



HM Treasury

Magenta Book 2020

Supplementary Guide:

Handling Complexity in Policy
Evaluation

March 2020



HM Treasury

Magenta Book 2020

Supplementary Guide:

Handling Complexity in Policy Evaluation

March 2020



© Crown copyright 2020

This publication is licensed under the terms of the Open Government Licence v3.0 except where otherwise stated. To view this licence, visit nationalarchives.gov.uk/doc/open-government-licence/version/3.

Where we have identified any third party copyright information you will need to obtain permission from the copyright holders concerned.

This publication is available at: www.gov.uk/official-documents.

Any enquiries regarding this publication should be sent to us at public.enquiries@hmtreasury.gov.uk.

ISBN 978-1-913635-23-7 PU 2962

Authors

This report has been developed by CECAN – the Centre for Evaluating Complexity Across the Nexus¹:

Martha Bicket, University of Surrey

Ian Christie, University of Surrey

Professor Nigel Gilbert, University of Surrey

Dr Dione Hills, Tavistock Institute of Human Relations

Dr Alexandra Penn, University of Surrey

Dr Helen Wilkinson, Risk Solutions

Acknowledgements

The authors would like to thank ESRC for funding provided under grant number ES/N012550/1 and our advisory group and other members of the CECAN team who commented on earlier drafts of this annex: in particular:

Elizabeth Robin, Beverley Bishop; Siobhan Campbell, Steven Finch, Simon Maxwell, Wendy Vander Neut, Mike Daly, Candice Howarth, Ben Shaw and Ben Fagan Watson.

And also:

Dr Barbara Befani (University of Surrey), Professor Stephen Morris (Manchester Metropolitan University), Chris Rees (Risk Solutions) and Emma Uprichard (University of Warwick) for their useful comments on Appendix 1.

Any outstanding errors are the fault of the authors.

¹ <https://www.cecan.ac.uk/>

Contents

Authors	5
Acknowledgements	5
Preface.....	8
1. Why complexity matters.....	9
1.1 Complexity and complex adaptive systems	10
1.2 Implications for policy making	17
2. The challenges of complexity to evaluation.....	19
2.1 Context and boundaries matter	22
2.2 Problems of standardising and controlling	23
2.3 Dealing with change	24
2.3 The role of gatekeepers	24
2.4 Meeting the challenges	25
3. Commissioning and managing evaluations.....	26
3.1 Commissioning and managing a complex evaluation	27
3.2 Building understanding	30
3.3 Designing an evaluation	30
Evaluation purpose.....	31
System attributes.....	32
Feasible designs	33
3.4 Resourcing an evaluation	34
Budget and proportionality.....	34
Expertise.....	35
Access to data.....	36
Time and timing.....	36
3.5 Dissemination and embedding learning	37
3.6 Questions for commissioners.....	37
4. Selecting complexity-appropriate approaches.....	40
Terminology.....	41
4.1 Evaluation approaches	42
Different ways of framing evaluation	42
Methods and tools	43
Hybrid designs	43

Evolution over time.....	43
4.2 Selecting an approach.....	44
Purpose	45
System attributes.....	47
Feasible approaches.....	48
Appendix 1: Overview of available approaches.....	52
Appendix 2: Resources	63
General.....	63
Academic articles	64
Practitioner guides.....	65

Preface

Twenty-first century policymakers in the UK face a daunting array of challenges: an ageing society, the promises and threats for employment and wealth creation from artificial intelligence, obesity and public health, climate change and the need to sustain our natural environment, and many more. What these kinds of policy challenges have in common is *complexity*. Their implications spill over and transcend established boundaries between departments, policy domains, sectors and research disciplines. These problems pose risks and opportunities that demand integration of insights from multiple bases of evidence, many of which are incomplete, requiring reliance on modelling and scenarios with all their inherent uncertainties. The challenges in question also interact in highly complex ways with one another - for example, climate disruption poses many risks to other systems in our natural and built environments, and for human and animal health. Finally, these kinds of policy challenges have implications and effects over the long term - reaching far into the future of our communities and environments and calling for policymaking with a perspective over decades, if not half-centuries and more.

A case in point is the UK Government's 25-Year Plan for the Environment, *A Green Future: Our 25 Year Plan to Improve the Environment*, published in January 2018. This ambitious document sets out a broad strategy "for improving the environment, within a generation, and leaving it in a better state than we found it". The 25-Year Plan acknowledges the extreme complexity of the task it sets for integrated environmental policymaking. It sets out a range of goals and identifies many links with other policy domains and a very wide variety of stakeholders. The 25-Year Plan outlines a set of policy goals to be applied to highly complex systems that are co-evolving and whose stakeholders span many sectors, levels of activity and environments, and whose policy interventions are intended to have impacts extending over years and decades. It is a vision for policy whose ambition and implications point to the importance of an appropriate framework for appraisal and evaluation.

The example of the 25-Year Plan highlights with particular force the significance of *complex systems thinking* and its implications for policymaking. As such, it can be taken as an example of a major shift in policy challenges, analysis, design and evaluation that cuts across many areas of policymaking and governance. In understanding and working with these challenges, complex systems thinking and complexity-appropriate tools for policy can be invaluable and are of great significance. This Annex will explain what complexity thinking is, what the features of complex systems are, and how new methodologies and tools can equip policymakers to work *with* unavoidable complexity. In particular, the annex highlights ways in which complexity appropriate evaluation strategies can be used to gather timely and rigorous information about the implementation of an intervention, making a positive contribution to enhancing the process of adaptive policy making.

1. Why complexity matters

Key points:

- Complex systems have characteristics that make their behaviour hard to predict and which present challenges to policy making and evaluation:
 - Complex systems may be in a state of continual change and may also resist change.
 - Context and history matter, the same intervention will often have different outcomes in different contexts, or if delivered in a slightly different way.
 - Policy interventions in complex domains will often need to evolve over time in response to the way in which the system is adapting.
 - This highlights the importance of a continuous process of evaluation and learning, to enable flexible or adaptive management in complex, evolving environments.
 - Appreciation of how complexity can affect the policy process provides the opportunity to enhance effectiveness both in the design and delivery of the policy, and in its evaluation.
 - An appropriate evaluation strategy, in support of a learning or adaptive management approach, can help to track changes arising from a policy intervention over time, increase understanding of unexpected effects, and enable plans to be adapted if things take an unexpected course.
-

"...it is complex interventions that present the greatest challenge for evaluation and for the utilization of evaluation, because the path to success is so variable and it cannot be articulated in advance." – Patricia Rogers²

² Rogers, P, Westhorp, G, Walker, B. (2015). Dealing with complexity in a realist synthesis: community accountability and empowerment initiatives. In: Dealing with Complexity in Development Evaluation: A Practical Approach, Sage Publications, United States

1.1 Complexity and complex adaptive systems

The concept of complexity and the characteristics of complex adaptive systems have been studied in many different disciplines, from thermodynamics to economics, and have influenced research ranging from understanding weather patterns and ecosystems to studies of organisations and social behaviour.

There is no single agreed definition of complexity, and to choose one would not do justice to the diversity of work in this area. This annex therefore provides a list of widely agreed characteristics of complex systems, and then discusses their relevance to policy making and to policy appraisal and evaluation.

The defining feature of a complex system is that it is made up of (and emerges from):

- many diverse, interacting components,
- non-linear and non-proportional interactions between these components;

and in complex *adaptive* systems,

- adaptation or learning by the components in response to change.

Simple, complicated and complex

A distinction is sometimes made³ between *simple*, *complicated* and *complex* problems. Typical examples are following a recipe (simple), sending a rocket to the moon (complicated) and raising a child (complex). In reality, few policy interventions are truly simple. Usually, the level of complexity increases according to the number of different elements (actions, layers, organisations, government departments) that are involved.

For the sake of brevity, the term complex system is used throughout this annex, but most systems of interest to policy, including all social systems, are in fact *complex adaptive systems*. It is the unpredictability of complex adaptive systems that makes them particularly challenging, both in the implementation of effective policy interventions and their evaluation. For this reason, policy problems involving high complexity are sometimes referred to as 'wicked issues'.

³ See Glouberman, S and Zimmerman, B (2002) *Implicated and Complex Systems: What Would Successful Reform of Medicare Look Like?* DISCUSSION PAPER NO.8. [pdf] Commission for the future of health care in Canada, and

Complexity and wicked issues

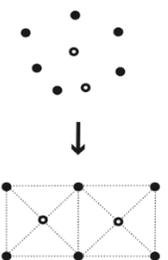
The 2007 Australian Public Service's *Tackling wicked problems; a public policy perspective*⁴ described several policy areas, such as climate change, obesity, indigenous disadvantage and land degradation, that had a level of complexity that made them particularly difficult to tackle. The term 'wicked' in this context is used, not in the sense of evil, but as a crossword puzzle addict or mathematician would use it: i.e. it is an issue highly resistant to resolution.

Although high complexity and the unpredictability of systems can be a major challenge for policy makers and evaluators, an understanding of some of the key properties of such systems can provide the opportunity to increase the effectiveness and impact of the policy.

Table 1 lists eleven properties of complex adaptive systems that have particular relevance to policy making and policy evaluation and provides definitions, examples and illustrations drawn from the natural, social and policy worlds.

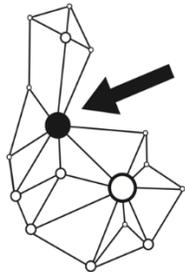
⁴ Australian Public Service (2007). Available at: [Tackling wicked problems, a public policy perspective](#) [Accessed 8th November 2019]

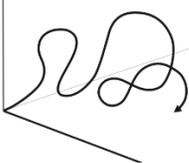
Table 1: Properties of complex systems with definitions and examples

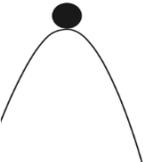
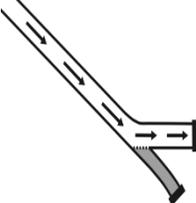
Property of Complex System ⁵	Definition	Examples
<p>Adaptation</p> 	<p>Components or actors within the system are capable of learning or evolving, changing how the system behaves in response to interventions as they are applied.</p>	<ul style="list-style-type: none"> • In the natural world: species evolve in response to change in their environment. For example, bacteria evolve resistance to antibiotics. • In the social world: people communicate, interpret and behave strategically to anticipate future situations. • In the policy world: the setting of policy 'targets' may result in efforts to individually or collectively 'game the system' (for example, NHS waiting time targets introduced in the 1990s alongside additional investment, resulted in significant reductions in NHS waiting times. However, in some places, the introduction of new screening processes and changes to reporting practices meant that targets could be met without much reduction in the overall waiting time for patients).
<p>Emergence and self-organisation</p> 	<p>New, unexpected, higher-level properties can arise from the interaction (and self-organisation) between the components (individuals, groups or organisations) within a system. These properties are said to be emergent if they cannot easily be predicted from the properties of the lower level components.</p>	<ul style="list-style-type: none"> • In the natural world: the resilience of an ecosystem to external change is an emergent property of the interactions between its species • In the social world: emergent properties can be seen in the formation of social movements, social norms and new markets, or even in the formation of a queue. • In the policy world: policies often aim to encourage emergence by, for example, imposing tariffs to help markets form; economic policy relies on emergence in the form of the "invisible hand"; regulation may also be needed to protect us from emergent phenomena.

