables and explanation rests on distinguishing real from spurious observed relationships between two variables through experimental manipulation, for example, within a randomised control trial (Pawson, 2008). Such observed relationships between two variables are taken as evidence that there is a causal relationship between them.

In contrast, realists hold a generative view of causality which assumes that not all causal forces are observable.

> *The basic idea is that things that we experience or can observe are caused by 'deeper', usually non-observable processes. So, for example: we can open our hand and observe the tennis ball we held fall to the ground, but we cannot 'see' the gravity that causes the ball to fall. Similarly, we can experience a training programme and observe that participants use different language at the end of it than they did at the beginning, but we cannot 'see' the new content being stored in memory or the new connections being forged in the brain that enables them to do so. That is, the causal processes happen at a different level of the system than the observable outcomes. In realist philosophy, the underlying causal process is known as a 'mechanism'. (Westhorp, 2014)*

Pawson and Tilley (1997) employed this broad concept of mechanism within realist evaluation to understand how programmes give rise to intended and unintended outcomes. The specific mechanisms of interest in realist evaluation are therefore 'programme mechanisms' (Pawson and Tilley, 1997).

Programme mechanisms describe how it is that programmes and interventions contribute to outcomes (Pawson and Tilley, 1997). Realist evaluators do not subscribe to the successionist view that interventions directly cause outcomes; CCTV does not directly cause a reduction in crime, counselling does not directly cause a reduction in teenage pregnancies, clinical practice guidelines do not directly cause a reduction in the prescription of anti-biotics. Rather, realists understand that programmes offer (and sometimes take away) resources to (and from) participants. These resources may be material, social, emotional, political and so on. The interaction between these resources, and how participants interpret and act upon them, is known as a 'programme mechanism'. This is the pivot around which realist evaluation revolves.

> **"** The concept of a mechanism lies at the heart of realist explanations of how and why change occurs."

Programmes are complex and there are always multiple stakeholders involved in their implementation. For example, in antibiotics prescription interventions, key interest groups are clinicians, patients and their relatives, pharmaceutical companies, practice managers, and so on. Different stakeholders may respond in a multitude of ways and therefore the mechanisms through which a programme may work are likewise multiple. Furthermore, these responses are also constrained and supported by wider contextual features (see 1000 words on Context). For example, clinicians may decide to follow the guideline, they may choose to ignore it or they may try to 'adapt' the guideline to 'fit' the patients in front of them by following some, but not all of its recommendations. Their

reasoning will be shaped by context, for instance the age of the patient (or how demanding the patients' parents/carers are in the case of children), the interaction of the antibiotic with other medications the patient uses, their expertise, previous experiences of using the drugs for certain illnesses, broader incentive schemes, and the like.

What a mechanism is…. and is not

Pawson and Tilley's (1997) formulation of the concept of a mechanism tends to focus on changes in the reasoning and behaviour of individuals; however, others have also argued for this concept to be extended to understand how mechanisms work at different levels within systems (Westhorp 2014 a and b). Debates about a finer grained understanding of this concept are an ongoing feature of the realist literature, conference question and answer sessions and realist-friendly dinner party conversations. We expand on a few points within these debates here.

A common misunderstanding is to mistake programme activities or components with programme mechanisms (Pawson & Tilley, 1997, Leeuw and Astbury, 2010, Weiss 1997). To use the antibiotics example above, the programme resources are the guideline itself and in some cases a financial incentive systems used to encourage support clinicians' adherence to the guidelines. The context includes, to give a few examples, the pre-existing organisational features of the primary care, the existing IT systems that are expected to accommodate access to the guidelines and the capacities and preferences of patients. The intended programme mechanisms are clinician's awareness of and trust in the guidelines as a research-based guide to support clinical decision-making.

Pawson and Tilley (1997) (and others for example, Leeuw and Astbury (2010)) also caution against directly equating programme mechanisms with variables. Programme mechanisms are not "observable attributes of some unit of analysis" (Leeuw and Astubry, 2010) but are explanations of why and how outcome patterns occur. The search for programme mechanisms moves beyond the directly observed and measured variables that are indicative of outcome patterns to provide an explanation of those relationships. These programme mechanisms may not be directly observable, but this does not mean that they are not real. We still accept the pull of gravity as an explanation of why an object falls to the ground even though we cannot directly observe gravity itself.

Furthermore, programme mechanisms do not occur in a vacuum. Mechanisms exist whether or not they operate; however, whether or not they operate is shaped by context. For example, Dalkin (2014) in her realist evaluation of a palliative care pathway, found that there were fewer non-cancer than cancer patients registered on the palliative care register (an outcome pattern). Registration on a palliative care register is

intended to trigger a discussion about advanced care planning so that patients' wishes are made clear and care can be planned accordingly. In the context of cancer patients, who often have a somewhat predictable illness trajectory, this was indeed the programme mechanism through which palliative care registers worked, as registration lead to an increased discussion of patients' needs at multidisciplinary meetings.

However, in the context of people who had chronic conditions other than cancer, their illness trajectories were perceived to be less predictable. Consequently, clinicians were less confident that they could predict whether a patient with a non-cancer diagnosis was nearing the end of life and experienced greater anxiety about including these patients in the register.

> " ... mechanisms are not always activated but are 'causal tendencies' whose activation depends on supportive contextual conditions"

This was because they were concerned that, patients would be classified as nearing the end of life when they were not. This had both emotional consequences for the patient and family (people had been told they had a few weeks to live when this was not the case) and resource implications (they did not want to overpopulate palliative care registries unnecessarily, which could divert resources away from those who were truly nearing the end of life). This anxiety led to them avoid registering non-cancer patients. In summary, programme mechanisms are not always activated but are 'causal tendencies' whose activation depends on supportive contextual conditions (Leeuw and Astbury, 2010).

Thus, to summarise, programme mechanisms are:

- the interaction between program resources and the ways that participants interpret and respond (or not) to them
- an explanatory account of how and why programmes gives rise to outcomes
- hidden, but still real shaped by and interconnected with context.

They are not:

- Programme components, inputs or outcomes
- Variables, mediators or moderators.

Everytime I throw the ball up, it just comes back down again.

That's just gravity.

References

Astbury B and Leeuw FL (2010) Unpacking black boxes: mechanisms and theory building in evaluation, *American Journal of Evaluation* 31(3): 363–381

Dalkin S, Greenhalgh J, Jones D, Cunningham B, Lhussier M (2015) What's in a mechanism? Development of a key concept in realist evaluation. *Implementation Science,* 10:49: DOI 10.1186/s13012–015–0237–x

Dalkin S (2014) *The Realist Evaluation of a Palliative Integrated Care Pathway in Primary Care: What Works, For Whom and in What Circumstances?* Doctoral thesis, Northumbria University.

Gill M and Turbin V (1999) Evaluating 'Realistic Evaluation': Evidence from a Study of CCTV. In Painter K and Tilley N (eds) *Surveillance of Public Space: CCTV, Street Lighting and Crime Prevention.* Crime Prevention Studies Volume 10. Monsey, N.Y: Criminal Justice Press.

Pawson R and Tilley N (1994) What works in evaluation research?. *British Journal of Criminology*, 34(3), pp.291–306.

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

Pawson R (2008) Causality for beginners. In: NCRM Research Methods Festival 2008. http://eprints.ncrm.ac.uk/245

Weiss C (1997) Theory based evaluation: Past present and future. *New Directions for Evaluation* (No 76). San Francisco, CA: Jossey Bass.

Westhorp G (2014) *Realist Impact Evaluation: An introduction* ODI Research and Policy in Development, Australian Government Department of Foreign Affairs and Trade. www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/9138.pdf

---

Project team:

Professor Trish Greenhalgh
UNIVERSITY OF OXFORD
www.phc.ox.ac.uk/team/trish-greenhalgh

Professor Ray Pawson
UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/people/staff/pawson/

Dr Geoff Wong
UNIVERSITY OF OXFORD
www.phc.ox.ac.uk/team/geoffrey-wong

Dr Gill Westhorp
CHARLES DARWIN UNIVERSITY
www.cdu.edu.au/northern-institute/our-teams/603/5928#gillian-westhorp

Dr Joanne Greenhalgh
UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/people/staff/greenhalgh

Dr Ana Manzano
UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/people/staff/manzano/

Dr Justin Jagosh
UNIVERSITY OF LIVERPOOL
www.liv.ac.uk/cares

Realist evaluation, realist synthesis, realist research – what's in a name?

The RAMESES II Project

Realist research

Realist research is a collective name for research that is underpinned by the principles of realist philosophy (see 'Philosophy' in this series). Realists argue that there *is* a real world out there but our understanding of it is filtered through our senses, cultures and experiences (Wong et al 2012; Westhorp, 2014). Realist research uses a generative understanding of causality (Bhaskar, 1975, Pawson, 2008). That is, the outcomes we observe are generated by causal processes and forces that we cannot see (see 'What is a Mechanism?' in this series), and which operate (or not) according to contexts in which they occur (see 'What do realists mean by context?' in this series). Consequently, outcomes vary in different contexts.

> **"** Realist research uses a generative understanding of causality..."

Realist social research

Realist social research seeks to explain how and why the social world works and uses the relationships between contexts, mechanisms and outcomes to do so. For example, Fitzpatrick (2005) discusses the causes of homelessness. She notes that some positivists (see 'Philosophy', in this series) (for example, Randall and Brown, 1999) argued:

> *…there are continuing problems of rough sleeping in areas with no housing shortage. Equally, the great majority of people in poverty or with mental health, or substance abuse problems, do not sleep rough. … It follows that housing shortages, poverty, mental health and substance misuse problems cannot be said to cause rough sleeping. (p. 5)*

This view, Fitzpatrick argues, is underpinned by a successionist understanding of causality (if we see x and then we see y, on a regular basis, then x causes y.) In

research, this translates as seeking correlations between observable or measurable variables. However, Fitzpatrick (2005) argues that poverty and housing shortages are "real causes of homelessness if they can be shown to have a tendency to bring about homelessness, even if they only bring about actual homelessness on some occasions".

This is an example of realists' recognition that in an open social system, there are multiple different causal mechanisms operating at different levels of systems. Those mechanisms interact in different ways in different contexts, which means that only some people who experience poverty end up homeless. Furthermore, Fitzpatrick (2005) highlights that while positivists conceptualise poverty as a variable and measure what proportion of people experiencing poverty are homeless, realists ask "what is it about poverty that could cause homelessness?" Thus, realists seek to explain what it is about poverty (context) that gives rise to a causal pathway (mechanism) which leads to homelessness (an outcome).

Realist evaluation and realist synthesis

Realist evaluation (Pawson and Tilley, 1997) and realist synthesis (also known as realist review, Pawson, 2006) are both specific forms of realist research. They focus on understanding social programmes, initiatives or interventions. This is what makes them different from other forms of realist social research, which seek to understand other social phenomena. Realist evaluation is, as the name suggests, an approach to evaluation; realist review is an approach to literature review and synthesis of existing evidence.

Realist evaluation and realist synthesis are underpinned by a premise that all policies, programmes and initiatives combine activities, roles and resources which are designed to solve a social problem. However, they rely on human volition as well as a range of other factors to make them work. Some of those factors are social. For example, the success of an exercise routine to

reduce back pain may depend on the adherence of the person with pain to the programme (which will in turn be affected by a range of social and personal factors), access to available facilities, a willing coach, a group of like-minded exercise buddies, and so on. A realist approach recognises that programmes are not universally successful and work better in some circumstances than in others. Programmes implemented in different contexts work through different mechanisms and produce different patterns of outcomes. Thus, realist evaluation and synthesis do not seek to determine the 'average effect' of a programme or answer the question 'what works?'. Rather they try to explain in what respects, for whom, in what circumstances and why a programme or policy works.

> ❝ A realist approach recognises that programmes are not universally successful and work better in some circumstances and conditions than in others.❞

Realist evaluation uses mainly primary data. The evaluator ventures into the field and collects data in order to develop, test and refine a programme theory to explain for whom and in what circumstances and why an intervention or programme works. Realist evaluation is usually a multi- or mixed methods approach and usually involves both qualitative and quantitative data. The sorts of data that might be collected include (but are not limited to) routinely collected or survey data, interview data with stakeholders and participants, preferably using realist interview methodology (Pawson, 1996; Manzano, 2016); and documentary or observational data.