⁵ Images by, or adapted from, Dr Joanna Boehnert created for CECAN: published under a Creative Commons license: Attribution-NonCommercial-ShareAlike 4.0 International

Property of Complex System ⁵	Definition	Examples
Unexpected indirect effects	Long causal chains within systems, generated by multiple interactions between components, can mean that intervention or change in one part of the system can lead to unexpected change in another, seemingly remote, component.	<ul style="list-style-type: none"> • In the natural world: reintroduction of wolves to Yellowstone led to revegetation, more beavers and changing river flow, as elk avoided grazing in valleys in response. • In the social world: decreasing popularity of smoking and more women working indirectly contributed to rising obesity levels. • In the policy world: the interaction between changing agricultural practice (increased winter planting), climate change (more extreme rainfall) and housing policy (building in floodplain) may have decreased resilience to flooding.
Feedback (and feedback loop)	Feedback occurs when the result or output of a process influences the input into the next iteration of the same process. This can happen either directly or indirectly, and can work to both increase and accelerate or to suppress the changes taking place.	<ul style="list-style-type: none"> • In the natural world: sweating or shivering helps the body maintain a constant temperature (example of negative feedback). • In the social world: a crowd may stampede if the panic of one individual spreads to others, creating panic and a rush to escape throughout the crowd (example of positive feedback). • In the policy world: increasing gentrification in a previously poor area is a phenomenon caused by positive feedback through the social transmission of popularity locations. People increased their cigarette consumption when nicotine levels were reduced, a negative feedback stabilising nicotine consumption.
Levers and hubs	Some components of a system may have a disproportionate influence over the whole because of the structure of their connections. Their activity may help to mobilise or slow down change, and their presence or absence make a system vulnerable to disruption.	<ul style="list-style-type: none"> • In the natural world: if a 'keystone' (highly influential, but low abundance) species in an ecosystem becomes extinct, there may be cascading extinctions amongst connected species and significant change in ecosystem structure and function. • In the social world: a well-connected and highly motivated individual or group may be mobilised to champion a particular cause. Alternatively, an individual or organisation may become a major obstacle to change through vetoing or blocking this (e.g. NRA in USA in relation to gun regulation). • In the policy world: statutory instruments, markets, regulations and protocols are examples of policy levers that can be used to produce significant social and environmental outcomes.



Property of Complex System ⁵	Definition	Examples
Non-linearity 	A system behaving in a non-linear fashion is one in which the effect of inputs on outcomes is not proportional: small changes may lead to large effects in one situation, but have little impact in another. It can also lead to sudden large-scale changes, or reverses in direction despite small or consistent changes in inputs.	<ul style="list-style-type: none"> • In the natural world: A species' population size does not increase without bounds as food sources increase, but will plateau as it is limited by other factors such as build-up of wastes or lack of space. • In the social world: a new product may be slow to take-off but after a certain point sales accelerate, before slowing again as the market is saturated. • In the policy world: in a weight loss programme, it was found that a six week programme had an impact, but a three week programme with similar content had little impact (i.e. its benefits were not half of those of the 6 week programme).
Domains of stability 	Systems may have more than one relatively stable state (called attractors in complexity science) and these may change as the context evolves. Complex systems will tend to gravitate towards these states, and then remain in them until some external change causes significant perturbation. If the system has multiple domains of stability, it can mean that once a change in the system has moved beyond a certain threshold (or tipping point) the system can slide rapidly into another state, a change that may be very difficult to reverse.	<ul style="list-style-type: none"> • In the natural world: the planet may exist stably with or without ice caps, but not at intermediate states; as the polar caps shrink, less sunlight is reflected and warming accelerates and vice versa as ice cover increases. • In the social world: a community of small businesses is temporarily cleared from an area for redevelopment, which is then delayed - it is extremely difficult for businesses to become sustainable again after a long hiatus, loss of social connections and customers. • In the policy world: the level of public transport provision in a suburban area may be stable either at a very low level, with few, marginalised, users or at a high level, at which the service makes money, is reliable and its use becomes habitual; providing support for an insufficient level of transport for an insufficient time may not generate the positive feedbacks needed stabilise high provision.

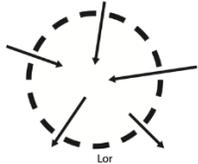
Property of Complex System ⁵	Definition	Examples
Tipping points	 <p>Closely linked to the idea of 'domains of stability', tipping points refer to the threshold beyond which a system goes through rapid change into a different state. It can be seen in situations in which change has initially been quite slow, but suddenly increases in pace.</p>	<ul style="list-style-type: none"> • In the natural world: A forest ecosystem may be stable over a large range of average rainfall, but may rapidly become desert as rainfall decreases beyond a certain threshold • Social world example: The gradual, then sudden, gentrification of a neighbourhood changing the demographics and character rapidly. Social media 'storms' in which minority opinions become the majority. • In the policy world: the sale of solar panels to householders increased very slowly over several years until suddenly taking off in response to a change in tariffs and word of mouth (across neighbourhoods).
Path dependency	 <p>The future development of a complex system depends on its history - how it got to its present state – as well as where it is currently. The order in which policy instruments or decisions are introduced may affect their cumulative impact.</p>	<ul style="list-style-type: none"> • Natural world example: Evolution is a highly path-dependent process. Organisms cannot radically change from their predecessors but change and modify themselves by mutations of adaptations that already exist. This is why evolution seldom finds optimal solutions. • Social world example: The health over the whole of the lifespan of an individual can be influenced by the diet and wellbeing of their parents and the conditions under which they were born and brought up (one of the causes of health inequality). • In the policy world: The choice of an organisation to lead a new policy initiative, and their history and reputation, may have a powerful influence over the way in which the policy is delivered, and how other organisations behave in relation to the policy.

Property of
Complex
System⁵

Definition

Examples

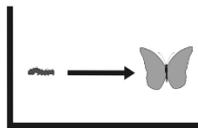
Openness



An open system is a system with many links and connections into its wider environment, which means that it can be powerfully affected by changes happening elsewhere. The links may take many forms including the exchange of information, inflow and outflow of material or energy, or of individuals and social groups and money.

- In the natural world: Invasive species such as grey squirrels or Himalayan Balsam arrive in ecosystems and out-compete similar native species. This may or may not have profound consequences for ecosystem function depending on the differences between the behaviour of the original and the new species.
- In the social world: A food production company may change rapidly in response to changes in food fashion or in the cost and availability of key ingredients.
- In the policy world: A 'delayed transfer of care' occurs when a patient is ready to leave a hospital but is still occupying a bed. While the NHS is responsible for the majority of delays, the social care system is responsible for a substantial proportion, as much as 38% at the beginning of 2017/18. Longer stays in hospital can affect a patient's health and impact waiting times in A&E departments and for planned surgery.

Change over
time



Complex systems develop and change their behaviour over time. This is due to their openness and the adaption of their components, but also the fact that these systems are usually out of equilibrium and are hence continuously in a process of change.

- In the natural world: Global ecosystems have formed, changed and developed over time, and continue to do so, from the origins of life, through the oxygenation of the atmosphere caused by the evolution of photosynthesis in bacteria, to the vast complexity of multicellular life in existence today.
- In the social world: Social norms, customs and cultures change radically over time and can never be said to have reached an end point. On a small scale, a local community partnership changes direction when one of the constituent partners changes its policies.
- In the policy world: New technological and social developments constantly drive policy change. For example, social media, the mass availability and use of individuals' data within a globalized economy have led to new behaviours and business models, with huge policy and legal implications.

1.2 Implications for policy making

Even a relatively straightforward policy intervention can have some elements of complexity. However, the complexity becomes greatly increased when one or more of the factors listed below are present:

Box 1: Factors driving complexity

- The problem being addressed has multiple causes;
- The environment in which the policy is being introduced is, itself, in a state of flux, or there are already several other initiatives taking place;
- The policy is being delivered at more than one level, involving a range of different interventions;
- There are large numbers of actors (organisations or individuals) who need to be engaged in delivery of the intervention, increasing the likelihood of conflicts of interest and the presence of different perspectives on the intervention, its outcomes and value
- The issue covers more than one policy domain.
- No one organisation, department or agency has overall control over the intervention and its likely outcomes

The last of these points is well illustrated by the 2007 'system map' that sets out many of the factors that can lead to obesity⁶. The factors operate in a wide range of policy making domains, including education, health, food regulation and business. While the problems are 'joined up' in complex ways, many of the policy actors in the system are not well coordinated and may not be in contact at all.

Example⁷:

Climate change is a pressing and highly complex policy issue involving multiple causal factors and high levels of disagreement about the nature of the problem and the best way to tackle it. The motivation and behaviour of individuals is a key part of the solution as is the involvement of all levels of government and a wide range of non-government organisations (NGOs). The debate has been characterised as falling into three competing 'stories' which emphasise different

⁶ Vandenbroeck, P., Goossens, J., and Clemens, M. (2007). *Tackling Obesities: Future Choices – Building the Obesity System Map*. [pdf]. Crown Copyright. Available at https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/295154/07-1179-obesity-building-system-map.pdf [Accessed 8 November 2019]

⁷ Assc.gov.au. (2018). *Tackling wicked problems: A public policy perspective* [online] Australian Public Services Commission. Available at <https://www.apsc.gov.au/tackling-wicked-problems-public-policy-perspective> [Accessed 8 November 2019]

aspects of the climate change issue: profligacy, lack of global planning and scaremongering.⁸ The three stories tell plausible but conflicting tales of climate change. None are completely wrong, yet at the same time none are completely right—each story focuses on some partial aspect of the debate. The stories’ proponents are unlikely to agree on the fundamental causes of and solutions to the global climate change issue. And since these stories contain normative beliefs they tend to be immune to enlightenment by scientific facts. This leaves the policy maker with a dynamic, plural and argumentative system of policy definition—typical of many wicked policy problems.

Because of the multiple interactions and influences in complex systems, the relationship between a policy intervention and its outcomes may be far from straightforward. For example:

- Complex Systems may be in **continual change**, or might **resist change**, as different parts respond at different times, or adapt to maintain the ‘status quo’ (linked to the characteristics of *adaptation, feedback loops, emergence and self-organisation and domains of stability*);
- **Context (and history) matters**: the same intervention will often have different outcomes in different contexts, or if delivered in a slightly different way (linked to the characteristics of being an *open system and path dependency*);
- The nature of the **change is unpredictable**, going faster or slower, or taking a different path to the one expected (linked to the characteristics of *non-linearity, unexpected indirect effects, feedback loops, levers and hubs and tipping points*);
- **Multiple perspectives**: different actors within the system will often have different views of what is happening, and this can influence the way they respond to an intervention.

It can be difficult to **describe and communicate** complexity if the prevailing model of change is essentially linear, i.e. assuming that if *this* element is put in place, *that* outcome will be the result. As yet, there is no wide understanding of, or language to describe, complex adaptive systems and their characteristics.

However, putting in place an appropriate evaluation strategy can help increase understanding if a complex policy, introduced in a complex setting, does not function quite as planned. Combined with a learning or adaptive management approach, this can help to track changes arising from a policy intervention over time and enable plans to be adapted if things take an unexpected course. For example, evaluation feedback can be used to update the design and strategy as understanding of the system develops, or the system itself changes, monitor aspects of the system that analysis indicates might be subject to change, or if unexpected or indirect outcomes become apparent.

The rest of this Annex describes how evaluation can help to address complexity. First, in Section 2, some of the challenges posed for evaluation are described. Section 3 discusses how these can be addressed, through the management of the evaluation; and Section 4 introduces evaluation approaches that can help with evaluating complex domains.

⁸ M. Thompson and M. Verweij. (2004). the Case for Clumsiness. *Singapore Management University, Humanities and Social Sciences* [online] . Working Paper Series, No. 5, pp. 12–23. Available at https://ink.library.smu.edu.sg/soss_research/25/ [Accessed 8 November 2019]

2. The challenges of complexity to evaluation

Key points

- Complex systems intensify the challenges for evaluation.
 - Complex systems can be particularly sensitive to context and to how the boundary and scope of the evaluation are defined.
 - Complexity can make causality hard to prove (i.e. whether the policy led to a particular outcome) because of the difficulties it creates for standardising an intervention or isolating a control group.
 - Because complex systems are constantly changing, the design of the evaluation may also need to be changed over its course.
 - Change may continue in difficult to predict ways after the evaluation has finished.
 - Some components of a complex system can have a disproportionate influence over the whole, their activity may help to mobilise or slow down change, and make a system vulnerable to disruption – evaluation can help identify these, but they can also significantly affect the evaluation through enabling or obstructing evaluation activities.
-

By its nature, complexity confronts the evaluator, and the commissioner of evaluation, with challenges. A summary of these, linked to the features of complexity, is given in Table 2

Table 2: Challenges for evaluation and its commissioning and management linked to characteristics of complexity

Complex system challenges	Linked to which characteristics of complexity	Challenges for evaluation	Implications for commissioning and managing an evaluation
Multiple interactions and influences which may be non-linear	This is the central definition of a complex system – all other features flow from this	<ul style="list-style-type: none"> • Long, highly interconnected, indirect causal chains linking inputs to impacts can make demonstrating attribution and causality challenging 	<ul style="list-style-type: none"> • Ensure appropriate evaluation approaches used • Involve expertise in specialist evaluation approaches to explore causality
Systems may be in continual change, or may resist change	<ul style="list-style-type: none"> • Adaptation • Emergence and self-organisation • Change over time • Feedback loops • Domains of stability 	<ul style="list-style-type: none"> • Objectives, design and data requirements of the evaluation may change over time • The intervention and outcomes may not be at a ‘final state’ or may revert to the previous state after the end of the evaluation 	<ul style="list-style-type: none"> • Consider reviewing the evaluation design as time goes by (adaptive management approaches) • Use specialist approaches that enable ‘findings’ about further change to be postulated, with clear caveats around uncertainty
Context (and history) matters Openness to outside influences	<ul style="list-style-type: none"> • Open systems, • Path dependency 	<ul style="list-style-type: none"> • Establishing a clear boundary around the intervention is difficult • It is not easy to standardise the intervention and hence establish clearly defined treatment and control groups • Outcomes in one setting may be different to those in another setting (external validity challenge) 	<ul style="list-style-type: none"> • Ensure appropriate stakeholder involvement including other departments/external stakeholders with view of wider context • Involve expertise in specialist evaluation approaches to explore causality and the influence of context • Ensure sufficient resources to enable collection of data on context and history
Multiple perspectives	<ul style="list-style-type: none"> • Multiple actors and relationships 	<ul style="list-style-type: none"> • No ‘one correct purpose’ for the evaluation • No ‘one correct understanding’ of the intervention and its setting • No ‘one correct answer’ to questions 	<ul style="list-style-type: none"> • Ensure alignment of understanding between all system stakeholders • Consider use of participative approaches to generate shared understanding

Complex system challenges	Linked to which characteristics of complexity	Challenges for evaluation	Implications for commissioning and managing an evaluation
The nature of the change is unpredictable Multiple causality	<ul style="list-style-type: none"> • Non-linearity • Unexpected indirect effects • Feedback loops • Levers and hubs • Tipping points • Domains of stability • Emergence 	<ul style="list-style-type: none"> • Specialist methods needed to tease out issues of causality and attribution • Evaluation plans may need to change to address the emergence of unexpected features 	<ul style="list-style-type: none"> • Involve experts with knowledge of a range of specialist evaluation approaches • Consider review of evaluation approaches as time goes by (agile management approaches) • Give careful consideration to system-wide metrics and responsive approach to emergent system characteristics and knowledge required to capture unpredicted features emerging
Complexity is difficult to communicate	<ul style="list-style-type: none"> • Features above are not widely understood 	<ul style="list-style-type: none"> • Particular challenge when communicating findings 	<ul style="list-style-type: none"> • May need additional time, and expertise in participative approaches, to explain/ensure alignment of understanding

The table highlights three key issues: (i) defining the scope and boundary of the intervention and the evaluation; (ii) establishing causality: did the intervention lead to observed effects; and (iii) dealing with change: managing the impact of questions, data needs and designs changing over the course of an evaluation.

2.1 Context and boundaries matter

Due to the characteristics outlined in Section 1, interventions into complex policy domains depend more on the context in which they are undertaken than simple environments.⁹

Example:

In healthcare, the effectiveness of a drug is influenced by factors such as the patient's other existing medical conditions, their lifestyle and consistency in taking the drug; these can all have implications for its overall success.

Example¹⁰:

An evaluation adopted a semi experimental design comparing one group of participants who received new (financial) resources and another that did not. The evaluation approach aligned well with an initial logic model but failed to show any significant difference in outcome between the two groups. Exploring why the evaluator realized that they had failed to identify that the intervention was just one part of a larger system (had placed the boundary in the wrong place) that there were important system level interconnections between all recipients whether in receipt of the new resources or not and that those not receiving new resources were receiving other kinds of support (other than the specific benefits being assessed).

This makes it more difficult to decide how and where to draw a boundary and set the scope of the evaluation. The way that the boundaries are drawn can make a big difference to the conclusions, since including or omitting factors with non-linear influences on outcomes may fundamentally alter the evaluation's findings and conclusions.

⁹ Datta, J and Petticrew, M (2013) Challenges to evaluating complex interventions: a content analysis of published papers. *BMC Public Health* [online] 13(568). Available at: <http://www.biomedcentral.com/1471-2458/13/568> [Accessed 8th November 2019]

¹⁰ Gauthier, B. (2019). Chapter 19. In: H. Hutchinson .ed. *Evaluation Failures: 22 Tales of Mistakes Made and Lessons Learned*. Sage Publications.

Example:

An evaluation of a new training initiative discovered that the location of the training, the background and experience of the trainers and how training participants were recruited appeared to have had a major influence on the outcome: potentially more important than the actual training content. Because this was not anticipated at the start of the evaluation, and the evaluation approach was not reviewed during its delivery, insufficient data was collected to test the influence of these factors rigorously.

2.2 Problems of standardising and controlling

Counterfactual approaches, such as experimental methods and randomised controlled trials (RCTs) are widely viewed as providing the most robust evidence, but rely on being able to control extraneous variables around the intervention, while manipulating others in order to establish causality. This helps to maximise internal validity – the confidence that can be placed in the cause and effect relationship between variables in that specific context. However, in a complex intervention, the factors ruled out may be highly relevant either to the overall effectiveness of the intervention, or help to explain variations in outcome which arise in different settings. ‘Controlling out’ many aspects of the intervention can make it difficult to apply evaluation findings to a different context, i.e. address the external validity of results.

Another difficulty is standardising a complex intervention, since it may be constantly changing, and highly dependent on other factors in its wider environment. In some cases it is inappropriate or counterproductive to try to standardise the intervention, or to isolate a control group, making it hard to meet the requirements of an experimental design.

Example:

An experimental design used to evaluate a programme designed to help disabled people into work foundered when it was discovered that organisations operating in different locations delivered the intervention in different ways, tailoring this to the types and levels of disability they were dealing with, and the kind of work opportunities available locally. They also differed in how they targeted and recruited participants (i.e. in terms of their level of disability, and ‘readiness to work’). As well as making it difficult to clearly articulate, or standardise, the intervention, these factors also made it difficult to identify and collect uniform data from an intervention group and a control sample.

This is not to say that experimental designs cannot be useful in complex environments, if one element can be isolated and is relatively standardised, and the wider context stable. However, they may need to be combined with other approaches (see Section 4). The challenge of establishing a counterfactual - that is finding a reliable way of determining whether an intervention has made a real difference, or whether the change, or some part of the change, would have happened anyway - is central to many complex evaluations. Section 4 describes a range of different approaches to achieving this.

2.3 Dealing with change

Because complex evaluations and the systems they seek to evaluate are likely to change over time, new questions, causal pathways, stakeholders or even objectives may emerge during the evaluation which the original evaluation design does not accommodate. The validity of the evaluation may also be quite limited. For example, an intervention might actively change in response to an ongoing evaluation, and change may continue, or even accelerate in the 'intervention site' after the evaluation comes to an end, or regress to an earlier state once funding and support has been withdrawn.

Example¹¹:

In a four year public health programme promoting community responses to cardiovascular disease delivered across a number of local districts. The evaluation design included both process and impact elements and a strong participative element. After working well for the first year, a number of disruptive policy and organizational changes took place in the wider context forcing sites to change their plans. Staff turnover was rampant. Although regularly adapting the evaluation to respond to these changes, in year three the whole evaluation approach, and some of the key findings, were challenged in a letter of complaint by one of the new site managers who was unfamiliar with participative research approaches. The evaluation contract was abruptly terminated.

2.3 The role of gatekeepers

One part of the overall system or programme, for example key individuals can obstruct or undermine evaluation activities because they affect many other parts of the system (high connectivity), or due to their role as a "gatekeeper" to large parts of the system. Such individuals can make it difficult for evaluators to collect data – or by rejecting the veracity of the data, lead to the findings of the evaluation also being rejected. This may be particularly the case where data being collected by an evaluation or results emerging from it, are seen to threaten the current 'status quo' or equilibrium of the system. This focus on data usually points to much deeper problems in the system. In the second example below, addressing the deeper, underlying issue helped this new information to be accepted. To be effective, evaluations must be sensitive to points of energy and influence in the system, as well as ways in which momentum and power flow within the system.

¹¹Barrington, G. (2019) In Chapter 3, In: H. Hutchinson .ed. *Evaluation Failures: 22 Tales of Mistakes Made and Lessons Learned*. Sage Publications

*Example*¹²:

An evaluator ran into difficulties when presenting baseline findings from the evaluation of a child labour project run by an international NGO. Government officials challenged the methodology and failure of the evaluators to consult adequately on this. However, a key problem for the government department was that the data indicated levels of child labour that were potentially damaging to the reputation of the companies involved and the country itself.

*Example*¹³:

Programme managers of a multisite programme evaluation vigorously rejected extensive evidence (drawn from interviews, a survey and case studies) that local staff wanted to have greater contact and communication with the central team. This masked the concern from the central team that they didn't have the capacity to provide this level of communication with sites. When presented with more contextual information and suggestions of how this challenge could be addressed without a major investment of resources, the evidence was (reluctantly) accepted.

2.4 Meeting the challenges

Although complex policy environments can be challenging to evaluate, there is a wealth of evaluation good practice and complexity-appropriate evaluation approaches that can help, supporting its effective delivery, optimising its impact, and providing transferable lessons for similar policies. These are discussed in the following sections.

¹² Muramutsa, F. (2019). Chapter 8, In: H. Hutchinson .ed. *Evaluation Failures: 22 Tales of Mistakes Made and Lessons Learned*. Sage Publications

¹³ Tindall. D. (2019). Chapter 20, In: H. Hutchinson .ed. *Evaluation Failures: 22 Tales of Mistakes Made and Lessons Learned*. Sage Publications

3. Commissioning and managing evaluations

Key points:

- A good evaluation can help in understanding, and managing, an intervention by providing regular and rigorous feedback, and opportunities for ongoing learning and reflection.
 - In complex policy environments commissioners, evaluators and the policy clients should expect to challenge traditional notions of evaluation and evaluation design:
 - Evaluative activities should become integral to intervention implementation, building on modelling and analysis carried out as part of policy design
 - Evaluations should explore not just how well the intervention is working and how it can be improved, but also question whether quite different approaches may have produced better results
 - In this way, appraisal and evaluation will merge into a continuous process of learning and policy evolution, recognising that complex systems need to be continuously monitored and adapted (as exemplified in adaptive management).
 - The inclusion of key stakeholders in planning and 'mapping' of the intervention helps to increase our understanding of complexity and any challenges this might pose.
 - Stakeholders may have different views on complexity and appropriate evaluation strategies, so expectations and assumptions will need to be managed carefully.
 - Commissioners and other key stakeholders need to be aware that the level of quantitative rigour and certainty of outcome may be limited, even when using sophisticated evaluation methods; they need to be realistic about what can be achieved
 - Governance and management of evaluations need to be flexible to respond to emergent changes to the intervention, or to system responses to the intervention, or as new understanding evolves.
-

This section indicates ways that a well-planned and managed evaluation can help with understanding the challenges posed by complexity and provide opportunities to anticipate and deal with these. The points below build on the guidance on planning and management given in the main Magenta Book, highlighting areas that are particularly important when dealing with a complex policy intervention.