> ❝ Realist evaluation uses mainly primary data.❞

The term realist evaluation is also sometimes used when primary data is used to develop programme theory (a necessary first stage before testing and refining it). Depending on the nature of the data used, this can be the first stage of evaluation, developing the programme theory
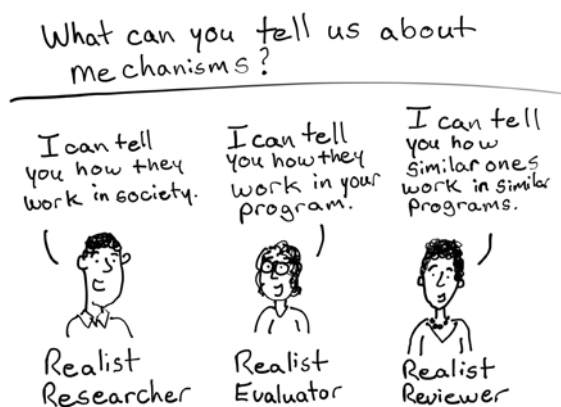
so that appropriate indicators and data collection tools can be developed in the next stage. Alternatively it can be the realist equivalent to an ex-ante evaluation, predicting the likely outcomes of a programme prior to its implementation.

> ❝ Realist synthesis uses mainly secondary data.❞

Realist synthesis uses mainly secondary data. It is a form of systematic literature review (Pawson, 2006). The process of *theory elicitation* (identifying the theory) is carried out by consulting any relevant material – policy documents, grey literature, editorials, think pieces – and stakeholders to identify the ideas and assumptions about how a programme is intended to work. The process of *theory testing* is undertaken by synthesising existing research into elements of the programme theory, including evaluations of programmes or interventions that share the same programme theory. This might include studies which identify outcomes (for example, outcome evaluations or research trials); studies which identify that interventions have a variable outcomes patterns (for example, sets of studies across a large programme or literature reviews); studies which compare interventions in different contexts (for example, comparative case studies); or those which examine the mechanisms through which a programme works (for example, case studies). The purpose is to synthesise findings from these studies and other relevant data to test and refine theories which explain in what circumstances and through what underlying causal processes interventions produce intended and unintended outcomes.

One of the features that distinguishes realist synthesis from other forms of systematic review is that the focus of the study is not a programme per se, but the programme theory. Because the same programme theory can be used in different kinds of programmes, the range of studies included can be much wider than some other kinds of reviews.

Sometimes realist synthesis and realist evaluation are combined in the same project. For example, a realist synthesis may be used to develop a programme theory and then primary data collected to test and further refine it. Other researchers may develop a programme theory through a realist synthesis, develop a programme to put it into practice, and then evaluate it through a realist evaluation.

To summarise, realist thinking is used to understand causation in both the material and the social world. This can be through realist evaluation, realist synthesis, and other forms of social research which are neither evaluation nor literature review. Realist evaluation and realist synthesis, however, usually focus on the evaluation of policies, programmes, initiatives and interventions.

References

Fitzpatrick S (2005) Explaining homelessness: a critical realist perspective. *Housing, Theory and Society* 22(1): pp. 1–17.

Manzano A (2016) The craft of interviewing in realist evaluation. *Evaluation*, 22(3), pp.342–360.

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

Pawson R (1996) Theorizing the interview. *British Journal of Sociology*, pp.295–314.

Pawson R (2008) *Causality for beginners*. In: NCRM Research Methods Festival 2008. http://eprints.ncrm.ac.uk/245/

Pawson R (2006) *Evidence based policy: A realist perspective*, London: Sage.

Westhorp G (2014) *Realist Impact Evaluation: an introduction* ODI Research and Policy in Development, Australian Government Department of Foreign Affairs and Trade.
www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/9138.pdf

Wong G, Greenhalgh, T, Westhorp G, Pawson R (2012) Realist methods in medical education research: what are they and what can they contribute? *Medical Education*, 46(1):pp.89–96.

Project team:

Professor Trish Greenhalgh
UNIVERSITY OF OXFORD
www.phc.ox.ac.uk/team/trish-greenhalgh

Professor Ray Pawson
UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/people/staff/pawson/

Dr Geoff Wong
UNIVERSITY OF OXFORD
www.phc.ox.ac.uk/team/geoffrey-wong

Dr Gill Westhorp
CHARLES DARWIN UNIVERSITY
www.cdu.edu.au/northern-institute/our-teams/603/5928#gillian-westhorp

Dr Joanne Greenhalgh
UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/people/staff/greenhalgh

Dr Ana Manzano
UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/people/staff/manzano/

Dr Justin Jagosh
UNIVERSITY OF LIVERPOOL
www.liv.ac.uk/cares

What realists mean by context; or,
Why nothing works everywhere or for everyone

The RAMESES II Project

A key premise of realist evaluation is that the mechanisms through which programmes work will only operate if the circumstances are right (Pawson and Tilley, 1997). For example, a tennis ball released in the air on earth (Context 1) falls to the ground (Outcome 1) due the force of gravity (Mechanism 1). Releasing a tennis ball in water (Context 2) results in it floating to the surface (Outcome 2) due to buoyancy (Mechanism 2). In fact, both gravity and buoyancy operate in both contexts but in air, the mechanism of gravity is stronger, whereas in water, the mechanism of buoyancy is stronger. The two mechanisms compete and the context (air or water) determines which 'wins' (Westhorp, 2014). Thus, it is "*the contextual conditioning of a causal mechanism which turns (or fails to turn) a causal potential into a causal outcome*" (Pawson and Tilley, 1997, p69).

> " …the mechanisms through which programmes work only operate if the circumstances are right…"

Realist evaluation applies this same reasoning to policies and programs. It seeks to understand "*why does a programme work in Wigan on a wet Wednesday and why does it then fail in Truro on a thunderous Thursday?*" (Pawson and Manzano-Santaella, 2012).

A particular challenge for realist evaluators is wrestling with the complexity of context (Pawson, 2013). Contexts are not just material and social but psychological, organisational, economic, technical, and so on. These different 'types' of context interact and influence each other. Moreover, contexts operate at all levels of systems, from the atomic to the cosmic, and the different levels also interact and influence each other. The challenge is to identify what is relevant for the particular investigation.

A deeper understanding of context

For policies and programmes, context describes those features of the situations into which programmes are introduced that affect the operation of programme mechanisms. The settings into which programmes are introduced do not, in and of themselves, constitute context in the realist sense. However, things about the way those settings operate can. Health promotion programs, for example, may be implemented in hospitals, General Practice (Family Medicine) clinics, prisons, and schools. There are differences *between* those types of settings that may affect how the programme works; but there will also be differences *within* each type of setting. The programme may work in some hospitals but not others, and some prisons but not others. A realist investigation seeks to identify what it is within the setting that affects whether and how the programme works. Perhaps it is whether the practitioners involved volunteered for the role or were directed to do it. Perhaps it is the motivation of the programme recipients, and how high the issue falls on their list of priorities. Perhaps it is something about the way power relationships work in the organisation.

In the social sciences, context also refers to the sets of "social rules, values, sets of interrelationships" that operate within times and spaces that either constrain or support the activation of programme mechanisms (Pawson and Tilley, 1997, p70). Sayer (2010, p75) defines context as the "material resources, social structures, including conventions, rules and systems of meaning in terms of which reasons are formulated". It is these systems of meaning, rules and sets of relationships that shape stakeholders' reasoning in response to programme resources and consequently, influence programme outcomes.

As Maxwell (2012) points out, realists are not just

arguing that causal relationships differ across contexts (they do) but are making a more fundamental claim that "the context within which a causal process occurs is, to a greater or lesser extent, intrinsically involved in that process" (Maxwell, 2012, p 40). In other words, context is inextricably enmeshed with the mechanisms through which a programme works. This means that we need to think of context in relation to particular mechanism(s), and as a constituent and interconnected element of a context-mechanism-outcome configuration, not as a separate entity. Novice realist researchers sometimes fall into the trap of producing a catalogue of contexts, without theorising how such contexts trigger particular mechanisms and thus outcomes (Pawson and Manzano-Santaella, 2012). The result is that explanatory elements become disconnected and do not explain how the context affects mechanisms and thus outcomes. Realist evaluations need data on contexts but crucially, also need analytic strategies to examine the interaction between context, mechanism and outcome.

For whom: context affects reasoning

Different people respond to the resources offered by programmes in different ways. This is in part because of the social norms affecting and adopted by different groups. The classic 'big picture' examples are culture, class, gender, religious and political beliefs, power, status and individual capacities – all things which shape what individuals value and how they reason in response to the programme's offerings.

For a specific example, take the case of programmes offering self-management support for people with diabetes. Newly diagnosed adolescent patients may struggle with self-management because they prefer to, or feel pressured to, fit in with peer or family activities like drinking alcohol or consuming sugary food. Of course, other peer groups and families may routinely engage in healthy eating and fitness behaviours which may support people to self-manage. As people grow into adulthood, they often become less dependent on peer and family approval and their 'reasoning' in relation to self-management may change.

Context affects how things are done, which influences how people respond

Programs are introduced into existing settings, and interact with existing policies, procedures, attitudes and beliefs, and priorities. These features of context affect how programmes are implemented, which in turn influences how people respond.

For example, in Spain, doctors are in charge of organ donation and consent processes and organ donation rates are high. In other countries, including the UK, with lower rates of organ donation, nurses are responsible for obtaining consent from families (Manzano and Pawson, 2014). It is not being a nurse or a doctor per se that explains these differences. Nurses have more limited power over decision making in professional hierarchies and systems of care and have less resources to manage the process, which likely underpins the variation in donation rates.

> "Programmes are introduced into existing settings, and interact with existing policies, procedures, attitudes and beliefs, and priorities. These features of context affect how programmes are implemented, which in turn influences how people respond."

Context affects whether resources are available for people to put new decisions into action

Context also affects whether resources are available for people to put new decisions into action. Prashanth et al (2014) showed that health managers in a remote area actively participated in a capacity building intervention designed to improve their skills in planning and managing health services and that they intended to make improvements. However, there were no changes in outcomes, as measured by indicators including budget utilisation, maternal antenatal care visits and infant still birth and mortality rates. The remoteness of the area meant that it was not a popular place for staff to work and there was a high turnover, which meant that managers had few staff to implement changes. In this example, it is not remoteness per se that explained the lack of success, but the effect of remoteness on staffing, which reduced the resources available to implement changes.

Contexts operate in time

There is a dynamic interplay between programmes and contexts which evolves and changes over time. Programmes may change the context in which they are implemented, which in turn may prompt changes to the resources offered by the programme, which subsequently shapes a different set of mechanisms and thus outcomes. Ball et al (2016) found that the aims and functions of referral management centres (RMCs) evolved over time in response to changes in the local service network that occurred as a result of their implementation. Some RMCs were originally developed to collect data on referral volumes by speciality and the availability of this information enabled health commissioners to introduce new services to deal with the demand. This in turn necessitated a change in the function of the RMC to one of clinical triage to ensure the 'right' patients were referred to the new services.

> " There is a dynamic interplay between programmes and contexts which evolves and changes over time."

Changes in the context over time can also constrain the intended mechanisms through which the programme is expected to work and thus limit its impact. For example, the majority of weight-loss programmes are well-known to fail in the long-term as the context experienced by the dieter (life events, social engagements, holidays) increasingly constrains motivation over time.

In summary, context is an essential component of any realist explanation of for whom, in what circumstances and why interventions or programmes 'work'. Contexts do not refer to places, people, time or institutions per se, but to the social relationships, rules, norms and expectations that constitute them, as well as the resources available (or not). Contexts are therefore bound up with the mechanism(s) through which programmes work, and need to be understood as an analytically distinct but interconnected element of a Context-Mechanism-Outcome configuration. The task of realist evaluators is to understand what in particular it is that functions as a context that shapes the mechanisms through which a programme works.

References

Astbury B and Leeuw FL (2010) Unpacking black boxes: mechanisms and theory building in evaluation, *American Journal of Evaluation* 31(3): pp.363–381

Dalkin S, Greenhalgh J, Jones D, Cunningham B, Lhussier M (2015) What's in a mechanism? Development of a key concept in realist evaluation. *Implementation Science*, 10:49: DOI 10.1186/s13012–015–0237–x

Dalkin S (2014) The Realist Evaluation of a Palliative Integrated Care Pathway in Primary Care: *What Works, For Whom and in What Circumstances*? Doctoral thesis, Northumbria University.

Gill M and Turbin V (1999) Evaluating 'Realistic Evaluation': Evidence from a Study of CCTV. In Painter K and Tilley N (eds) Surveillance of Public Space: CCTV, *Street Lighting and Crime Prevention*. Crime Prevention Studies Volume 10. Monsey, N.Y: Criminal Justice Press.

Pawson R and Tilley N (1994) What works in evaluation research?. *British Journal of Criminology*, 34(3), pp.291–306.