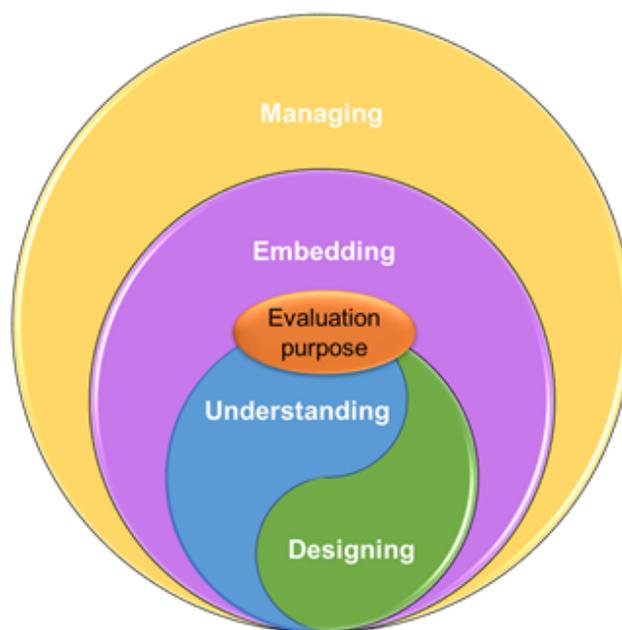
A list of questions is provided at the end of this section that commissioners can use to aid planning at each stage of the evaluation planning process.

3.1 Commissioning and managing a complex evaluation

The Magenta Book identifies a number of key stages in the commissioning and management of an evaluation: scoping, design, choosing appropriate methods, conducting the evaluation, and disseminating and using the learning. When the context and the policy are complex, the stages are likely to be less clear-cut, with a central task throughout planning and delivering both the intervention itself and the evaluation, being to gain insight into the system itself, and respond to new learning and developments as these emerge.

This close, and ongoing, inter-relation between developing understanding and the evaluation design are illustrated in Figure 1 taken from the Complexity Evaluation Framework commissioned by Defra. The figure also shows how these activities are run alongside the ongoing processes of embedding learning, and managing (including the commissioning process). This shows the evaluation process as a series of nested activities, carried out iteratively and in parallel, rather than sequential steps.

The Defra Complexity Evaluation Framework GUIDANCE FOR COMMISSIONING EVALUATIONS IN DEFRA



In this framework, evaluation is described as a set of nested processes.

The evaluation is centred around and defined by the **EVALUATION PURPOSE**. This purpose informs an iterative process of **UNDERSTANDING** the system and intervention and **DESIGNING** and adapting the evaluation, which will continue throughout the evaluation.

These activities are conducted with the ongoing engagement of stakeholders, and understanding and learning are fed back and **EMBEDDED** into relevant processes both inside and outside of the evaluation.

All of these interacting components of an evaluation are actively **MANAGED** in response to evaluation findings and evolving needs.

Figure 1: Nested components of a complexity-appropriate evaluation

A high level of complexity – and the adoption of a more integrated approach to evaluation – also highlights the importance of having a commissioning and management style that supports rather than restricts the ‘emergent’ elements of an adaptive system. This is a growing theme in management literature, particularly where high levels of uncertainty and a lack of consensus between stakeholders around appropriate interventions and expected outcomes makes the management task itself increasingly complex.

The Stacey diagram is one way of capturing these dimensions. Figure 2 shows how increasing uncertainty and lack of agreement increase the complexity of the situation and shows the types of management activities that can be adopted to tackle these.

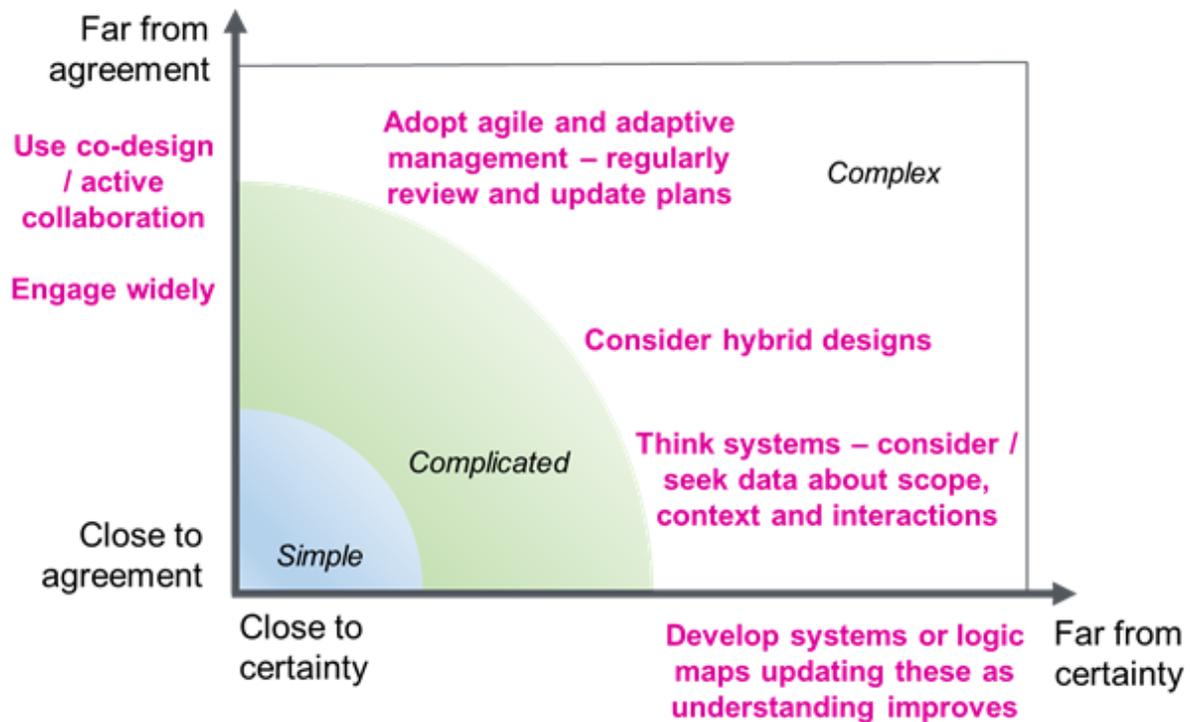


Figure 2: Planning and managing in complex situations¹⁴

Lack of consensus, uncertainty and greater complexity require more attention to be paid to supporting understanding and learning, building consensus between different parts of the system, and having regular points of review, when plans can be changed to respond to changes taking place, either on the ground, or in the wider context. These are all features of a more 'agile' or 'adaptive' management style (see box). Evaluation activities alongside this can provide useful data to support the process of learning and adaptation of plans.

¹⁴ This diagram has been adapted from the original matrix developed by Stacey – see Stacey, R.D. (1996). *Strategic Management & Organisational Dynamics*. London: Pitman. Pp 47

Box 2: Adaptive and agile management

'What is distinctive about **adaptive management*** is not that interventions may be changed in response to monitoring and learning. It is that those interventions are based on an understanding that engaging with a complex problem or system will require an exploratory, flexible approach because solutions are not known in advance.'

'Organisational leadership and culture need to facilitate and promote adaptation. This means accepting uncertainty at times and accepting that plans and budgets will need to change. It means putting in place HR, finance, communications and other systems that facilitate learning, reflection and course-correction to ensure that the best results possible are achieved, even if they are not the results that were expected.'

*Quoted from Bond (2016) Adaptive management: what it means for CSOs¹⁵

'**Agile project management*** is an approach based on delivering requirements iteratively and incrementally throughout the project life cycle. At the core of agile is the requirement to exhibit central values and behaviours of trust, flexibility, empowerment and collaboration.'

*From the Association of Project Managers¹⁶

Even if a formal adaptive or agile management approach is *not* used, the management of a complex evaluation *will* require regular opportunities for review, ensuring that the methods being used remain suitable, and new learning and insights are being incorporated into decision making. This is likely to require additional meetings, presentations and interim reports and close working with the evaluators. This needs to be taken into account in drawing up the plan and the budget for the evaluation.

This can present challenges, particularly in the commissioning stage. Government commissioning processes usually assume that bidders will be able to draw up and cost a detailed plan of action at the outset, which will remain relatively unchanged during delivery. The risk of this leading to an inappropriate evaluation will be compounded if those involved in the commissioning process have a limited knowledge of complexity, or complexity-appropriate evaluation designs, and opt for 'tried and tested' evaluation approaches in order to minimise risk. The risk is further increased if the true level of complexity is not communicated to bidders, and these bidders are given limited opportunities to discuss this and possible approaches to deal with the complexity with the commissioners. Adopting an adaptive or agile management style implies having a more 'open' specification, building in an initial exploratory or scoping phase after which plans, at least for the initial phases of the evaluation, can be firmed up, and building formal review points into any

¹⁵ Bond (2016). *Adaptive management: what it means for CSOs*. [pdf]. Available at: https://www.bond.org.uk/sites/default/files/resource-documents/adaptive_management_-_what_it_means_for_csos_0.pdf [Accessed 8th November 2019]

¹⁶ Apm.org.uk. (undated). *What is agile project management?* [online]. Available at: <https://www.apm.org.uk/resources/find-a-resource/agile-project-management/> [Accessed 8th November 2019]

contracts drawn up. Having a detailed risk register and a 'contingency' element in the budget will also help address any unforeseen developments without seriously derailing evaluation plans.

3.2 Building understanding

One of the first steps in the commissioning and management of an evaluation is to gain an initial **understanding** of the policy or intervention to be evaluated, and the context in which it is taking place. This forms the basis for the initial design of the evaluation and for the more detailed understanding delivered through the evaluation. The Magenta Book recommends developing a 'map' of the intervention and its outcomes through use of *logic or policy mapping*, or developing a theory of change, particularly if this was not done at the policy planning or appraisal stage. Where the policy or its setting is complex, this mapping stage will be particularly important, with some additional points that need to be borne in mind.

Some mapping tools assume a relatively linear progression from input to outcomes. However, change in complex systems is often far from linear, and can be unpredictable. The presence, and impact, of complexity characteristics, such as feedback loops, tipping points, levers and hubs (see Table 1) often only become apparent over time.

The mapping process can therefore be greatly enhanced by:

- Involving stakeholders with a detailed understanding of the context in which the policy is being delivered, or with experience of similar actions undertaken elsewhere, helping to anticipate how key features of complexity might affect delivery and outcome.
- Treating any maps drawn up at an early stage as a 'theory to be tested', subject to revision as more information becomes available.
- The use of system mapping and modelling approaches (introduced in Section 4), either before or alongside the logic mapping, to encourage an appreciation of the wide range of factors that can impact on delivery and outcomes and how these interact.

If system mapping has already been undertaken during the appraisal stage, this can provide a good foundation for the design of the evaluation. Indeed, in a complex intervention, the transition between appraisal and evaluation stages might be much less clear cut, with changes taking place over the life of the policy requiring regular review and even reappraisal of any new options that emerge, and the evaluation actively feeding into this process. This is discussed further under *evaluation purpose* below.

3.3 Designing an evaluation

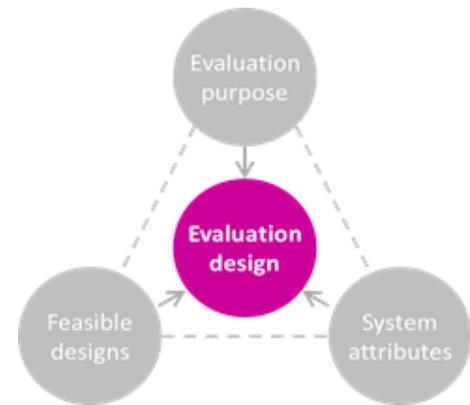
Section 4 deals in depth with the issue of choosing an appropriate evaluation approach given the **purpose** of the evaluation, the **attributes of the system** being evaluated, and the **feasibility** of possible designs, illustrated in the design triangle below. This section addresses management and commissioning of the evaluation design from the same three perspectives. It focuses specifically on the challenges of working in policy contexts that are unfamiliar with complexity or complexity-appropriate evaluation approaches and dealing with multiple stakeholders.