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

Pawson R (2008) Causality for beginners. In: NCRM Research Methods Festival 2008. eprints.ncrm.ac.uk/245/

Weiss C (1997) Theory based evaluation: Past present and future. *New Directions for Evaluatio*n (No 76). San Francisco, CA: Jossey Bass.

Westhorp G (2014) *Realist Impact Evaluation: An introduction ODI Research and Policy in Development*, Australian Government Department of Foreign Affairs and Trade. www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/9138.pdf

---

Project team:

Professor Trish Greenhalgh
UNIVERSITY OF OXFORD
www.phc.ox.ac.uk/team/trish-greenhalgh

Professor Ray Pawson
UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/people/staff/pawson/

Dr Geoff Wong
UNIVERSITY OF OXFORD
www.phc.ox.ac.uk/team/geoffrey-wong

Dr Gill Westhorp
CHARLES DARWIN UNIVERSITY
www.cdu.edu.au/northern-institute/our-teams/603/5928#gillian-westhorp

Dr Joanne Greenhalgh
UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/people/staff/greenhalgh

Dr Ana Manzano
UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/people/staff/manzano/

Dr Justin Jagosh
UNIVERSITY OF LIVERPOOL
www.liv.ac.uk/cares

Frequently Asked Questions about Realist Evaluation

The RAMESES II Project

1. What is the difference between realist evaluation and realist synthesis?

Realist evaluation is, as the name implies, a form of evaluation. Data is collected from primary sources. This can include field research, interviews, focus groups and documentary analysis. Realist synthesis – also known as realist review – is a form of literature review. Realist review can use published peer-reviewed articles, evaluation reports, other gray literature, existing data sets (for example census information) and in some cases, interviews to supplement the literature. Many realist projects involve a mix of realist evaluation and realist synthesis. For more information, see 'What's in a Name' in this series.

> ❝ Causation is at the heart of realist evaluation. Causation is about examining 'what causes what to happen"

This is an example of realists' recognition that in an open social system, there are multiple different causal mechanisms operating at different levels of systems. Those mechanisms interact in different ways in different contexts, which means that some people who experience poverty end up homeless. Furthermore, Fitzpatrick (2005) highlights that while positivists conceptualise poverty as a variable and measure what proportion of people experiencing poverty are homeless, realists ask "what is it about poverty that could cause homelessness?" Thus, realists seek to explain what it is about poverty (context) that gives rise to a causal pathway (mechanism) which leads to homelessness (an outcome).

2. How does realist evaluation differ from other types of theory-driven evaluation?

Realist evaluation uses realist causal explanation, which asserts that outcomes are caused by underlying mechanisms rather than directly by the programmes themselves (See 'What is a mechanism' in this series for further information). Whether or not these mechanisms operate depends on supportive or disabling contexts. Realist programme theory describes how the underlying mechanisms are expected to operate in different contexts and what outcomes will be generated if they do (or don't). Many other forms of programme theory do not address underlying causal processes. (See Funnell and Rogers, 2011 for examples of different kinds of programme theory).

3. Why would I want to use realist evaluation?

Realist evaluation is useful if you want to understand 'what works, for whom, under what circumstances and why?'. Realist evaluation is not required for evaluations that do not seek to answer questions about 'how' or 'why' programmes work the way they do.

4. What does the word 'causation' mean and why is it important in realist evaluation?

Causation is about 'what causes what to happen'. There are different understandings of causation in different evaluation paradigms. Some paradigms are based on what is called 'a successionist view'. In this view, a correlation between 'independent' variables (things that are varied or controlled – think 'input' variables) and 'dependent' variables (the outcome of interest) is examined. If the relationship is confirmed across many cases, causation is taken to be established. Realist evaluation assumes that there is something underlying

that relationship – a mechanism that causes the outcome – which will only operate if the circumstances are right. (See 'What is a mechanism?' in this series). Consequently, realists take the perspective that "what causes something to happen has nothing to do with the number of times we observe it happening" (Sayer, 2000). Realist evaluators are trying to do more than identify that there is a relationship – they are trying to explain it.

> **"Realist evaluators are trying to do more than identify that there is a relationship – they are trying to explain it."**

To take an example, a study could take 1000 university students and ask half of them to study the night before an exam, and the other half to not study the night before the exam. Then if you find that the students who studied the night before the exam did 80% better than students who did not, you would say you have evidence to suggest that studying the night before an exam improves exam results. This is based on a successionist view of causation that confirms A leads to B if we see it happening many times. The realist approach would seek to understand why it is the case that studying the night before an exam leads to the improved results. It could be because of better retention of information, or greater confidence and less stress, any or all of which could impact on performance. The realist investigation could also be used to try to explain what happened for those who studied but did not do better (lack of sleep, increased sense of inadequacy leading to greater stress and so on).

6. What kinds of questions should I pose in a realist evaluation?

Realist evaluation does not ask: 'Does this intervention work'? or even 'Did the intervention work this time?'. Those questions ask for an 'on average' answer. Realists assume that every programme will work for some people and in some contexts, but not others. The 'on average' answer will be hiding some for whom it did, some for whom it didn't, and potentially, some for whom it caused harm. Realist evaluations ask some or all of the elements of the question 'What works for whom, in what respects and to what extent, under what circumstances, and how?'. The question is not just 'does it make a difference' but 'what is it about X that makes a difference to whom, why (or why not)?' Understanding this helps inform decisions about whether and when to use particular kinds of interventions, and how to adapt them to local circumstances.

> **"Realist evaluations ask some or all of the elements of the question 'What works for whom, in what respects and to what extent, under what circumstances, and how?"**

7. What is a CMO configuration?

CMO configuration (sometimes written as CMOC) stands for context-mechanism-outcome configuration. It is the basic causal explanatory framework for realist evaluation and realist reviews. Stated as a sentence, it means 'In this context, this mechanism generates this outcome.' It is a tool to help remember what needs explaining: all outcomes are a result of interactions between contexts and mechanisms.

> **"CMO configuration (sometimes written as CMOC) stands for context-mechanism-outcome configuration"**

9. What does it mean to say that the CMO configuration is a heuristic?

A heuristic is a model or framework that is designed to support investigation or discovery, even if the framework itself is used in an incomplete way or is messy. In a realist evaluation, the CMO configuration is used in different ways. It can be used to frame evaluation questions; to predict outcomes in different circumstances (and thus to select outcome indicators to look for), and as a framework for analysis. The CMO as a heuristic is simply a reminder of the realist causal explanatory framework.

References

Funnell S and Rogers P (2011) *Purposeful program theory. San Francisco:* Jossey-Bass.

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

Project team:

Professor Trish Greenhalgh
UNIVERSITY OF OXFORD
www.phc.ox.ac.uk/team/
trish-greenhalgh

Professor Ray Pawson
UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/
people/staff/pawson/

Dr Geoff Wong
UNIVERSITY OF OXFORD
www.phc.ox.ac.uk/team/
geoffrey-wong

Dr Gill Westhorp
CHARLES DARWIN UNIVERSITY
www.cdu.edu.au/northern-institute/
our-teams/603/5928#gillian-westhorp

Dr Joanne Greenhalgh
UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/
people/staff/greenhalgh

Dr Ana Manzano
UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/
people/staff/manzano/

Dr Justin Jagosh
UNIVERSITY OF LIVERPOOL
www.liv.ac.uk/cares

Philosophies and Evaluation Design

The RAMESES II Project

There are many different schools of philosophy, each of which makes different assumptions about the nature of the world (ontology) – including how causation works the nature of knowledge (epistemology) and what constitutes 'value' (axiology). These assumptions have powerful implications for evaluation design, data requirements, analysis and so on.

Here we select just three broad schools of philosophy – positivism, constructivism and realism – and do gross injustice to each by over-summarising gross injustice to each by over-summarising them. We discuss both ontology – the philosophy of the nature of reality – and epistemology – the philosophy of knowledge.

> " There are many different schools of philosophy, each of which makes different assumptions about the nature of the world, the nature of knowledge, what constitutes 'value' and how causation works."

Positivist *ontology* holds that there is an objective reality, existing independently of human interpretation, and governed by natural laws. Positivist *epistemology* holds that the observer and the observed are independent of each other, and that we can identify 'facts' about reality through a cycle of observation and theory. Only that which can be observed is considered a valid source of knowledge. The role of the positivist evaluator is, therefore, to identify and report observable facts. Post-positivism shares the same assumption about an independent reality, but accepts that the theories, values, experiences and perspectives of the researcher will influence what is observed and how it is interpreted. It relies on triangulation and dispute to come 'as close as possible' to objectivity. The logic that underpins randomised control trials is (post)positivist, as is the logic that underpins meta-analytic reviews. (For a readable discussion of the

difference between positivism and post-positivism, see www.socialresearchmethods.net/kb/positvsm.php)

Constructivism is an epistemology (philosophy of knowledge). It lacks a clear ontology. It asserts that all knowledge is both socially and individually constructed and interpreted: as a logical consequence, we can never be quite sure what reality is like or even whether it exists. Neither (in many forms of constructivism) can we prove what is 'true' or 'not true': 'facts' are just things that are accepted to be true (at least by many in a particular context). Many constructivists are primarily concerned with the meanings and interpretations that human actors give to experiences and use exclusively qualitative methods to investigate them. The role of the constructivist evaluator is therefore to identify and report these meanings.

Realism sits somewhere between the two, closer to post-positivism in ontology; closer to some forms of constructivism in epistemology[1]. What distinguishes realism is its particular understanding of how causation works. The role of the realist evaluator is to explain how and why programmes or policies cause their various outcomes in different sets of circumstances.

> " What distinguishes realism is its particular understanding of how causation works."

Key assumptions in realist philosophy and their implications for evaluation design[2].

1. *Both the material and the social worlds are 'real': anything that can have real effects is itself real.* Thus, for example: gender – as distinct from sex – is real, and we know that because it has real effects. That is to say: while sex is biological and gender is socially constructed, both have real effects. Constructions of gender vary across time,

culture, and sub-group, and therefore so do their effects – but the effects are still real, and therefore, so is gender. Other examples include culture, class, and ideas: they are real even though they are not material objects. This means, firstly, that programmes and policies are real and can have real effects; and secondly, that social institutions and constructs (culture, class, gender, religion, economic systems...) can have real effects on whether and how programmes work. Realist evaluations are designed to understand what these effects are and how and why they vary across contexts.

2. ***Mind-independent reality***. Both the natural and the social worlds are independent of, and inter-dependent with, our understandings of them. For realists, natural and social systems (think the banking system, the industrial relations system…) exist and exert their own powers, despite the fact that different humans make different judgements of them. In this sense, they are independent. However, humans can affect natural or social systems, and they are affected by them: in this sense, they are inter-dependent. Therefore, both the powers and liabilities of the systems and the interpretations that people make may be relevant in an evaluation.

3. ***All enquiry and all observations are shaped and filtered through the human brain. There is, therefore, no such thing as 'final' truth or knowledge.*** While we can never reach absolute certainty, realism argues that it is possible to work towards a closer understanding of the nature of reality, because reality itself constrains the interpretations that can reasonably be made of it. We can make observations and conduct tests that help us to judge between competing interpretations. Evaluations can therefore work towards better understanding of whether, how and why programmes work, but can never provide 100% 'proof' of any conclusion.

4. ***All social systems are open systems.*** Their boundaries are porous and flexible: people, ideas, information and resources flow in and out. Different social systems interact and influence each other. Further, the systems will change over time, in complex and interactive ways – regardless of whether a programme or policy is introduced. Perhaps most importantly: any outcome that is observed will be a result of interactions within and across systems – not simply an outcome of the programme or policy.

Evaluations will only ever be able to show that a policy or programme *contributed* to an outcome.

5. ***Realism offers a particular understanding of how causation works.*** The basic idea is that things that we experience or can observe are caused by 'deeper', usually non-observable processes. (See 'What is a mechanism?' in this series). Realist evaluations try to identify the mechanisms that cause programme outcomes, not just an association between 'the programme' and 'the outcome'.

" Realist evaluations try to identify the mechanisms that cause programme outcomes, not just an association between 'the programme' and 'the outcome'."

6. ***Realism provides a specific way of thinking about 'context'.*** Whether mechanisms generate outcomes depends on the context. If I am standing on land (context) when I release a tennis ball, gravity (mechanism) will draw it to the ground (outcome). If I happen to be underwater (context), buoyancy (mechanism) will cause the ball to float (outcome). Realist evaluations identify *what it is* about the context that determines whether, and which, mechanisms 'fire'.