The design triangle

The 'design triangle' illustrates three inter-related factors that should be considered in establishing or reviewing an evaluation design. The diagram emphasises that many of these decisions are interconnected. For example, the kinds of evaluation questions that can be asked partly determines the selection of approach and methods, which also has to take account of system attributes in understanding the kinds of questions that can be answered. It is an adaption of the design triangle developed by Stern*.

* Where Stern uses programme attributes, we have used system attributes to emphasise that complexity aspects of the wider system, within which the policy is implemented, will also affect the choice of evaluation design.

See for example: https://www.bond.org.uk/data/files/Impact_Evaluation_Guide_0515.pdf,



Evaluation purpose

Evaluations are undertaken for a range of purposes. The Magenta book mentions two key purposes in Chapter 2 (see Table 2.1): **accountability** and **learning**, and also mentions the value of supporting evaluative thinking (see Box 3 below). There is a further discussion about the different purposes of evaluation in Section 4, which includes the fact that evaluation can also support **listening and building**: ensuring diverse voices are heard and building trust and legitimacy across stakeholders.

Box 3: Evaluative thinking¹⁷

Evaluative thinking is systematic results-oriented thinking about:

- What results are expected
- How results can be achieved
- What evidence is needed to inform future actions and judgments, and
- How results can be improved in the future.

'Evaluative thinking becomes most meaningful when it is embedded in an organization's culture. This means that people in the organization expect to engage with each other in clarifying key concepts, differentiating means and ends, thinking in terms of outcomes, examining the quality of evidence available about effectiveness, and supporting their opinions and judgments with evidence. Evaluative thinking is what characterizes learning organizations ... Keeping up with research and evaluation findings becomes part of everyone's job'

¹⁷ Quinn Patton, M. (2014). *Evaluation flash cards: Embedding Evaluative Thinking in Organizational Culture*. [pdf]Utilization-Focused Evaluation. St. Paul, Minnesota. Otto Bremer Foundation. Available at <https://iprc.iu.edu/spf/docs/Evaluation%20Flash%20Cards.pdf> [Accessed 8th November 2019]

The purpose will be reflected in the evaluation questions that key stakeholders want to see addressed. In a complex policy situation, there may be a gap between what key stakeholders would like to know, and what it is possible to conclude from the evaluation.

Where the policy is cross-sectoral and/or involves several different stakeholder groups, differences can also emerge between these about the purpose of the evaluation, and how important it is to have:

- A learning element to the evaluation (as well as demonstrating accountability and evidence of impact) - learning aims are particularly important in the evaluation of a complex intervention, supporting an adaptive or agile management approach to policy implementation, and enabling the evaluation design to respond to emerging findings (for more on agile and adaptive management see under managing the evaluation above)
- A process element in the evaluation, as well as assessing outcomes and impacts. The Magenta Book distinguishes between process, impact and value for money evaluation. When a policy or its setting are complex, a process element to the evaluation will be important for tracking changes to the implementation process over time, helping to explain why impacts have, or have not been achieved.

Commissioners should be prepared to explore not just how well the intervention is working and how it can be improved, but also question whether quite different approaches may have produced better results. In this way, appraisal and evaluation will merge into a continuous process of learning and policy evolution.

Some evaluation approaches are particularly useful in supporting learning and adaptation and listening and building. **Participative methods** such as participative system mapping and modelling approaches, and **developmental evaluation** provide regular opportunities for stakeholders to engage in collecting, reviewing and responding to emerging findings. **Developmental evaluation** was specifically designed for complex, innovative and emergent interventions, through the adoption of regular cycles of evaluation and decision making. However, as in other areas, stakeholders may vary in their willingness to engage in learning and reflection and accept high levels of risk and uncertainty.

System attributes

As has already been noted, a key stage in planning an evaluation is to understand the intervention itself, and the setting into which it is being delivered, ideally through some kind of system, logic or theory mapping. Knowing about any characteristics of a complex adaptive system present will be an important element in choosing an evaluation approach. Box 1 in Section 1 identifies factors that increase the complexity of a system. Table 1 illustrates the characteristics of complexity.

Not all of the characteristics will be apparent at the outset of an evaluation. Different elements of complexity (feedback loops, sensitivity to context, levers and hubs) will often emerge over the course of policy delivery, or as the evaluators find out more about the setting in which it is being delivered. This may be unwelcome news if those who planned the policy were assuming a generally 'linear' model of policy implementation and can throw an inappropriate evaluation strategy off course.

The involvement of stakeholders at the start of the evaluation planning process who have experience or 'local' knowledge about the kind of intervention being used, and its setting, will be particularly helpful. In policy arenas where understanding about complexity is still at an early stage, it can be useful to engage experts with understanding of complexity to advise on planning the evaluation.

Feasible designs

Many of the familiar evaluation methods and approaches described in the Magenta Book can generate useful findings as long as the complexity of the policy and its setting, is acknowledged, and the findings treated appropriately.

However, there is now a growing body of additional evaluation approaches that are particularly useful in addressing complexity, although knowledge of these remains patchy. Adoption of appropriate methods can be particularly challenging if the prevailing culture favours evaluation approaches that do not address the challenges of complexity very well.

The following issues may be particularly challenging for some stakeholders:

- **Evaluator objectivity:** some of the methods described in the next section are more explicit in their acknowledgement that evaluators become part of the situation as they interact with people on the ground.
- **Participatory evaluation** approaches can be particularly helpful in enabling local stakeholders to play a role in the design and 'sense making' aspects of the evaluation, but may be unfamiliar to some stakeholders.
- **The relative value of qualitative and quantitative data:** several of the approaches described in Section 4 rely on qualitative as well as quantitative data; this will be challenging in policy environments, or with stakeholders, that view only quantitative findings as robust.
- **The level of precision and detailed explanation that can be achieved:** The level of quantitative rigour and certainty of outcome may be limited, even when using sophisticated evaluation methods.

*Example:*¹⁸

Three evaluations taking place in a high stakes public policy environment illustrate the tension for evaluators between meeting stakeholder expectations at the same time as maintaining their professional integrity. In one case 'difficult' findings were challenged because they were based on qualitative rather than quantitative data, the situation becoming more high profile when 'challenging' findings received media coverage. 'To them (the committee receiving the report) numbers constituted 'real' data and qualitative information was nothing more than anecdotes and opinions. At times our conversations felt like we were speaking two different languages. We had different views of data evidence and evaluation methodologies'.

Strong leadership from project managers, a *clear explanation* about why a particular approach is being adopted, and the *opportunity to air differences of perspective* will be important. Otherwise, there is a risk inappropriate evaluation approaches will be selected, or of evaluation findings being dismissed as irrelevant or invalid by stakeholders who might otherwise have much to learn from both the evaluation process and its findings.

As the approaches and methods used may need to change as understanding of the system and intervention develops, and as new players may become involved over time, this process of communication needs to continue throughout the evaluation.

Resources will be an important consideration in deciding how feasible an approach is - this is discussed in the next section.

3.4 Resourcing an evaluation

As noted earlier, many of the evaluation methods and approaches listed in the Magenta Book may work well in a complex evaluation, so long as the challenges arising from this complexity is acknowledged. However, even in a relatively straightforward evaluation design, issues related to expertise, access to data, timing and the amount of time may require additional thought, as well as the overall budget.

Budget and proportionality

Some of the more sophisticated evaluation approaches being developed to address complexity, described in Section 4 are quite expensive to implement. Costs can be higher than other evaluation approaches because:

- The methods themselves can be time consuming
- They require costly expertise or software
- The data required may be difficult to access

¹⁸ Mohan, R. (2019). Chapter 7, In: H. Hutchinson .ed. *Evaluation Failures: 22 Tales of Mistakes Made and Lessons Learned*. Sage Publications

- Ongoing dialogue (meetings, interim reports) between evaluators, research managers, policy or implementation teams is required to support a flexible or adaptive management approach.

*Example:*¹⁹

An evaluation of a professional development programme for government employees had a large number of stakeholders involved. Soon after the start there were a number of changes in key personnel followed by a change in government with accompanying changes in policy and restructuring. Significant evaluation resources had to be used in briefing incoming staff and adapting the design to accommodate shifts in focus and interest leaving little left in the budget for the final analysis and report writing.

Key considerations in setting budgets for a complex situation are understanding how the evaluation findings are to be used, and the level of risk associated with using the wrong kind of evaluation approach, leading to misleading findings. A proportionate evaluation delivers findings that are fit for purpose given the risks.

Expertise

Understanding of complex adaptive systems, and their implications for evaluation, remains quite limited. Added to this, some complexity-appropriate methods described in Appendix 1 are relatively new, and may require specialist knowledge and experience. This may mean looking beyond the usual evaluation providers and tapping a wider field of expertise, including practitioners and academics with expertise in complexity sciences, new research methods or modelling skills. Involving experts with experience of this kind during the early planning (prior to commissioning the evaluation) can help ensure that the commissioning brief is sufficiently open to allow the use of appropriate methods.

The expertise needed to undertake a complex evaluation is also likely to go beyond having appropriate technical knowledge and experience. Understanding the theory underpinning different methods (e.g. different approaches to understanding attribution) may be required to make the case for alternative evaluation strategies, and sophisticated analytic skills required to make sense of complex data. Good interpersonal and management skills are also required. Negotiating with multiple stakeholders and keeping the evaluation on track in complex settings is particularly challenging, and some of the more interactive methods (e.g. running a theory of change, system mapping or dependency modelling workshop with stakeholders) will also require good (and specialist) facilitation skills.

¹⁹ Hutchinson, H (2019). *Evaluation Failures: 22 Tales of Mistakes Made and Lessons Learned*. Sage Publications

Access to data

A complex intervention often requires a wide data 'net' to ensure that sufficient information is collected about the multiplicity of factors influencing outcomes (and different stakeholder views of these) and to enable 'emergent' features to be captured. While good initial mapping can provide a sound basis for focusing on what is the best data that will address the evaluation questions posed, it may still be necessary to adapt the data collection methods in the light of learning taking place: for example, if individual or organisational 'champions' (*levers and hubs*) emerge as key elements.

Some data may only be available from another sector or government department (where the intervention crosses different areas of responsibility) requiring new data sharing arrangements to be put in place. The time to establish this should not be underestimated.

Example:

The evaluation of the impact of a new 'holistic needs assessment' service for people with long term conditions on service delivery required referral data from multiple organisations, including clinical and social care services, and information about housing, benefits advice and transport use. Data from some of these sources were easier to access than from others.

Time and timing

Time can be an important issue in a complex policy evaluation. How long it will take for change to become apparent (emerge) and measurable may be unpredictable at the outset of an intervention and change may also continue longer than expected. It can also take time for 'complexity' features to become apparent.

Starting the evaluation late can risk losing opportunities to understand the initial 'starting position', build in data collection method from an early stage or track the delivery process over time. Ending the evaluation too soon raises the danger that changes continue to take place (which might be quite dramatic, if a 'tipping point' has yet to be reached) after the evaluation comes to an end. Alternatively, changes can reverse, if the system adapts and reverts to an earlier 'domain of stability'. An example of this is when changes to individuals or communities are not sustained after a short-term intervention ends.

Time may also be required before some aspects of complexity (such as feedback loops) and their impact become apparent. This will require a level of patience – and trust – on the part of the commissioners, in allowing evaluators time to immerse themselves in the situation being evaluated, and to analyse the data. The **modelling approaches** introduced in Section 4 can provide evaluators with tools to extrapolate from experience to date to explore how outcomes might evolve in the future.

3.5 Dissemination and embedding learning

The unpredictability of complex systems, the challenges around demonstrating causality and generating transferable findings can make the task of communicating evaluation findings particularly difficult. Commissioners are often keen to have a set of clear and relatively brief findings to disseminate to their policy colleagues and wider stakeholders. Instead, the findings may indicate patterns, principles, and potential scenarios, rather than straightforward answers to the question of whether the intervention worked.

The adoption of an *iterative approach*, with regular opportunities for sharing and reflection, is one way of testing the robustness of these. This also provides the opportunity to **embed** learning as this emerges, perhaps through making changes to the delivery of the intervention and the evaluation design. Early reporting (using interim reports or presentations) help to inform the audience about the complexity in the system, and understand why the findings are less clear cut and definitive than they might have liked.

3.6 Questions for commissioners

Table 3: Useful questions for commissioners for each activity in the evaluation planning and delivery process

Activity	Useful questions to address
Understanding the policy	<p>To what extent does the policy or programme, or its context, demonstrate the features of complexity outlined in Table 1?</p> <p>Have variations in the outcomes of the policy or programme, depending on the different contexts in which it is delivered, been considered?</p> <p>Would it be useful to involve additional expertise or stakeholders who can contribute to the understanding of this complexity?</p> <p>Would system modelling tools be useful for drawing up an initial 'map' of the policy or programme and how it is expected to work?</p>
Clarifying the purpose of the evaluation	<p>Does the policy or programme overlap with, or have implications for other government departments? Have these been consulted, or invited to contribute to planning its evaluation?</p> <p>Have different interests at a local level been considered?</p> <p>Have efforts been made to establish a common understanding between different stakeholders of the aspects of complexity which pose a challenge for the evaluation and how these can best be addressed?</p>

Evaluation design	
Evaluation purpose and questions	<p>Is it clear how the results of the evaluation will be used?</p> <p>Do the evaluation objectives and questions take into account the complexity identified?</p> <p>Have opportunities been built in to enable these to be revisited – and potentially revised – so that emergent changes in the policy and programme (or its context) can be accommodated?</p>
Choice of evaluation approach	<p>Are the evaluation methods and approaches being considered appropriate in terms of:</p> <ul style="list-style-type: none"> • the purpose of the evaluation? • the features of complexity identified? • feasibility, including the resources available and other constraints: • the expertise and/or software required, • the data available or obtainable, • budgetary constraints? <p>Are key stakeholders aware that the level of quantitative rigour and certainty of outcome may be limited, even when using sophisticated evaluation methods? Are they comfortable with the approach?</p> <p>Has there been a discussion about the trade-off between the quantitative rigour of findings and accuracy in terms of reflecting the complexity (and any associated uncertainty) present - given how findings will be used and the risks of providing misleading information?</p>
Data collection methods	<p>Have the data requirements of the methods chosen been considered and fully costed (taking into account the level of complexity)?</p> <p>Have opportunities been built into the plan to reconsider the data collection methods, in order to accommodate any unexpected elements emerging during the course of the evaluation?</p>
Conducting the evaluation	
Management	<p>Has the commissioning process taken the above points into account (e.g. is the specification sufficiently broad to allow for initial exploratory activities, the use of new and relatively untested evaluation approaches and adaptation of these as new information emerges)?</p> <p>Have opportunities for regular discussion between the evaluators, commissioners and other key stakeholders about any emerging developments been built into the plans?</p> <p>Has flexibility been built in to allow for changes to be made to the approach or time scale in order to reflect these developments?</p> <p>Has an adaptive management or agile process been considered?</p> <p>Have differences of view between members of the advisory or steering group been brought to the surface and discussed?</p>

Using and disseminating

Were recipients of the evaluation findings:

- given the opportunity to be involved in the evaluation design and dissemination?
 - kept informed of any changes in the programme or its evaluation?
 - given an indication of the complexity of the policy or programme, and how this might impact on the findings, or recommendations arising from these?
 - alerted to the fact that there might be further changes resulting from the policy or programme which, at the time of completion of the evaluation, are hard to predict?
-

4. Selecting complexity-appropriate approaches

Key points:

- There is a wealth of evaluation approaches and methods available that work well with complexity, particularly when implemented within a theory of change framework.
 - Which approach is chosen will depend on the complexity characteristics of the system, evaluation purpose and the feasibility of the available approaches.
 - Particularly useful ways of framing evaluations for complex settings include:
 - Participative approaches, including system mapping, which can bring actors together to generate deeper, shared understanding and provide a safe space in which participants can expose and air differences, and developmental approaches that involve stakeholders in the evaluation and as agents for change and are particularly useful supporting adaptive management approaches
 - Qualitative, theory-based approaches, which can be used to explore whether the policy is contributing to change, in what way, and the underpinning mechanisms of change, to provide rich information and potentially useful lessons for similar policies and contexts
 - Configurational (case-based) approaches, which help identify those factors, or combinations of factors, that appear necessary or sufficient, to success including contextual factors.
 - Computational system modelling, which can provide a 'virtual' counterfactual, a vision of what might have happened in the absence of the policy when it is not possible to establish an experimental counterfactual and allow the evaluator to project forward into the future and explore what further change may happen.
 - As noted in Section 3, detailed methodological requirements may only emerge over time, and evaluator and commissioners should regularly review the design to determine how well it is working and whether it should be modified.
-

This section provides guidance on how to choose the approach or combination of approaches appropriate for a particular evaluation. It should be used to supplement the guidance provided in the Magenta Book.

Before setting out the guidance, it is important to give some general points about terminology, different evaluation approaches and methods, and how these can be combined.

Terminology

For this Annex we have used the terms as shown in the box below.

Box 4: Strategy, approach, methods and tools

Strategy: An evaluation strategy refers to the whole evaluation plan, including the choice of approach, methods and tools, and how these will be put in place.

Approach: The approach refers to the overall, conceptual perspective adopted for the evaluation – for example, a theory-based approach articulates and tests a theory of how the policy causes or contributed to observed results. Different approaches can be combined in a hybrid design.

There can be a number of ways of delivering a particular approach. Examples of theory-based evaluation approaches include: realist evaluation, which focuses on identifying and confirming the underlying (decision making) mechanisms that lead to change, and contribution analysis, a systematic six step process that examines the causal chain, intermediate outcomes, risks and assumptions to build confidence around the contribution a policy is making to outcomes.

Method or technique: A method is a way of gathering, analysing or making sense of data. Methods can be, but are not necessarily, independent of the approach adopted. Examples of data collection methods include: surveys, questionnaires, interviews, realist interviews (specific to realist evaluation), desk reviews, and (critical) observation.

Tools: Tools help facilitate implementation of methods and techniques, for example NVivo, (thematic analysis); SPSS, STATA and R (statistical analysis); NetLogo (agent based modelling); OpenMarkov, Hugin and AgenaRisk (dependency modelling); Tosmana and EvalC3 (configurational analysis), Survey Monkey and Smart Survey (web-based surveys); checklists, templates etc.

These terms are not mutually exclusive, for example, qualitative comparative analysis, a way of establishing the necessary conditions, and the sufficient (good enough) combinations of conditions, for an outcome to emerge, can be applied as one method, alongside others, to process and make sense of the data emerging in an evaluation, or form the overall conceptual framework of the evaluation.

4.1 Evaluation approaches

Different ways of framing evaluation

There are a number of different ways approaches can be categorised to describe the overall conceptual approach to an evaluation. In this Annex, evaluation approaches have been clustered into six groups, differing mainly in how they establish conclusions. These are:

- **Participatory, emancipatory and adaptive approaches** such as **developmental evaluation**, **action research** and **peer challenge**, highly responsive and exploratory approaches in which stakeholders take an active part in the delivery of the evaluation providing real time feedback on the policy.
- **Theory-based approaches** which articulate a theory of how the policy is working to deliver change, then seek to test this to investigate whether, why or how the policy causes or contributes to observed results, and whether alternative explanations can be ruled out. Approaches include:
 - **Systems mapping and modelling**, used to generate, progress and test the theory of change through an iterative process of developing and testing a formal model of the system, and
 - **Generative causation approaches**, such as **realist evaluation** and **contribution analysis**, that seek to articulate underlying mechanisms or processes of change, and test the theory empirically to investigate whether, why or how the policy causes or contributes to observed results, and how context influences these.
- **Configurational case-based approaches** such as **qualitative comparative analysis (QCA)**, which support systematic analysis of a number of cases of the intervention being evaluated to identify the configuration of factors, or combinations of factors, that appear necessary or sufficient, to success.
- **Counterfactual approaches, including:**
 - **Experimental approaches** such as **randomised control trials**, and **quasi-experimental** such as **difference in difference**, which provide a usually quantitative measure of the extent to which any observed changes in an outcome of interest were caused by the intervention (or treatment) by means of a comparison of results obtained for a treatment group with those in a non-treatment control, and
 - **Predictive approaches**, which attempt to predict what would have happened in the absence of the treatment using statistical or simulation modelling; in its simplest form, predictive approaches use informant opinion as to whether impacts would have happened in the absence of an intervention.
- **Statistical association approaches**, that look for correlations between cause and effect or between variables, to explore the influence of (usually) isolatable multiple causes on a single effect, while controlling for 'confounders'.
- **Synthesis designs**, such as **realist synthesis**, which seek to draw conclusions by combining results from evaluations drawn from several contexts.

All of these approaches can, and for complex settings should, be applied within a *Theory of Change framework* as recommended in the Magenta Book. They can be more or less **participatory** in nature, depending on how far stakeholders involved in the design and delivery of the evaluation. In complex settings, involving many stakeholders with differing perspectives, participatory methods are particularly useful and may be essential. Participatory methods are particularly recommended for constructing the theory of change. **Participatory system mapping** and approaches such as

outcome harvesting and **most significant change** can help develop the theory of change where there are multiple, perhaps conflicting, viewpoints and where there are high degrees of uncertainty.

Table 7 in Appendix 1 provides additional descriptions of each approach, setting out their main strengths and weaknesses with respect to complexity-appropriate evaluation. **Table 8** provides more detail on some specific variants of these that can be particularly useful in complex settings.

Methods and tools

There are a very wide range of different data collection, analysis and synthesis methods available to evaluators. Often these are independent of the design selected. The Magenta Book provides an excellent overview of commonly used and some emerging methods.

Table 8 in Appendix 1 identifies some methods that are less commonly used but can be particularly helpful in tackling the challenges inherent in complex systems. They include techniques such as system mapping and modelling, smart data and data visualisation and the use of narratives, which can help generate deeper and more nuanced understanding about the system and the policy. The table does not seek to be comprehensive, but gives a flavour of the range of different types of method available.

Throughout this section additional information about designs, approaches or methods highlighted in **bold** can be found in Table 7 or 8.

Hybrid designs

None of these approaches or methods is mutually exclusive. In evaluation it is common to combine different methods so that conclusions are based on several different sources of information gathered in different ways. For complex evaluations, often what is required is a hybrid design in which two or more approaches are combined and tailored to meet the needs of the evaluation.

Example of a hybrid design

The Environment Agency and Greater Manchester Combined Authority (GMCA) required an evaluation methodology to assess the impact of embedding Environment Agency staff within a local authority such as the GMCA. A range of designs and methods were used in combination to measure tacit knowledge and impact. These included: **process tracing** to measure influence; **contribution analysis** to assess other impacts on observed change, **systems mapping and analysis** to understand the overall systems dynamics and effects, risk analysis to identify the risks to the outcomes and cost benefit analysis to quantify and monetise impact.

Evolution over time

Finally, as noted in Section 3, detailed methodological requirements may only emerge over time, as understanding of the system improves and the theory of change develops. It is therefore important

that the evaluator and commissioners regularly review the design and methods to determine how well they are working and whether they should be modified.