7. ***Mechanisms, operating in contexts, generate outcomes.*** Because different mechanisms operate in different contexts, programmes will generate different outcomes for different groups in different contexts. Realist evaluations seek to identify and explain how and why different outcome patterns are generated.
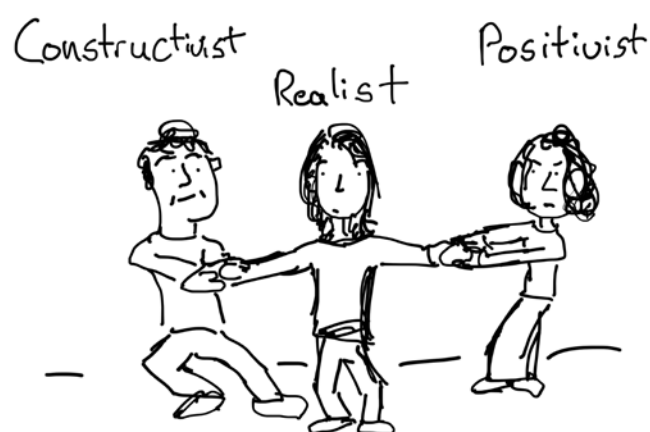
Further reading

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

Westhorp G (2014) *Realist Impact Evaluation: an introduction* ODI Research and Policy in Development, Australian Government Department of Foreign Affairs and Trade.
www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/9138.pdf

Project team:

Professor Trish Greenhalgh
UNIVERSITY OF OXFORD
www.phc.ox.ac.uk/team/trish-greenhalgh

Professor Ray Pawson
UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/people/staff/pawson/

Dr Geoff Wong
UNIVERSITY OF OXFORD
www.phc.ox.ac.uk/team/geoffrey-wong

Dr Gill Westhorp
CHARLES DARWIN UNIVERSITY
www.cdu.edu.au/northern-institute/our-teams/603/5928#gillian-westhorp

Dr Joanne Greenhalgh
UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/people/staff/greenhalgh

Dr Ana Manzano
UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/people/staff/manzano/

Dr Justin Jagosh
UNIVERSITY OF LIVERPOOL
www.liv.ac.uk/cares

A realist understanding of programme fidelity

The RAMESES II Project

The traditional idea of 'programme fidelity' is used in two different circumstances. One is in experimental designs which aim to assess whether X (the programme) leads to Y (a desired outcome), and to what extent it does so. The aim here is to establish that 'X causes Y'. The other is after an 'evidence-based programme' has been developed. It has already been 'established' that 'X causes Y'; the programme is now being implemented more widely. Here, the aim is to ensure that 'the same' programme is implemented, on the assumption that if the same things are done, the same outcomes will be achieved. Evaluations of this type tend to focus on detailed process evaluation: were the right things done, in the right order, with the right people, to the right extent… and so on. Traditionally, programme fidelity is the measure of the standardisation process, that is, the "extent to which delivery of an intervention adheres to the protocol or program model originally developed" (Mowbray et al, 2003).

> " Traditionally, programme fidelity is the measure of that standardisation process, that is, the "extent to which delivery of an intervention adheres to the protocol or program model originally developed"."

In experimental designs, fidelity is a holy grail. Testing the proposition that "X causes Y" requires that X is always the same; the intervention should be identical in every way and in every application. In its purest form, it also requires a closed system, so that it is possible to isolate the effects of 'X' from everything else that might impact on 'Y'. Measuring the extent of impact also requires a counterfactual – circumstances where 'X' does not apply (but that is a separate issue which is not addressed here.)

Randomised controlled trials (RCT) of medications are the usual example for this logic. The manufacturing process turns out uniform copies of the drug and so it can be assumed that X (the treatment) is always the same. To show that the medication leads to its intended outcome, people can be randomly allocated to the medication or not.

A realist evaluation approach makes different assumptions. The medication may be the same, but the same amount of the same medication does different things in different people (which is why side effects vary and medications for adults are not necessarily appropriate for children). Even more importantly for realists, implementation and effectiveness still involve human volition. Patient compliance with medication prescriptions varies considerably (Nieuwlaat et al 2014). Expectations of effectiveness may increase the likelihood of desired outcomes (the placebo effect). Staff handing out the medication may consciously or unconsciously affect the desired outcome, and so on.

People running trials respond to these potential problems with supplementary methods to ensure and check whether patients are taking the medications as prescribed and whether, for example, staff conscious and unconscious biases are eliminated. This follows the classic empirical formulation 'specify-standardise-check': precisely design and specify an intervention and take steps to ensure that the actual intervention is faithful to plan by making sure those steps are always the same and are always followed.

This 'specify-standardise-check' approach is more contentious in the evaluation of social programmes and policies, applied healthcare and psychological interventions, and so on. As with medication trials, the logic has been used both during research to test programme effectiveness and in process-fidelity evaluations. Despite the best implementation efforts, it is impossible to standardise the delivery of complex interventions. The personnel responsible for delivering the intervention (the police officers, psychologists, field staff) are diverse, as are the contexts in which they work (the crimes spots, their psychological models,

the countries they work in). There are differences among those receiving the intervention (the criminals, the people who need counselling, the communities being developed) in terms of their circumstances and capacities. And there are differences in the implementing organisations (organisational settings, culture, and resources) all of which shape how the programme is delivered and responded to (see 'Context' in this series).

❝ Context is not a set of variables which can be 'controlled'."

Realists also argue that context is not a set of variables which can be 'controlled'. Rather, it is the interaction between contexts at individual and social levels and aspects of the programme that determines whether, and shapes how, the programme mechanisms work. This in turn shapes the intervention outcomes. From a realist perspective, people who share the same measurable individual characteristics can respond to the intervention in different ways and thus generate different outcomes. Furthermore, realists argue, it is impossible to equalise all aspects of context between two groups; and without understanding *what it is about the context* that matters (that is, what aspects of context affect which mechanisms), it is not possible to determine which elements of context should be equalised between trial and control groups.

Programme fidelity from a realist perspective

Realist evaluation takes a different approach to programme fidelity. Realists do not expect standardised programmes. Programmes start to change and adapt as soon as implementation begins, because of differences in local conditions, funding, programme support systems, political impetus, staff turnover, policy fashions, attempts to turn the ideal model into a feasible model on the ground, and so on.

Sometimes adaptations represent attempts to tailor the programme to specific contexts in order to enable the programme to work. Other times, adaptations reflect resource constraints, a lack of engagement or attempts to 'work around' a programme deemed 'unworkable' by key stakeholders. Realist evaluation seeks to understand when and why these adaptations occur and the nature of the adaptations. 'Fidelity failures' do not necessarily lead to programme failure (in fact, they can in some circumstances improve outcomes). Dfferences in implementation become a source of data for understanding how and why programmes do or do not work.

❝ Fidelity should be re-articulated in terms of programme theory."

In summary, the concept of fidelity, as traditionally understood, is unhelpful in realist evaluation. Fidelity should instead be re-articulated in terms of programme theory: fidelity to underlying causal processes, context sensitivity, and adaptation. In other words, in a realist evaluation the goal is to understand and explain how, why, for whom, in what contexts and to what extent outcomes have occurred, compared to what might have been expected based on the programme theory. Variation is a resource that can be used to help understand what works, in what circumstances, for whom and why.

An Example

A new programme aims to implement guidelines for stroke care coordinators to improve rehabilitation outcomes. Care coordinators are expected to meet with people who have had a recent stroke and their families and go through the exact process specified in the guidelines so that the desired outcome (increased rehabilitation for people with strokes) is achieved. When these programmes are implemented in different wards, the care coordinators adapt them to their local circumstances. The standardised model may be adapted because it is impossible to follow the model strictly (e.g. if it contradicts other guidelines). Or some of the questions may be skipped or adapted (e.g. those related to sexual activity or advanced end of life decisions may not be relevant).

A realist investigation will explore how contexts affect the work done by, and outcomes from, different care coordinators in different services. For example, novice staff are more likely to engage with the new guidelines because they provide a safe tool and offer a perception of following quality professional standards. Senior care coordinators may not feel they need that safety net and continue to work according to their usual practice.

So we just use the hose to precisely control the amount of water each plant receives. But it's raining!

References

Hope T (2009) The illusion of control: A response to Professor Sherman. *Criminology and Criminal Justice*, 9(2), pp.125–134.

Mowbray C, Holter M, Teague G, & Bybee D (2003) Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation, 24*, pp. 315–340.

Nieuwlaat R, Wilczynski N, Navarro T, Hobson N, Jeffery R, Keepanasseril A, Agoritsas T, Mistry N, Iorio A, Jack S, Sivaramalingam B, Iserman E, Mustafa RA, Jedraszewski D, Cotoi C, Haynes R. (2014) Interventions for enhancing medication adherence. *Cochrane Database of Systematic Reviews* 2014, Issue 11. Art. No.: CD000011. DOI: 10.1002/14651858.CD000011.pub4

Paul G (1967) Strategy of outcome research in psychotherapy, *Journal of Consulting Psychology*, 31(2):pp.109–118.

Project team:

Professor Trish Greenhalgh
UNIVERSITY OF OXFORD
www.phc.ox.ac.uk/team/
trish-greenhalgh

Professor Ray Pawson
UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/
people/staff/pawson/

Dr Geoff Wong
UNIVERSITY OF OXFORD
www.phc.ox.ac.uk/team/
geoffrey-wong

Dr Gill Westhorp
CHARLES DARWIN UNIVERSITY
www.cdu.edu.au/northern-institute/
our-teams/603/5928#gillian-westhorp

Dr Joanne Greenhalgh
UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/
people/staff/greenhalgh

Dr Ana Manzano
UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/
people/staff/manzano/

Dr Justin Jagosh
UNIVERSITY OF LIVERPOOL
www.liv.ac.uk/cares

The Realist Interview

The RAMESES II Project

Realist interviews are theory-driven interviews, meaning that theory should be used explicitly and systematically throughout the interview process. Realist interviews are qualitative in nature: participant views are explored through conversations.

However, the purpose of the realist interview is different to other types of interviews. In constructivist interviews, researchers tend to explore participants' views and experiences of the topic under investigation. The aim is to elicit and understand the respondent's world view and experiences. Realist interviews investigate propositions about how, where, when and why programmes are and are not effective. It is the programme's story that is pursued. This is achieved by capturing the participants' stories about the programme because those experiences can illuminate the varying processes (mechanisms and contexts) and manifold outcomes of the programme. To do so the interviewer should relate with respondents in a distinctive process called the 'learner-teacher cycle'.

> **" Realist interviews investigate propositions about how, where, when and why programmes are and are not effective. It is the programme's story that is pursued."**

The teacher-learner cycle

In realist interviews, the subject matter of the interview is the programme theory (or some aspect of it). Theories are placed before the respondents for them to comment on, with a view to confirming, denying and refining the theory. The realist premise is that the research process will usually start by theorising, then will test those theories; the theories will be refined and tested again. In this iterative process, our understanding of the real world is also refined.

In realist studies, qualitative data are not considered simply to be constructions (as they would be in some

kinds of constructivist interviews). While the subjective nature of human knowledge is acknowledged, data are considered to be evidence for real phenomena and processes, and are used to make inferences about these processes. The realist investigator tries to understand how each fragment of evidence contributes to their interpretations and explanations and how their ideas are tested and refined within and between those fragments. The interview is one of many testing tools.

While realist interviews can be used to develop programme theory, the teacher-learner cycle usually involves teaching the interviewee the particular programme theory under test. This implies a key pre-condition: that the evaluator has already established those theories. The realist investigator should also, whenever feasible, arrive at the interviews armed with some understanding about what happens in the natural setting, because it supports a conversational style of interview. This understanding can be more or less well developed depending on resources and time constraints. It may be grounded in previous experience in a sector or issue area, come from prior interviews (for example, while developing the programme theory), and/or from synthesising grey and academic literature and other data sources.

Interviews should start with general questions about interviewees' role in, experiences of and views about the programme. Subsequent questions follow up their responses, asking them to tell their stories about specific experiences or issues with the programme, its participants, and constraints. Each of these questions investigates something about the programme theory being tested. Specific elements of programme theory can then be explicitly introduced and tested with the respondent. In order to avoid allegations of 'leading the interview', it is a good idea to test multiple, including contradictory, theories about the same aspect of the programme with the same respondent.

In the first round of interviewing, questions will be mainly

exploratory. Questions will try to ascertain how the programme works for whom and in what circumstances. As the evaluator becomes more knowledgeable about programme nuances, the questions evolve. They become less standardised and more tailor-made to refine specific context-mechanism-outcome (CMO) configurations. The ideas in these configurations are made explicit to the interviewee, although not necessarily described as "context", "mechanism" and so on. The respondent can then teach the evaluator about those specific components of the programme in their circumstances and help the refinement process or they can refute the interviewers' hypotheses with their examples.