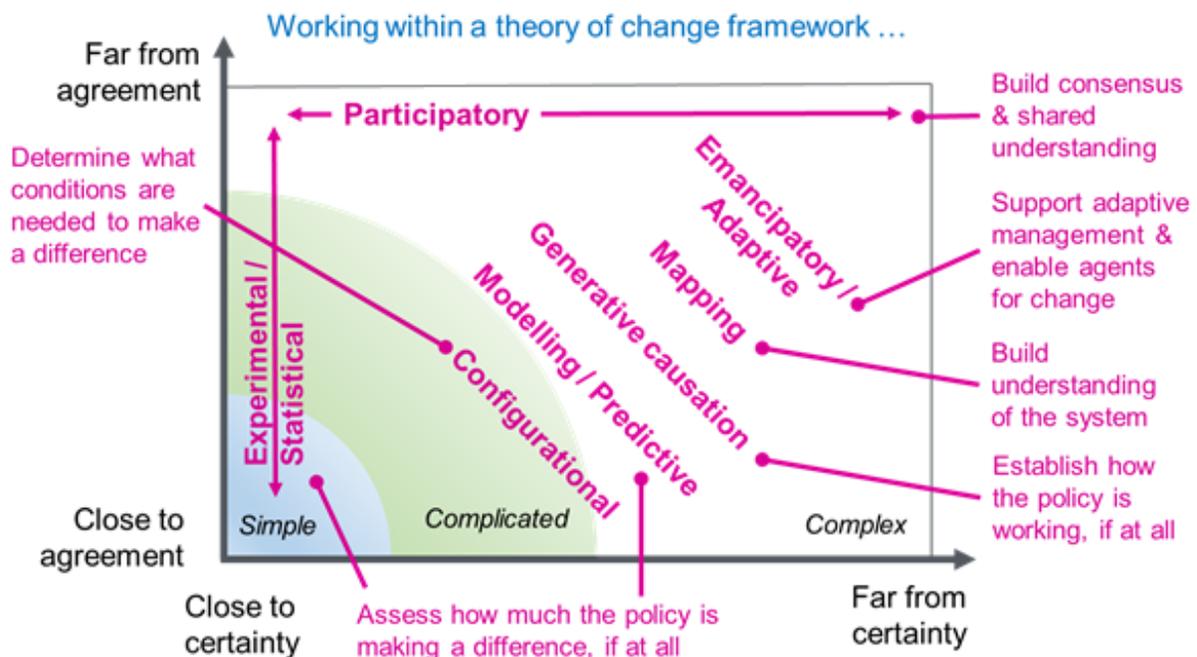
4.2 Selecting an approach

There is no simple, mechanistic way to make design decisions. Section 3 introduced three key, interrelated considerations that should inform the design of the evaluation: the purpose of the evaluation, the system attributes, including both the complexity and level of understanding of the system, and which designs are feasible with the resources available.

Figure 1 shows how consideration of the *purpose* of the evaluation and *system attributes* relevant to complexity inform which of the available approaches are likely to be most useful, subject to the final test of *feasibility*. So, for example:

- Where the policy is relatively simple (left hand portion of the figure), **randomised control trials, quasi-experimental or statistical approaches** may provide information on *whether, and to what extent*, the policy worked. To understand *why* the policy worked and *what aspects of it may work elsewhere*, it will be necessary to consider using e.g. **generative causation or configurational approaches** as well.
- Moving towards the upper right hand area of the figure, where there are low levels of understanding and agreement about the system and the policy, **participatory system mapping** will be useful to help develop a common understanding of the system and how the policy is intended to deliver impact (the theory of change), **adaptive approaches** can be used to generate information to support adaptive management in a timely manner.

Figure 2: Identifying potentially useful approaches



The rest of this section provides further guidance on the three key aspects of design.

Purpose

The choice of approach will depend on the purpose of the evaluation. This is often articulated as a set of questions to pose to the evaluators, but these must relate to how the evaluation findings will be used. This will enable the evaluation team, in discussion with the client, to focus resources on the most useful activities and outputs, responding flexibly to the changing needs of the evaluation.

Three reasons for evaluation identified in the Magenta Book are:

- **Learning.** The Magenta Book highlights the role of evaluation in providing evidence to help manage risk and uncertainty, especially in areas that are innovative or breaking new ground. This is clearly important in complex environments where there is a need to understand from an early stage whether a policy is working as intended and what can be done to improve it. Evaluation can inform decisions as to whether to stop and invest elsewhere. Evaluations can also contribute to the broader *knowledge base* providing information that decision makers can use to help inform future policy design.
- **Accountability.** Government makes decisions on people's behalf and has a responsibility to inform the public about the outcomes and value of the initiatives they put in place. Evaluations provide evidence of policy effectiveness for Spending Reviews and in response to scrutiny from bodies such as the National Audit Office and Public Accounts Committee.
- **Listening and building.** ensuring diverse voices are heard and building trust and legitimacy across stakeholders. This purpose can be particularly important for complex systems, where there may be many perspectives and low levels of agreement, and for policies that are participatory or have empowerment objectives.

In complex environments, learning and knowledge generation, and listening and building are likely to be important from the early stages of policy implementation, allowing the policy to be adapted in response to change. Later, the purpose may shift to examining whether the policy is having the desired impacts, whether any unintended consequences are emerging, and whether it represents good value for money.

Table 4 matches approaches and some particular complexity-appropriate methods against a range of typical evaluation questions.

Table 4: Answering evaluation questions

Evaluation question	Approach / method	Benefits
What is important to different groups, who can champion change? What levers are generating change, what may be inhibiting change?	Most significant change Participatory system mapping	Most significant change is an iterative participatory process that aims to clarify the values held by different stakeholders Participatory system mapping brings stakeholders together to build a system map and develop trust and mutual understanding Structured conversations about whether and how the policy is delivering change, can be used to develop the theory of change If begun at the option appraisal stage, forms a consistent framework for design, monitoring and evaluation through piloting and full implementation
	Big data and associated methods	Might ultimately allow local emergence of system dynamics that subsequently spread throughout systems to be understood Can provide near real time data to support learning
How well was the policy implemented? How can this be improved?	Participatory, adaptive approaches	Generates trust and shared understanding, champions and agents for change
	Experimental approaches	Provides robust evidence of whether a policy has made a difference, and to what extent, in a specific context
Is the policy making a difference, by how much? Is it delivering value for money?	Statistical association approaches	Weaker than experimental designs, but can provide a quantitative measure of the extent of impact, where it is not possible to define a counterfactual
	Predictive modelling approaches	Using computational system modelling to predict a counterfactual has the advantage of being able to account for contextual factors and deal with the emergent properties of complex systems
Is the policy making a difference, how? What conditions are needed?	Generative causation approaches	Explores the causal mechanisms (decision making and behaviours) or processes leading to change and the impact of contextual factors
	Configurational approaches	Recognises that different combinations of factors can lead to change and that these may include factors external to the policy
How sustainable is change likely to be? How can we adapt the policy to work	Generative causation approaches	Provides learning that is transferable to other contexts, and provides a basis for discussing sustainability, by uncovering the underpinning processes or mechanisms that lead to change and exploring the influence of context

elsewhere? How can the policy be improved?	Computational system modelling	Using information available on system behaviour to date, modelling can confirm whether the policy is contributing to change and how, and project forward to explore whether change is likely to be sustained in different scenarios
---	--------------------------------	---

System attributes

The complexity challenges inherent in both the policy intervention and the setting in which it is being introduced should be key determinants when selecting an approach. The evaluation design should be sensitive to the aspects of complexity that may be present, how well these are understood, and the range of different perspectives and levels of agreement or controversy that exist around the issues.

Table 5 summarises which approaches or methods are useful for different aspects of complexity.

Table 5: Tackling different aspects of complexity

Complexity challenge	Approach / method	How it helps
Sensitivity to context	Generative causation, configurational and system mapping and modelling	Treats context as a variable affecting outcomes, rather than a factor to be isolated and controlled, which in complex systems is often not possible
Openness/ open system	System mapping	Can guide division of a programme wide theory of change into multiple 'nested' theories to split complex programmes into more readily manageable segments without losing sight of the interactions between sub-systems and between the system and the wider environment
Multiple interactions and influences Long, indirect causal chains linking inputs to impacts	System mapping and modelling	Can capture the key influences and interactions and guide construction of complexity-appropriate theories of change Provides a framework for exploring the strength and importance of relationships affecting outcomes and impacts
Continual change, difficult to predict outcomes arising from e.g. feedbacks, non-linearity, tipping points, thresholds, emergence, path dependence, attractors	Computational system modelling	Provides exploratory tools in domains that are complex and "theoretically-insecure" (i.e. where there is no widely agreed and accepted theoretical underpinning; agent based modelling can bring emergent properties to the surface)
	Predictive modelling approaches Scenario analysis	Computational system modelling, including agent based modelling, can be used both to predict: <ul style="list-style-type: none"> what would have happened in the absence of a policy (a virtual counterfactual), and how outcomes may continue to unfold into the future

Complexity challenge	Approach / method	How it helps
		Scenario analysis, involves exploration of potential impacts under different alternative futures
	Big data and associated methods	Support modelling and might ultimately allow the evaluator to understand the local emergence of system dynamics that subsequently spread throughout systems
	Adaptive approaches	Enable rapid response to changes, supports adaptive management
Multiple-causality	Configurational approaches	Identify factors, or combinations of factors, that appear necessary or sufficient for to the success of a policy
	System models	Provide a way of representing and exploring multiple causality
Multiple perspectives	Participatory and emancipatory approaches	Generate understanding of multiple perspectives and mutual trust Feed the experiences and knowledge of informants into the evaluation to provide a source of evidence (key informants) Engage those stakeholders who are essential to delivering changes as active agents in its delivery.
Communication challenges	Participatory approaches	Generate deeper, shared understanding and trust among those involved
	Narrative methods	Help people engage effectively with the evaluation and communicate findings relating to the impacts on people and communities more meaningfully
	Agent based modelling	Provides a method of generating narratives that explain results in ways people can relate to

Feasible approaches

As well as considering which approaches are best suited to the aims of the evaluation and the level of complexity in the system, the resources available (and proportionate) in terms of funding, time, data and skills must also be taken into account.

Table 6 summarises the factors affecting feasibility, including the specialist skills and levels of resource required.