The theory-testing purpose of evaluation shapes the interview relationship. The roles of teacher and learner are not static, but fluid. At some stages the interviewer is teacher ("here is an element of programme theory") and at other times the interviewee is ("and here's how it does or doesn't work here"). The idea is that the interview evolves into a discussion, thinking through the complexities of the programme.

Theories will be surfaced, refined and finally consolidated. Sometimes these three processes will happen within the same evaluation – or the same interview – but due to constraints, some evaluations or interviews may have to stop at theory gleaning or early stages of refinement. Other evaluations start from a more advanced base and move from theory refinement to consolidation.

How many realist interviews are enough?

Realist hypotheses are not confirmed or abandoned through saturation in interviews, but using multiple types of data obtained in a multi-method strategy. A theory may be gleaned, refined or consolidated not just in the next interview(s), but also while digging for nuggets of evidence in other sources of data (e.g. documents, routinely collected administrative data).

At the design stages of realist evaluation projects, it can be difficult to establish a definite number of interviews required. Evaluators become more knowledgeable of programme successes and barriers as they start conversing with staff and stakeholders. It is at that point that theories take firmer shape and a more definitive interview sample can be developed.

However, more is not necessarily better. Because the unit of analysis is not the person, but the events and processes around them, every respondent (every staff member, every participant) can uncover a collection of micro events and processes, each of which can be explored in multiple ways to test theories. Depending on the evaluation design, this can mean that a relatively small number of participants

with detailed knowledge of the programme are interviewed multiple times to fully test and refine the initial programme theories, rather than interviewing many people once.

Who to interview and who knows what?

Sampling for realist interviews is theory based. That is, respondents are selected because they are in a position to cast light on a hypothesis or a particular aspect of programme theory.

" Sampling for realist interviews is theory based."

Different points of views must be pursued, not to ensure 'balance' or to achieve 'consensus', but because different perspectives are needed to investigate patterns and uncover unintended outcomes. Specifically, respondent selection should be based on their 'CMO investigation potential'. It is necessary to work out 'who can tell you what' about context, mechanisms and outcomes.

For initial theory gleaning, it is better to start by interviewing practitioners rather than users of the programme: people who know the programme well. In some programmes, this will be those whose job it is to monitor what goes on (e.g. middle managers, ward managers); in others, it might be senior practitioners.

Frontline practitioners could be the next set of interviewees because they are good sources of information about the participants for whom interventions do and don't work, and about programme barriers and unintended consequences of implementation at the 'coalface'. Different practitioners will have different experiences and this may shed useful light on multiple aspects of programme theory.

Participants of the programme (e.g. service users, patients, claimants) can generally talk about their own outcomes and the change processes to which they attribute those outcomes. They can also tell you about their own circumstances, which may be relevant to the 'individual' components of context.

If the programme theory suggests that the programme will work differently for particular sub-groups or in particular sets of circumstances, it is important to ensure that those sub-groups or contexts are represented in the sample. It is also important to ensure that the evaluator can identify which set of interview data relates to which sub-group or context. Without that, realist analysis is all but impossible.

Westhorp and Manzano have developed a 'starter set' of questions for realist evaluation interviewing (See 'Realist Evaluation Interviewing – A 'Starter Set' of Questions' in this series).

Further Reading

Manzano A (2016) The craft of interviewing in realist evaluation. *Evaluation* 22(3): pp.342–360

Pawson R (1996) Theorizing the interview. *British Journal of Sociology*, pp.295–314.

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

I have an idea. Let's all show our cards. It will be much easier to decide if we should fold.

Project team:

Professor Trish Greenhalgh
UNIVERSITY OF OXFORD
www.phc.ox.ac.uk/team/
trish-greenhalgh

Professor Ray Pawson
UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/
people/staff/pawson/

Dr Geoff Wong
UNIVERSITY OF OXFORD
www.phc.ox.ac.uk/team/
geoffrey-wong

Dr Gill Westhorp
CHARLES DARWIN UNIVERSITY
www.cdu.edu.au/northern-institute/
our-teams/603/5928#gillian-westhorp

Dr Joanne Greenhalgh
UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/
people/staff/greenhalgh

Dr Ana Manzano
UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/
people/staff/manzano/

Dr Justin Jagosh
UNIVERSITY OF LIVERPOOL
www.liv.ac.uk/cares

Retroduction in realist evaluation

The RAMESES II Project

What is retroduction?

There are different forms of reasoning in research and evaluation. The two most commonly described are induction, or inductive reasoning, and deduction, or deductive reasoning. Inductive logic derives conclusions from multiple observations: it builds theory from observation. Deductive logic starts from theory and tests propositions by seeing whether associations match expectations.

Realist research and evaluation uses 'retroduction'. Retroduction refers to the identification of hidden causal forces that lie behind identified patterns or changes in those patterns. It asks the question: 'why do things appear as they do?'(Olsen, 2010). The pre-fix 'retro-' comes from the Latin to mean, 'behind, beneath'.

> " Retroduction refers to the identification of hidden causal forces that lie behind identified patterns or changes in those patterns."

"Retroduction entails the idea of going back from, below, or behind observed patterns or regularities to discover what produces them"
(Lewis-Becket al, 2004)

The same idea is expressed clearly by Sayer (2000):

"Merely knowing that 'c' has generally been followed by 'e' is not enough: we want to understand the continuous process by which 'c' produced 'e', if it did. This mode of inference in which events are explained by postulating (and identifying) mechanisms which are capable of producing them is called 'retroduction'" (p 207).

> " Retroduction uses both inductive and deductive logic, as well as insights or hunches."

Retroduction uses both inductive and deductive logic, as well as insights or hunches. It involves thinking through what causal powers might be at work in producing observed patterns or changes in patterns. It is underpinned by a belief that an understanding of causation cannot be achieved using only observable evidence. Retroductive theorising requires that inquirers use their common sense, intelligence, expertise, and informed imagination to build and test theories about underpinning causal processes. These may not be able to be tested immediately: truly novel theories often precede the means or technologies to test them.

Charles Darwin's theory of species evolution via natural selection is an exemplar of retroduction. The theory of natural selection suggests that over time, individuals less suited to the environment are less likely to survive and less likely to reproduce. Individuals more suited to the environment are more likely to survive and more likely to reproduce and leave their heritable traits to future generations, which over time constitutes the process of natural selection. This slowly effected process results in populations changing to adapt to their environments, and ultimately, these variations accumulate over time to form new species (Mayr, 1982).

Darwin did not himself identify with realism or retroduction, yet his theory of natural selection explains the underpinning forces of evolution that explain species formation, evolution, and biodiversity. The theory helps us to 'see' evolution, even though it is difficult to observe. His theory was not produced through inductive and deductive reasoning alone. His was a deeply inspired view of evolutionary change which featured mechanisms that are not readily observable.

Retroduction in realist evaluation

In realist evaluation, the retroductive question is about the causal powers of the policy, intervention or programme, given the circumstances in which it is applied. How is it that Intervention X can produce outcomes $Y_{1..n}$ given conditions $Z_{1...n}$? Retroduction is used to theorise programmes and formulate context-mechanism-outcome configurations.

> **"** In realist evaluation, the retroductive question is about the causal powers of the policy or programme."

Where a programme theory has been developed in advance, the process starts with deductive reasoning (seeking evidence to test the theory). Cases are examined, preferably to the point of saturation, checking that the patterns of success and failure, intended and unintended outcomes are consistent with the theory. Inconsistent cases may require the theory to be refined. That is, new theory is generated on the basis of observations, or inductive reasoning. That new theory is then put to the test in further cases (deductive reasoning again). The process of confirming, refuting and refining the theory continues as further instances are examined. That is, retroduction moves back and forth between inductive and deductive logic.

Retroduction is also used in realist reviews and wider realist research. Bhaskar, in *The Possibility of Naturalism*, asked: 'what properties do societies possess that might make them possible objects of knowledge for us?' (Bhaskar, 2015 p. 25). The question is inherently retroductive because it assumes that societies have underlying causal properties and that the task of realist inquiry is to understand those properties.

References

Bhaskar R (1979; 3rd edition 2015.) *The Possibility of Naturalism: a philosophical critique of the contemporary human sciences*, Oxon: Routledge.

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

Lewis-Beck M, Bryman A & Liao T (eds.) (2004) Encyclopedia of Social Science Research Methods, Thousand Oaks: SAGE.

Mayr E (1982) *The Growth of Biological Thought*, Cambridge: Harvard University Press.

Olsen W (2010) Editor's Introduction: Realist Methodology – A Review. *In:* OLSEN, W. (ed.), Los Angeles: Sage.

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

Project team:

Professor Trish Greenhalgh
UNIVERSITY OF OXFORD
www.phc.ox.ac.uk/team/trish-greenhalgh

Professor Ray Pawson
UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/people/staff/pawson/

Dr Geoff Wong
UNIVERSITY OF OXFORD
www.phc.ox.ac.uk/team/geoffrey-wong

Dr Gill Westhorp
CHARLES DARWIN UNIVERSITY
www.cdu.edu.au/northern-institute/our-teams/603/5928#gillian-westhorp

Dr Joanne Greenhalgh
UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/people/staff/greenhalgh

Dr Ana Manzano
UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/people/staff/manzano/

Dr Justin Jagosh
UNIVERSITY OF LIVERPOOL
www.liv.ac.uk/cares

Realist Evaluation and Ethical Considerations

The RAMESES II Project

Realist evaluation is a form of primary research and will usually involve human participants. It is important that evaluations are conducted ethically. Ethical conduct includes but is not limited to complying with professional ethical standards and local ethics regulatory requirements. Approval by an official body is often required before collecting data from participants and in some cases staff. Realist evaluators come from a range of different professional backgrounds and work in diverse fields. Evaluators should ensure that they are aware of and comply with their professional obligations and local ethics requirements throughout the evaluation.

> " Ethical conduct includes but is not limited to complying with professional ethical standards and local ethics regulatory requirements."

A common question is whether realist evaluation studies have particular ethical implications compared to other evaluation designs. There are a number of assumptions about the relationship between evaluation approaches and ethical considerations. For example, some assume that positivist driven methods like some types of questionnaires have fewer and more manageable ethical issues than constructivist data collection tools (interviews, observations, visual methods). Others propose that constructivist methods are more ethical, especially those that include participatory approaches and try to minimise power differentials between evaluator and participants. Experienced evaluators know that ethical considerations arise with every data collection technique. They are also aware that conflicts of interest and ethical dilemmas may arise at every stage of the evaluation (Leone et al, 2016): the commissioning and contracting stage, design, programme theory development, data collection and analysis, presentation of results, utilisation of findings, and so on.

In all types of evaluation approaches, ethical issues emerge over time, can be generated by the evaluation, and can change as the evaluation progresses. Because realist evaluation is methodologically iterative and emergent (Pawson, 2013), it can be difficult to specify some aspects of how things will be done before the start of the evaluation. For example, the sampling criteria, how many interviews will be done, exactly who will be interviewed and what each of the stakeholder topic guides will look like may not be known at the beginning of the project. Even when they have been planned, they may change as findings emerge and theory is refined. One challenge that all realist evaluations face when seeking formal ethical approval is that legitimate changes may be required to the methods used and participants recruited as the evaluation evolves.

> " Because realist evaluation is methodologically iterative and emergent, it can be difficult to specify some aspects of how things will be done before the start of the evaluation."

The key realist principle to remember when thinking about the ethical implications of realist evaluation is how the project is "ethical for whom, in what contexts, in what ways and how?" (Williams and Westhorp, 2016). Proposals for ethical approval should clearly distinguish the implications of the evaluation for different groups (e.g. programme architects, commissioners, programme managers, frontline managers, service providers, patients or participants, caregivers, evaluators, and so on) and different contexts (across institutions, locations and time spans). This includes particular attention for vulnerable groups (e.g. children, prisoners, participants who are incapable of giving consent). Both benefits and risks for each group and each context should be considered: approval depends on the balance between risk and benefit.

> **The key realist principle to remember when thinking about the ethical implications of realist evaluation is how the project is "ethical for whom, in what contexts, in what ways and how?""**

Realist evaluators should be able to explain to an ethical review committee how and why their methodological decisions have been made, in terms that committee members will understand. Specific implications of the realist methodology will need to be explained. This is likely to include the need to link data across contexts, mechanisms and outcomes and specific strategies to address those implications should be included. For example, it might be necessary to be able to link different data types for the same participant, meaning that the data must be stored in such a way that those linkages can be made. However, depending on the nature of the data and the size of the sample, this may increase the risk of inadvertently identifying informants. This becomes an even bigger issue if data is being stored for future re-use or made available for other researchers later. Therefore, data should be anonymised before they are permanently stored and research data should be stored separately from personally identifying data.

Anticipating that changes may be needed is important when seeking ethical approval. Flexibility may need to be built into the project to allow for iteration to occur without the need to apply for an amendment to ethical approval while at the same time identifying what sorts of more major changes would require such amendments to be made. These distinctions will need to be explained and discussed with those who provide and monitor ethical approval.

Realist evaluation proposals should identify strategies for iteration included in the design and the steps that will be taken to manage ethics in relation to iteration. This flexibility implies a number of practical issues to take into consideration: costs of running the study, time-span of the research including the need for ethic committee approval amendments, and technical issues (language used in questionnaire items, realist interview guides, and so on).

> **Realist evaluation proposals should identify strategies for iteration included in the design and the steps that will be taken to manage ethics in relation to iteration."**

Another possible issue for attention is the role of the evaluator in relation to other stakeholders and the programme itself. Realists often walk a fine line, neither completely external and 'objective' nor internal. They may contribute to developing programme theory; bring research knowledge to the table; put aspects of theory to stakeholders or participants in interviews; and work with varying degrees of participation by stakeholders in evaluation design, data collection and analysis. These roles have power implications that should be thought through and addressed.

In summary, during the commissioning, design, conduct and impact of realist evaluations, ethical issues will emerge and transform over time. The evaluator should tailor what s/he does in order to achieve intended outcomes while maximising the benefits and minimising the risk of harms for different people and organisations in different circumstances.

References

Leone L, Stame N and Tagle L (2016) Exploring ethical issues and conditions for institutionalizing evaluation in the public sector. Evaluation 22(2): pp.149–167.

Pawson R (2013) The Science of Evaluation: A Realist Manifesto, London: Sage.

Williams E and Westhorp G (2016) Realist Ethics Roundtable. 2nd International Conference on Realist Evaluation and Synthesis, 3rd to 5th October 2016, London, UK.
www.liverpool.ac.uk/media/livacuk/instituteofpsychology/researchgroups/CARES,2016,Book,of,Abstracts.pdf

Further Reading

Henry G and Julnes G (1998) Values and realist evaluation. New Directions for Evaluation 78: pp.53–71.

Project team:

Professor Trish Greenhalgh
UNIVERSITY OF OXFORD
www.phc.ox.ac.uk/team/trish-greenhalgh

Professor Ray Pawson
UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/people/staff/pawson/

Dr Geoff Wong
UNIVERSITY OF OXFORD
www.phc.ox.ac.uk/team/geoffrey-wong

Dr Gill Westhorp
CHARLES DARWIN UNIVERSITY
www.cdu.edu.au/northern-institute/our-teams/603/5928#gillian-westhorp

Dr Joanne Greenhalgh
UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/people/staff/greenhalgh

Dr Ana Manzano
UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/people/staff/manzano/

Dr Justin Jagosh
UNIVERSITY OF LIVERPOOL
www.liv.ac.uk/cares

Protocols and realist evaluation

The RAMESES II Project

Introduction

The Medicines for Human Use (Clinical Trials) Regulations 2004 defines a protocol as "A document that describes the objectives, design, methodology, statistical considerations, and organisation of a clinical trial". In the last 20 years, the production and publication of protocols for randomised controlled trials (RCTs) and systematic reviews (SRs) has increased exponentially. More recently, the enthusiasm for protocols has been extended to other study designs, including realist evaluation and realist synthesis. Thus, in the last 5 years, there have been a number of protocols published for realist evaluations (REs) (e.g. Dalkin et al, 2012; Randell et al, 2014; Mukumbang et al 2016; Mirzoev et al 2016). Unlike the RCT literature, relatively little has been written concerning the role of protocols in realist evaluations and to date there has been no

> " …relatively little has been written concerning the role of protocols in realist evaluations…"

research to compare protocols of realist evaluations with subsequent papers reporting their findings. Here we consider the extent to which the assumptions underlying the production of protocols for RCTs can be applied to protocols for realist evaluation and suggest the amendments that are necessary for realist evaluations.

Why write a protocol?

There are several ideas and assumptions underlying the publication of protocols. For RCTs, these include reducing the potential for observer and researcher bias by ensuring that methods are appropriately designed and holding authors to account for following the protocol to reduce outcome reporting bias (McNamee, 1997; Jones and Abbasi 2004). Protocols are also argued to enable the replication of the study, which in some quarters

is considered to be the hall mark of 'good science' (Loscalzo, 2012). Protocols are also hypothesised to act as a planning and risk assessment tool, enabling teams to identify potential risks to the study and put in place plans to address these prior to the beginning of the study (Shamseer et al, 2015). The publication of a protocol is expected to stimulate debate amongst the academic community regarding the validity and value of specific studies or of research methodologies more generally, and thus advance knowledge. Finally, they are also hypothesised to serve as a means of reducing duplication of effort by alerting other research groups to ongoing research (Shamseer et al, 2015).

Protocols and validity: Closed systems vs open systems

The validity and credibility of RCTs rests on making decisions about key aspects of trial design to avoid influence from stakeholders on these decisions during the trial. Such influences are assumed to threaten the creation of a closed system, in which the effects of an intervention can (at least in theory) be isolated from its context in order to understand the contribution that the intervention alone makes to outcomes.

However, realists argue that it is not possible to create closed social systems based on experimental manipulation. Rather than treat contextual influences as contamination or bias, realist evaluators seek to understand how context shapes the mechanisms through which an intervention produces outcomes. As such, validity in realist evaluation does not rest on the achievement of a closed system. Instead, the validity of a realist evaluation rests on the application of realist logic to identify, test and refine theories which explain how the context in which interventions are implemented shape the mechanisms through which they work and thus their outcomes (although this idea of validity is still contested (Greenhalgh et al 2015)).

Planning for Emergence:
Flexibility and not Deviation

It is not always desirable or possible to make decisions about every aspect of study design in advance for a realist evaluation. Some aspects are expected to emerge as the theory develops, making their specification within a protocol challenging. Even where programme theories can be identified at the outset (which is not always the case), these theories are always provisional and subject to change. Like the programmes realist evaluation seeks to study, the methodology is emergent and iterative; new programme theories and 'sub theories' may surface during the evaluation and may require changes to data collection or analysis in order to test them (Manzano 2016). Therefore, protocols in realist evaluation cannot be written with the same degree of detail and precision as those for RCTs. Instead, flexibility needs to be written into realist evaluation protocols. For example, the number of interviews and specific participant groups may not be specified rigidly but described as a range (e.g. 20–35) with the option to reduce or extend if required ("these numbers may decrease or increase dependent of the process of theory testing").

Deviation from protocol as a
quality control process?

For RCTs, there is a clear consensus that deviation from the protocol on particular aspects of trial design and analysis, such as inclusion criteria and outcome reporting, constitutes bias (Schulz et al, 2010; McNamee 1997; Goldacre et al 2016). The flexibility built into a realist protocol poses challenges to this idea. Given that modes of data collection and analysis in realist evaluation are iterative and emergent, deviation from the protocol cannot be assumed to constitute bias. Instead, if deviations from the protocol are acknowledged and explained, then this enhances the transparency of reporting and readers can judge for themselves the extent to which this may compromise the study's validity.

> **"** …there are many different ways in which to conduct a robust and valid realist evaluation."

However, lurking underneath this logic is a more fundamental problem with the idea that comparisons between protocols and reported findings enable the detection of bias in realist evaluations. Realist evaluation is not 'a method' but an approach, variation in the practice of realist evaluation is evident (Pawson and Manzano 2012; Marchal et al 2012) and there are many different ways in which to conduct a robust and valid realist evaluation.

This makes it difficult to pinpoint the precise ways in which deviation from a realist evaluation protocol may constitute a threat to validity. The reporting and quality standards for realist evaluation produced by the RAMESES II study (www.ramesesproject.org) include a rubric against which the quality of realist evaluation study can be assessed.

The underlying assumptions of 'what constitutes validity' are to some extent different in realist as compared to other studies. In most approaches, validity refers to the ability of a study to accurately measure or assess what it sets out to assess. This remains a concern in realist evaluation. However, realist approaches acknowledge that some things (particularly, some mechanisms) may not be either observable or measurable; but nonetheless must be incorporated for a study to be considered a valid realist investigation.

Protocols for realist evaluations

Given the need for flexibility in conducting realist evaluations and accepting that there are many different ways in which to carry out a robust and rigorous realist evaluation, is there value in writing a protocol for a realist evaluation?

We argue that there can be, because:

- Writing a protocol requires rigour in planning, and realist evaluations (like all other methods) require good planning. This does not imply that the plans are rigid and cannot be changed, but that the plans provide the basis from which to adapt as required;

- Describing the methods to be used can identify 'non-realist' assumptions (for example, describing sampling methods that are not theory-based, or describing statistical methods that do not allow for disaggregation of outcomes by sub-groups or contexts);

- Realist evaluations do not expect to be replicable in the traditional sense of the word, but they do expect to be transparent. A protocol enables others to assess and critique the quality of the work, with the assumption that critique contributes to quality;

- A well-written protocol can provide the basis for other aspects of an evaluation project, including developing research ethics applications (where required) and writing the methods chapter of a report.

However, just as standards for realist evaluations are different from standards for other evaluation or research designs, the standards for realist protocols must also be different from those for other designs. The protocol, like the evaluation itself, should be consistent with the standards for realist evaluation (www.ramesesproject.org).

References

Dalkin S, Jones D, Lhussier M et al (2012) Understanding integrated care pathways in palliative care using realist evaluation: a mixed methods study protocol. BMJ Open 2:e001533. doi: 10.1136/bmjopen-2012–001533

Goldacre B, Drysdale H and Heneghan C (2016) The Centre for Evidence Based Medicine Outcome Monitoring Project (COMpare) Protocol: http://compare-trials.org/wp-content/uploads/2016/01/COMpare-Protocol-18.5.2016.pdf

Greenhalgh T, Wong G, Jagosh J, et al (2015) Protocol— the RAMESES II study: developing guidance and reporting standards for realist evaluation. BMJ Open 5:e008567. doi: 10.1136/bmjopen-2015–008567

Jones G and Abbasi K (2004) Trial protocols at the BMJ. BMJ 329: 1360.

Loscalzo, J. (2012). Irreproducible experimental results: causes, (mis) interpretations, and consequences. Circulation 125, 1211–1214. doi: 10.1161/CIRCULATIONAHA.112.098244

Manzano A (2016) The craft of interviewing in realist evaluation. Evaluation 22(3) 342–360.

Marchal B, van Belle S, van Olmen J et al (2012) Is realist evaluation keeping its promise? A review of published empirical studies in the field of health systems research. Evaluation 18(2) 192–212.

McNamee D (1997) Protocol reviews at the Lancet. Lancet 350: 6.

Mirzoev T, Etiaba E, Ebenso B et al (2016) Study protocol: realist evaluation of effectiveness and sustainability of a community health workers programme in improving maternal and child health in Nigeria. Implementation Science 11:83 DOI: 10.1186/s13012–016–0443–1

Mukumbang F, van Belle S, Marchal B et al (2016) Realist evaluation of the antiretroviral treatment adherence club programme in selected primary healthcare facilities in the metropolitan area of Western Cape Province, South Africa: a study protocol. BMJ Open 6:e009977. doi: 10.1136/bmjopen-2015–009977

Pawson R and Manzano-Santaella A (2012) A realist diagnostic workshop. Evaluation 18(2): 176–191.

Randell R, Greenhalgh J, Hindmarsh J, Dowding D, Jayne D, Pearman A, Gardner P, Croft J, Kotze A (2014) Integration of robotic surgery into routine practice and impacts on communication, collaboration, and decision making: a realist process evaluation protocol. Implementation Science 9:52 doi:10.1186/1748–5908–9–52

Schulz KF, Altman DG, Moher D and the CONSORT Group (2010). BMC Medicine 8:18 DOI: 10.1186/1741–7015–8–18

Shamseer L, Moher D, Clarke M et al (2015) Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: an elaboration and explanation. BMJ, 349 doi: doi.org/10.1136/bmj.g7647

Project team:

Professor Trish Greenhalgh
UNIVERSITY OF OXFORD
www.phc.ox.ac.uk/team/trish-greenhalgh

Professor Ray Pawson
UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/people/staff/pawson/

Dr Geoff Wong
UNIVERSITY OF OXFORD
www.phc.ox.ac.uk/team/geoffrey-wong

Dr Gill Westhorp
CHARLES DARWIN UNIVERSITY
www.cdu.edu.au/northern-institute/our-teams/603/5928#gillian-westhorp

Dr Joanne Greenhalgh
UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/people/staff/greenhalgh

Dr Ana Manzano
UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/people/staff/manzano/

Dr Justin Jagosh
UNIVERSITY OF LIVERPOOL
www.liv.ac.uk/cares

Developing Realist Programme Theories

The RAMESES II Project

Pawson has stated that programmes are 'theories incarnate' (Pawson, 2006) and that every programme has a theoretical underpinning, whether it is made explicit or not. One of the important activities in doing a realist evaluation is to 'surface' (i.e. bring to the surface, or develop) a theory or theories that explain what works, for whom, under what circumstances and how. Those theories are then 'tested' (confirmed, refuted or refined) using the best available evidence. A realist evaluation may examine a single programme theory, or a number of programme theories, depending on the goals of the evaluation.

A good brainstorming exercise to help in surfacing realist programme theories is to ask the question: 'What is it about programme X that makes it work?' This can be followed by other questions that help surface the other aspects of a realist programme theory: "In what circumstances will it work like that?" "Will it work like that for everyone?".

> " Programmes are 'theories incarnate'. Every programme has a theoretical underpinning, whether it is made explicit or not."

How is programme theory constructed?

There are many ways of going about identifying programme theory. These include reviewing documentation for the programme, interviewing people involved with the programme or experts in the content area, adopting theory from existing literature, and adding the evaluators' hunches ('informed guesswork').

Regardless of how the programme theories are developed, it is crucial that the theories reflect the realist understanding of causation. That means that the theory should not be limited to statements about whether or not the programme will lead to outcomes, but how and why it will do so, and for what kinds of settings and populations.

> " Regardless of how the programme theories are developed, it is crucial that the theories reflect realist understanding of causation. That means that the theory should not be limited to statements about whether or not the programme will lead to outcomes, but also why it will do so, and for what kinds of settings and populations."

There is often a significant difference between programme theory, which is specific to the particular programme, and formal theories in different disciplines (for example game theory in economics, constructivist learning theory in education, attachment theory in psychology and human development).

Formal theories drawn solely from the literature can end up being too abstract or distant from the intervention to be a perfect fit. More often than not it will be necessary to formulate customised theories that can then be backed up by formal theories in the literature.

What if the programme already comes with a theory?

If the programme has an established theory, it will need to be checked to see whether it can stand as a realist programme theory. In order for it to be a realist theory, it has to say something about contexts and mechanisms in relation to outcomes. Often, pre-existing programme theory comes in the form of a logic model, demonstrating inputs and outputs but not explaining how or why they work. In other cases, only a formal theory is provided (e.g. theory of diffusion of innovation, normalization process theory, or the behaviour change wheel). These are theories that

explain a general pathway, or a finite set of elements that previous research suggests are likely to be relevant, but they are not adapted to the particular programme, and they may not explain 'how and in what contexts'.

Realists need to make a claim about 'what is causing what to happen? For whom and how?'. Generative causation gets at the 'how and why' question. Realist programme theories need to be surfaced or developed if they are not already explicit.

How do realist programme theories incorporate political and economic elements?

Some interventions aim to change political and economic systems (think democratisation processes in developing countries, or changes to the powers of unions and employers in industrial relations systems). In these cases, the programme theory deals directly with political and economic issues.

In other cases, programme theories may be influenced by political or economic agendas. In other words, policies and programmes may be developed under the pressure of demonstrating that something is being done about an issue. For example, funds are allocated due to political pressures and not necessarily because there is proper evidence of need or evidence that the funded strategy is likely to be effective. Similarly, policy decisions can be the result of financial pressures – the need to cut costs and streamline budgets.

These agendas may inform theorising about the context, but programmes still have an intrinsic logic to them that is different from larger political or economic interests. The programme theory should reflect this intrinsic logic. For example, a policy to allow nurse-led prescribing of

medication may have been implemented as a low-risk cost cutting measure by reducing the number of visits required to the doctor. The programme theory should not be limited to statements such as: 'If you allow nurses prescribe, this will lead to economic benefit to the health system.' Instead it should ask 'what is it about nurse-led prescribing that works (or not?) and why?' This opens up a host of ideas and possibilities about the impact of nurse-led prescribing on the healthcare system, but also to impacts on nurse work experiences and work load, nurse-patient relations, nurse-doctor relations and so on. While it won't be possible to test all aspects of the theory in one evaluation, a wider examination of these elements of the programme will contribute to wider learning – and may in fact change the answer about the economic benefits to the health system by taking other issues into account.

What if the programme theory can't be established at the outset of the evaluation?

Pawson and Tilley (1997) emphasize that it is important to construct initial or candidate programme theories at the outset of the evaluation. This facilitates the development of data collection processes, realist interviews and context-mechanism-outcome (CMO) analyses. However, there are circumstances where this isn't feasible. Genuinely innovative programmes may be rare, but they do exist. There are also programmes which have run for a long time but without being theorised. In these cases, it is possible to undertake a 'theory building' evaluation. In this type or realist evaluation, the evaluation may start with a very general theory, not yet developed into a realist format. A realist theory is developed using the evidence collected during the evaluation and becomes part of the product of the evaluation.

References

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

Pawson R (2006) *Evidence based policy: A realist perspective*, London: Sage.

Project team:

Professor Trish Greenhalgh
UNIVERSITY OF OXFORD
www.phc.ox.ac.uk/team/
trish-greenhalgh

Professor Ray Pawson
UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/
people/staff/pawson/

Dr Geoff Wong
UNIVERSITY OF OXFORD
www.phc.ox.ac.uk/team/
geoffrey-wong

Dr Gill Westhorp
CHARLES DARWIN UNIVERSITY
www.cdu.edu.au/northern-institute/
our-teams/603/5928#gillian-westhorp

Dr Joanne Greenhalgh
UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/
people/staff/greenhalgh

Dr Ana Manzano
UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/
people/staff/manzano/

Dr Justin Jagosh
UNIVERSITY OF LIVERPOOL
www.liv.ac.uk/cares

Working with a Librarian on a Realist Review

Vicky Duncan, Tamara Waldron and Gary Groot, University of Saskatchewan

The RAMESES II Project

Literature searching for realist reviews differs significantly in some aspects from literature searching for other types of reviews, such as meta-analytic reviews (Cochrane reviews are the best known example of this type) or scoping reviews. If you are fortunate enough to work in an academic setting, and/or have access to the searching expertise of librarians, forming a collaborative librarian-researcher team will strengthen your project.

> " …forming a collaborative librarian-researcher team will strengthen your project."

How do you find a librarian?

Ideally, you want to find a librarian with expertise related to your research, recognizing that the search might need to expand into additional fields of study. It is useful if the librarian has some experience in working on teams in the role of "information specialist", but this isn't absolutely necessary. If you aren't sure which librarian at your institution has the most closely related expertise, browse through your library's website. Often the names of individual librarians will be posted beside a listing of each discipline's key resources. Once you've identified a librarian with relevant expertise, book an appointment to discuss your project. Don't be shy…librarians are very friendly!

When and How?

In order for the partnership to be successful, it's important to include the librarian as early in the project as possible. The ideal time would be after the team has identified the research question and at least one possible theory that might explain the mechanism(s) at work. Ensure that team members (including the librarian) have the same

> " …include the librarian as early in the project as possible."

understanding of the research question at the outset, and throughout the theory refinement process. It's common for members of the team to interpret the research question differently. Lack of clarity will make the librarian's job more challenging. Clarity about the question and the evolving theory will help the librarian recommend additional sources to search, including grey literature sources. Working closely together will help the librarian support your team.

Early discussion of authorship will also promote a collaborative team. More and more librarians at academic universities have faculty status, and co-authorship of a peer-reviewed paper will be attractive.

As the project progresses, it's important to ensure your team are using the same criteria to assess the quality of papers retrieved. If team members use different criteria, it could lead to mixed feedback to the librarian about the quality and relevance of the material and/or the search terms. While the librarian will not normally need to be present when the relevance of each document is assessed, he/she should be present at every meeting that discusses evolving program theory. A clear agenda that includes bringing the librarian up to date on the project, including any decisions likely to affect the refinement of search terms, will facilitate communication. The librarian can then adjust the search strategy accordingly.

As the team works through the project, it is imperative to document all decisions made in relation to program theory development. Decisions and reasoning for selection of databases or information resources, subject headings and keywords, dates covered by the search and other inclusion and exclusion criteria should be recorded at every meeting. This documentation will provide a journal of the team's thought processes and rationales as they work towards refining the emerging theory. This information will prove useful when it comes time to write up the research. It also facilitates clarity within the team to justify decisions made throughout the project.

What does the librarian need to know in order to best support you?

As well as your expected time frame for completing the project, it will be useful if you could provide your librarian with as many of the following as possible: key background documents on your topic; relevant articles that you are already aware of; authors that are known to be writing in the field; and what associations/organizations are likely to, or have published documents on your topic.

It is also essential to be candid about feedback on the databases searches performed by the librarian. If the majority of the articles missed the mark – tell the librarian – but also tell them why the information wasn't relevant. Work together to refine the search strategy until it reaches a good balance of specificity and comprehensiveness. The aim is to identify the most relevant information sources (realist reviews aim for theoretical saturation), without having to wade through an overwhelming number of irrelevant items.

> " Work together to refine the search strategy…"

Do you have a data management plan?

Data management is the latest buzz word in the world of research, and for good reason. Many funding agencies and ethics boards now require that research proposals include a data management plan. It will provide details on how you plan to organize, structure, store, and care for data used in or generated by a research project. The team should talk about issues such as: who will store the data for the project? Who can access the data? Would it be desirable to make the data publically available? Who can edit the data? How will the data be kept secure? How will you ensure that there are backups for the data? Some funding agencies require that the data be stored on at least three separate devices, such as the network, a laptop, and an external hard drive. How long does the data have to be kept for? It is best if the data management plan is in place at the beginning of the project. Making it up as you go is less than ideal.

> " Excellent communication can make the difference between a successful review, and an aborted one."

In short…

If we had to name one essential item for ensured success in working with a librarian, it would be fluid communication. Both librarian and researchers need to communicate frequently, meaningfully, and respectfully. Librarians have a lot to offer the realist review team. Excellent communication can make the difference between a successful review, and an aborted one.

Project team:

Professor Trish Greenhalgh
UNIVERSITY OF OXFORD
www.phc.ox.ac.uk/team/
trish-greenhalgh

Professor Ray Pawson
UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/
people/staff/pawson/

Dr Geoff Wong
UNIVERSITY OF OXFORD
www.phc.ox.ac.uk/team/
geoffrey-wong

Dr Gill Westhorp
CHARLES DARWIN UNIVERSITY
www.cdu.edu.au/northern-institute/
our-teams/603/5928#gillian-westhorp

Dr Joanne Greenhalgh
UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/
people/staff/greenhalgh

Dr Ana Manzano
UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/
people/staff/manzano/

Dr Justin Jagosh
UNIVERSITY OF LIVERPOOL
www.liv.ac.uk/cares

"Theory" in realist evaluation

The RAMESES II Project

Introductions to realist evaluation commonly start by noting that realist evaluation is one of the family of theory-based evaluations. But what exactly do we mean by 'theory' and how and why is it used in realist evaluation?

The term 'theory' means several different things in evaluation. Figure 1 presents four different categories of theory: philosophy, formal theory, evaluation theory and programme theory. Each of these has specific roles to play in evaluation practice.

Figure 1: Different types of theory that maybe relevant in realist evaluations
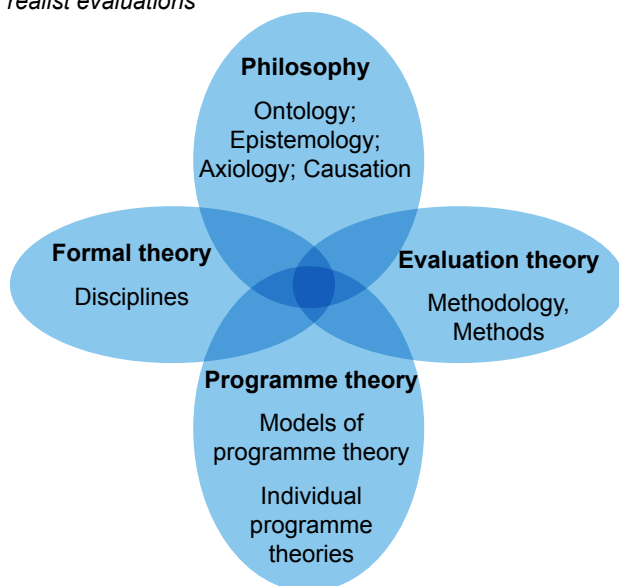
Philosophy

The first type of theory is philosophy, which deals with the 'fundamental natures' of things at such a level of abstraction that the ideas cannot be disproved. Different branches of philosophy consider different things: ontology deals with the nature of reality; epistemology with the nature of knowledge; axiology with the nature of value. Different philosophies also make different assumptions about how things are caused. These

four questions – what we consider to be 'real', what we can know about it, how we might go about valuing it, and how we think programs or policies cause change – are fundamental questions for evaluators.

For realist evaluation, the underlying philosophy is a realist philosophy of science. (See 'Philosophies and Evaluation Design' in this series.)

Evaluation theory

The second category is evaluation theory. These are theories about evaluation itself, relating to, for example:

- what evaluation is or isn't (e.g.: is it different from research? In what ways and why does that matter?)
- methodologies (the practical application of epistemological assumptions. See 'Philosophy and Evaluation Design' in this series for more detail);
- methods for data collection, analysis and interpretation. (Different schools of evaluation have different theories about what methods should be used, how and why they should be done and why each methods decision matters);
- the roles of evaluators (Independent? In house? Participant evaluators? Providing recommendations or just findings? Why?),
- the political stances that should or should not be taken (ranging from the politics of empowerment evaluation to that of randomised controlled trials).

Evaluation theories help us to "make good judgments about what kind of methods to use, under what circumstances, and toward what forms of evaluation influence" (Marks, 2005), learn from the past, and understand and contribute to debates in the sector.

Theoretical assumptions are apparent even at the level of specific methods. For example, many of the statistical methods commonly used in evaluation assume linear causation, even though there is

increasingly widespread acceptance that causation is not linear. Evaluators have to be able to justify and defend the evaluation theories and perspectives that they select and use for particular evaluations.

For realist evaluators, the overarching evaluation theory will be realist evaluation; and that will hold implications for every aspect of evaluation design and for each of the issues in the dot points above.

Formal theory

The third category of theory is 'formal theory' or 'substantive theory': the theories that operate in different domains or disciplines. Examples might include incentives theory in economics, attachment theory in human development or constructivist learning theory in education. Formal theories are often ignored in programme theory but probably should not be, for two reasons. The first is that programmes grounded in formal theories are sometimes more effective than those that are not. The second is that formal theories provide a bridge to a wealth of existing research and knowledge about a topic, which can inform several steps in many evaluation approaches: building programme theory, shaping questions for evaluations, interpreting findings and so on. In realist evaluation, formal theory is often used to identify mechanisms and features of context, and to explain how overall (often apparently disparate) sets of findings fit together. Theories written about a relatively specific phenomenon, such as Bandura's Social Learning Theory or Merton's Reference Group Theory, are usually more useful for this purpose than grand overarching theories such as Marx's theory of capitalism.

Programme theory

The final theoretical category is programme theory. This is what is usually referred to as being 'theory' in theory based evaluation. It is the description, in words or diagrams, of what is supposed to be done in a policy or programme (theory of action) and how and why that is expected to work (theory of change). Even here, traps await the unwary. Funnel and Rogers' book *Purposeful Program Theory* identified at least 15 different models of programme theory and 22 ways it is described (Funnell and Rogers, 2011, especially Chapter 9). The models have different purposes, contain different categories, imply different data needs, and justify different sorts of evaluative decisions.

> "Realist programme theory can be developed using relatively straightforward questions."

For realists, programme theory will explain how and why different outcomes are generated in different contexts.

Realist programme theory can be developed using relatively straightforward questions. The table below demonstrates.

Realist assumption	Focus Questions	Realist programme theory
1. Programmes intend to cause a change.	What change (outcome) does it intend to create?	Identifies *intended* outcomes.
2. The programme intends someone to do something different, or differently, to cause the change.	*Who* does it intend to do *what* differently? (May be several groups).	Identifies whose decision-making should be examined.
3. Programmes provide resources or opportunities, or change environments, to enable the different choice/ behaviour.	What does it provide to enable that choice or behaviour?	Identifies the 'resource' (in 'reasoning and resources'). Most programmes provide multiple resources each of which can trigger different reasoning.
4. Programme staff and participants make active choices and respond differently to resources.	How might different sub-groups of staff and participants respond to the resource?	Identifies various sets of 'reasoning' (in 'reasoning and resources'). Contributes to identifying 'for whom' programs may / may not work.
5. Participants' choices depend on/are affected by context.	What features of context affect how people respond to the resources? In what ways do those features affect responses?	Identifies what it is about context that matters. Contributes to identifying 'for whom' programmes may / may not work.
6. Context affects more than participant choices.	Which other features of context will affect whether & how the programme 'works'?	Identifies features of (e.g.) implementation, organisation, capacity, history that affect whether or not programmes work.
7. The choices that 'participants' make lead to different outcomes.	What outcomes would be generated by different decisions?	Identifies a wider range of potential outcomes (and thus indicators). Can include negative outcomes.
8. To evaluate programmes, collect information about context, mechanism and outcome.	What information will be needed and could be collected about contexts? Mechanisms? Outcomes?	Identifies the data necessary to test the programme theory.

Answering these questions provides both an overall programme theory and information about the data needed to test it. The overall programme theory comprises a number of sub-theories, any combination of which might be the focus of a specific evaluation. It is never possible to test all the sub-theories in a programme theory in one evaluation.

Finally, a point to note on the term 'middle-range theory'. These are theories that "... involves abstraction, of course, but they are close enough to observed data to be incorporated in propositions that permit empirical testing." (Merton, 1967). This term is often used in realist research and at times is confused as being programme theory or formal theory – or an additional theory required as well as both of these. However, the term "middle-range" is an adjective describing the level of abstraction of a theory. It is not another category or type of theory. Apart from philosophy, any of the types of theory mentioned above can (and probably should) be written as a middle-range theory. CMOs should also be written at a middle level of abstraction: specific enough to clearly explain the phenomenon, and general enough to apply across cases of the same type.

> "The term "middle-range" is an adjective describing the level of abstraction of a theory. It is not another category or type of theory. "

References

Funnell S and Rogers P (2011) *Purposeful program theory*. San Francisco: Jossey-Bass.

Marks M (2005) Evaluation Theory or What Are Evaluation Methods for?
www.hfrp.org/evaluation/the-evaluation-exchange/
issue-archive/evaluation-methodology/evaluation-
theory-or-what-are-evaluation-methods-for

Merton R (1967) *On Theoretical Sociology. Five Essays, Old and New*, New York: The Free Press.

Project team:

Professor Trish Greenhalgh
UNIVERSITY OF OXFORD
www.phc.ox.ac.uk/team/
trish-greenhalgh

Professor Ray Pawson
UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/
people/staff/pawson/

Dr Geoff Wong
UNIVERSITY OF OXFORD
www.phc.ox.ac.uk/team/
geoffrey-wong

Dr Gill Westhorp
CHARLES DARWIN UNIVERSITY
www.cdu.edu.au/northern-institute/
our-teams/603/5928#gillian-westhorp

Dr Joanne Greenhalgh
UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/
people/staff/greenhalgh

Dr Ana Manzano
UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/
people/staff/manzano/

Dr Justin Jagosh
UNIVERSITY OF LIVERPOOL
www.liv.ac.uk/cares

Realist Evaluation Interviewing – A 'Starter Set' of Questions

Gill Westhorp and Ana Manzano

The RAMESES II Project

The following table is designed to be read alongside the document 'The Realist Interview', in this series. It provides examples of basic questions that can be adapted for use in realist interviews, with an explanation or rationale for each.

Question / Question structure	Rationale
Can you tell me what your involvement in (or contact with) [programme / element of programme] has been?	Realist interviewing assumes that people know different things according to their roles. Use the answers to this question to tailor future questions to what it is that the respondent can be expected to know about.
What do you consider the outcomes of [programme / element of programme] to have been for [specific stakeholder group]? For programme participants, this question can be asked "for yourself", "for your family", and so on.	Realist evaluation assumes that programmes have different outcomes for different groups. The question should be asked repeatedly for different groups (e.g. children, parents, workers, the community as a whole) until the range of outcomes has been identified. If expected outcomes are not identified, it is OK to prompt for those outcomes. If unexpected outcomes are identified, prompt for greater description. Outcomes should be verified using other data types and sources where possible.
Can you give an example of [outcome named in previous question]?	Prompt for evidence of the nature and extent of the outcome. Also serves as a check on socially desirable responding.
Do you think that the outcomes have been the same for all [people within the specific stakeholder group – e.g. participants, workers]? In what ways have they been different?	Realist evaluation seeks to identify "in what respects, for whom" and "to what extent, for whom" programmes achieve outcomes. The question can be asked repeatedly for different groups – but to manage interview length, would usually focus on the stage of the implementation chain that is the focus of the evaluation
Have the outcomes been the same for [specific sub-groups identified in the programme theory – e.g. specific disadvantaged groups; different religions]. In what ways have they been different?	The focus of a realist interview is the programme theory. The question is seeking more specific information about "for whom" the programme has and has not been effective (in what respects, to what extent), but probing specifically in relation to sub-groups that are identified in the programme theory.

We are very curious about how [programme/ element of programme] causes its outcomes. How do you think the programme has caused, or helped to cause [outcome identified by respondent]?	Initial question leading into exploration of mechanisms. Many participants will identify programme activities (e.g. training) or resources (e.g. funding). It is ESSENTIAL to probe further – e.g. – So what did the training provide that was new? Was it mainly about new skills, do you think, or new attitudes? Or: So what exactly was the outcome of the training? How did that help cause (the later outcome)?
Do you think [the programme/element of programme] changed the way [stakeholder group] thinks or feels about [programme objective] in any way? In what ways? Can you provide examples? For participants: Has the programme changed the way you think or feel about [x] in any way?	Realist evaluation uses the construct of 'reasoning and resources' to explain how programmes cause outcome. This is an explicit probe for 'reasoning' in relation to a specific aspect of programme theory. E.G. – Programme may expect stakeholders to change their understanding of their roles or responsibilities and adapt their behaviour accordingly. So the question might be "Has programme affected how teachers think about their role in any way?"
There are lots of ideas about how [programme/ element of programme] actually works, and we think it probably works differently in different places or for different people. One of those ideas is [brief description of main mechanism]. Has it worked at all like that here/for you? Can you give an example?	The subject of a realist interview is the programme theory. The aim is to get the respondent to refine the programme theory for the particular context about which they know. This question should be asked about at least two different explanations of how the programme might work – as a check on socially desirable responding, and to elicit whether the programme works differently for different people.
What is it about the way [programme] was implemented that made a difference to how it worked? Or 'What is it about the way [Organisation] works that makes a difference to how it works?	Realist interviewing sees specific aspects of implementation as aspects of context (i.e. factors that affect whether and how mechanisms fire). The aim is to understand how implementation has affected mechanisms and therefore outcomes. Probe for both positive and negative aspects of implementation.
We've seen that this [programme] work differently in different places. What is it about this place that makes it work [so well, less well]?	Realist evaluation assumes context does affect outcomes (by affecting which mechanisms fire). Probe for aspects of culture, local resources/lack of them, local relationships, relationship between organisation and participants and so on.
If you could change something about this programme to make it work more effectively here, what would you change and why?	Often elicits understandings of why programme has not worked as effectively as it might (i.e. mechanisms not firing, aspects of context) as well as strategies for improvement.
What else do you think we need to know, to really understand how this programme has worked here?	Open probe that enables participants to comment on anything not covered by the interview. Structure of the question keeps the focus on 'how the programme works' and 'in this context'.

Project team:

Professor Trish Greenhalgh
UNIVERSITY OF OXFORD
www.phc.ox.ac.uk/team/
trish-greenhalgh

Professor Ray Pawson
UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/
people/staff/pawson/

Dr Geoff Wong
UNIVERSITY OF OXFORD
www.phc.ox.ac.uk/team/
geoffrey-wong

Dr Gill Westhorp
CHARLES DARWIN UNIVERSITY
www.cdu.edu.au/northern-institute/
our-teams/603/5928#gillian-westhorp

Dr Joanne Greenhalgh
UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/
people/staff/greenhalgh

Dr Ana Manzano
UNIVERSITY OF LEEDS
www.sociology.leeds.ac.uk/
people/staff/manzano/

Dr Justin Jagosh
UNIVERSITY OF LIVERPOOL
www.liv.ac.uk/cares