Table 6: Circumstances affecting feasibility

Approach	More feasible if ...	Less feasible if ...	Specialist skills and resources
Participatory and adaptive approaches	Appropriate range of stakeholders are willing and able to engage actively in the evaluation Useful when policy is participatory or has an empowerment objective	The findings must be seen to be rigorously objective and the evaluators need to retain independence from the system	Experience working embedded, or in partnership, with stakeholders May require experience working with hard to reach individuals and communities Support for evaluators immersed in what can be difficult environments should be provided to help evaluators maintain objectivity as far as possible, and manage stress Facilitation skills, the ability to work collaboratively and to keep strategic oversight of the work, will be important
System mapping and modelling	Appropriate range of stakeholders are willing and able to engage actively in the mapping or modelling exercise	More challenging where there is a high degree of ambiguity or many relevant influencing factors operating on different scales (time or geographical)	Specialist facilitators and system modellers Can be an efficient way of rapidly synthesising key informant knowledge and existing data (see synthesis below) to reach evidenced conclusions, so can be used where time and funding is limited or data is sparse
Generative causation	It is possible to formulate theoretical assumptions about the influence of context and the behaviours, attitudes and thinking of stakeholders (causal mechanism designs), or the causal processes or chains (causal process designs), leading to change	More challenging where there are complex (“distant”) relationships between outcomes of interest and drivers of interest	Experience in methods such as realist interviewing where relevant Investigating complex systems can be resource intensive due to the number of mechanisms, contexts or process steps involved, an ability to keep strategic oversight, to prioritise resource will be important

Approach	More feasible if ...	Less feasible if ...	Specialist skills and resources
Configurational approaches	Consistent information is available on (at least a small number of) factors that are assumed to affect the outcome	Fewer than 5 or 10 cases available for analysis	Software products are available to handle the mathematics required so that specialist skills are not required in this respect Collection of consistent information for large numbers of cases can be resource intensive
Experimental approaches	Groups of recipients and non-recipients can be established that are similar except for their receipt of the intervention, for example: <ul style="list-style-type: none"> • a pilot can be undertaken at the start including data collection in non-policy areas • a phased start across areas is possible • objective allocation is possible, for example using a cut-off score or random allocation • accidental factors influence allocation 	The relationship between outcomes of interest and drivers of interest is complex, with many potential confounding factors Small effect expected or effect builds up gradually over an extended period Data to support evaluation not sought until the policy is already established Data can be collected on only a small number of individual subjects in the recipient and non-recipient groups (less than around 30 in each group – but see this link for more advice)	Interpretation of experimental results for complex systems and for different contexts, will generally require some understanding of theory-based approaches Systems mapping can help identify parts of the system that are sufficiently isolated to allow an effective experimental design to be established
Quasi-experimental designs	A good understanding of relevant factors to ensure that systematic differences between the two groups are controlled for Good quality data is available for sufficient cases (beneficiaries, households, businesses etc.) for both the intervention and comparison groups and all significant factors	Small effect expected or effect builds up gradually over an extended time period Complex relationship between outcomes of interest and drivers of interest, with many potential confounding factors	Specialist knowledge can be required to deliver the detailed, complex analytical work required Again, interpretation of results and design is facilitated by some understanding of theory-based approaches or systems mapping

Approach	More feasible if ...	Less feasible if ...	Specialist skills and resources
Statistical association approaches	Good quality data is available for sufficient number of cases (beneficiaries, households, businesses etc.) for statistical analysis. Causal factors are independent of each other.	Small effect expected or effect builds up gradually over an extended time period Contextual factors such as cultural, institutional, historical and economic factors are expected to be important. Causal factors interact.	Interpretation of results for complex systems and for different contexts, will generally require some understanding of theory-based approaches

Appendix 1: Overview of available approaches

In **Table 7**, we describe the main strengths and weaknesses with respect to complexity of the conceptual approaches to complexity introduced in Section 4.1. Links to useful additional information are provided in this table.

Table 8 highlights some methods that are perhaps currently not widely used in support of evaluation in many domains, but that can be particularly helpful in tackling the challenges inherent in complex systems.

Table 7: The main conceptual ways of approaching evaluation

Conceptual approach	Description	Strengths and weaknesses with respect to complexity
<i>Participatory, emancipatory, adaptive approaches</i>		
Includes e.g. Developmental evaluation , Action research , Peer challenge , Most significant change (see below) , Outcome harvesting (see below)	<p>Stakeholders take an active role in the design and delivery of the evaluation.</p> <p>Includes:</p> <ul style="list-style-type: none"> <i>normative designs</i> where the experiences and knowledge of stakeholders provide a source of evaluation evidence (key informants) and <i>agency designs</i> where stakeholders are engaged as active agents for change. <p>A degree of participation is desirable in any evaluation and especially where understanding of the system, and how change will be achieved, is poor or where there are multiple perspectives.</p>	<p>Strengths: Allows deeper understanding of complex policy environments and interventions to be developed, based on the knowledge of people working within or affected by the system.</p> <p>Generates buy-in for the results of the evaluation, and champions for implementing recommendations. Supports continual learning and adaptive management of complex issues.</p> <hr/> <p>Weaknesses: Too deep involvement might be seen to compromise objectivity. Agency designs in particular blur the line between evaluation and system change.</p>

Conceptual approach	Description	Strengths and weaknesses with respect to complexity
<i>Theory-based approaches</i>		
<p>Qualitative system mapping approaches e.g. Participatory systems mapping (see also Table 8)</p>	<p>Used to generate, progress and test the theory of change in complex environments through an iterative process of developing and testing a formal model of the system (including the intervention). Qualitative mapping approaches can be used to generate deeper shared understanding of the system, the policy and how the policy is working to generate change. Qualitative mapping is a powerful tool for understanding and exploring complex systems. In complex settings it is also a good way of beginning the process of building a theory of change and a useful first step in building a quantitative, computational model of the system (see below) to further both qualitative and quantitative understanding of the system, develop and test the theory of change, explore outcomes over the longer term and construct virtual counterfactuals (see below).</p>	<p>Strengths: Recognises the high degree of interdependence and non-linearity in complex systems, and that complex systems undergo a continuous process of change. Helps identify metrics to support monitoring and evaluation. Can provide clear illustrations of why and how change is initiated and identification of levers and hubs that can be used to promote change. Provides a safe space to explore different perspectives. Enables a depth of shared understanding to be developed in a short time.</p> <p>Weaknesses: Requires specialist facilitation and mapping expertise not commonly available in the evaluation community. Will not pick up on the unknown, unknowns – emergent properties of systems that have not been observed, or cannot be postulated, by the people in the room – e.g. long timescale factors. May not provide simple definitive answers that are easy to communicate. Only the people involved in the process may find the maps useful.</p>
<p>Generative causation designs e.g. Realist evaluation, Contribution analysis</p>	<p>Articulates an intervention theory then seeks to test this theory empirically to investigate whether, why or how the policy causes or contributed to observed results and to see if alternative explanations can be ruled out. Includes:</p>	<p>Strengths: Strong on explanation (e.g. answering questions such as for whom does this work, under what circumstances does this work?) and on answering will this work elsewhere? Provides a framework within which many different types of information can be collated, synthesised and the findings clearly communicated.</p>

Conceptual approach	Description	Strengths and weaknesses with respect to complexity
	<ul style="list-style-type: none"> • <i>causal mechanism designs</i> that focus on identifying and confirming the underlying (decision making) mechanisms, and • <i>causal process designs</i> that focus on identifying and testing possible causal processes or chains. 	<p>Weaknesses: Generally weak on estimating quantities or the extent of impact. For complex systems can impose too linear thinking about relationships between cause and effect - use of systems modelling approaches (see previous) can mitigate this. Where there are complex (“distant”) relationships between outcomes and drivers of interest, data collection can be time consuming and will require careful prioritisation of resource.</p>
<i>Configurational approaches</i>		
<p>Case-based approaches e.g. Qualitative Comparative Analysis, Dynamic pattern synthesis (see Table 8)</p>	<p>Supports systematic comparative analysis of a number of cases of the intervention being evaluated to identify e.g. factors, or combinations of factors, that appear necessary or sufficient, to success. Cases that do not include the intervention can also be included as control cases showing what happens without the intervention. Generates a clear and explicit description of the combinations and disjunctions of factors that influence success.</p>	<p>Strengths: Good at dealing with moderate levels of complexity and interdependence, provides a practical and straightforward way to understand and communicate recommendations. Can be used to synthesise a small number of cases as well as larger numbers. Improves understanding of what contributes to change in different contexts and circumstances and allows factors of importance to success to be identified in complex multi-level interventions.</p> <p>Weaknesses: Not good at unpicking highly complex, highly interdependent combinations of causes. Cannot be used where there is only one or a very small number of cases. Not always clear how far findings are generalisable to other settings. Gathering consistent information on factors can be resource-intensive. This can be reduced where information collection systems are built into intervention design.</p>

Conceptual approach	Description	Strengths and weaknesses with respect to complexity
<i>Counterfactual approaches</i>		
<p>Experimental and quasi-experimental approaches e.g.</p> <p>Experimental:</p> <ul style="list-style-type: none"> • Randomised field experiments, • Randomised control trials (RCTs) <p>Quasi-experimental:</p> <ul style="list-style-type: none"> • Difference in difference studies, • Regression discontinuity designs 	<p>Provide a quantitative measure of the extent to which any observed change in an outcome of interest were caused by the policy, through comparison of results obtained for a treatment group that received the policy with those for a control group that didn't. Includes:</p> <ul style="list-style-type: none"> • <i>Experimental designs</i> where the causal effect of exposure to an intervention is estimated by comparing mean outcomes across groups created at random, where groups vary in their exposure to the intervention. It is usual practice for one group to remain unexposed and to act as a control group, and • <i>Quasi-experimental designs</i> that attempt to mimic the conditions of randomisation so any measured difference can be attributed to the intervention. This is typically done through matching or through a comparison of two groups where the outcome(s) of interest have historically moved in parallel. 	<p>Strengths: Good at answering questions such as 'did it work here', 'to what extent did it work here' and 'what would have happened without the intervention'? In the right circumstances can provide an unbiased estimate of the causal effect of exposure to an intervention. Obtaining measures of statistical uncertainty (for example confidence intervals) is in most circumstances relatively straightforward.</p> <p>Weaknesses: Weak at answering will it work elsewhere and at dealing with contextualisation: taking account of cultural, institutional, historical and economic settings. Cannot be used where there is only one or a small number of cases. Difficult to apply where causal pathways are complex, little understood and hard to unravel, making them unsuited to analysis through the experimental manipulation of single causal factors. Can, by their generalising nature, hide the initial emergence of new phenomena in a complex system.</p>

Conceptual approach	Description	Strengths and weaknesses with respect to complexity
<p>Predictive designs based on simulation modelling of policy processes e.g. Agent based models, Systems Dynamics, Markov Chain Monte Carlo.</p>	<p>Where research groups (e.g. intervention and control groups) cannot be created by random assignment, or mimicked using quasi-experimental methods, evaluators may attempt to model the effects of the intervention using statistical methods (see next), or simulation modelling. Simulation models that explicitly model the individual processes or mechanisms at work, and observe the behaviours of the system, can be used to develop a quantitative model of the system to:</p> <ul style="list-style-type: none"> • formalise and clarify understanding of the theory of change • construct virtual counterfactuals where insufficient data is available to use experimental methods • explore potential system behaviours over the longer term, which, depending on the modelling approach adopted, can include non-linear effects, tipping points, and potential unexpected outcomes, and • test variants to the policy design, effectively bridging the gap between evaluation and policy design and appraisal 	<p>Strengths: Recognises the high degree of interdependence and non-linearity in complex systems, and that complex systems undergo a continuous process of change. Can provide clear illustrations of why and how change is initiated and, depending on the modelling approach adopted, how unanticipated outcomes can emerge. Where models are developed collaboratively, can provide a safe space where different perspectives can be explored by groups. Quantitative modelling may be the only way of obtaining a numeric estimate of impact in very complex systems where the data to support experimental and statistical approaches is not available.</p> <p>Weaknesses: Choosing and implementing the best modelling approach requires specialist modelling expertise not commonly available in the evaluation community. The approach is called predictive. However, prediction, in the sense of predicting the future value of some measure, is often impossible in complex policy domains due to the characteristics of complex systems described in Section 1. Also, any model is necessarily an abstraction from reality, and since it is impossible to isolate sections of society, from outside influences, there may be unexpected exogenous factors that have not been modelled and that affect the outcome. Quantitative results should always be interpreted and used with this in mind. Models rarely provide simple definitive answers that are easy to communicate.</p>
	<p>Models should always be developed collaboratively with input from the policy client and stakeholders.</p>	

Conceptual approach	Description	Strengths and weaknesses with respect to complexity
<p>Predictive designs using statistical modelling approaches e.g. regression statistical models or structural statistical models (such as Bayesian Networks)</p>	<p>Where research groups (e.g. intervention and control groups) have not been created by random assignment, which is often very difficult in complex environments, evaluators may apply observational methods. These attempt to model the effects of the intervention using regression or structural statistical models (such as Bayesian Networks), etc.</p> <p>The evaluator essentially attempts to model exposure as well as outcomes as is the case with quasi-experimental methods, even though they are effectively in an observational setting.</p> <p>Where data to characterise the system is missing, variable or uncertain specialist modelling methods such as Bayesian belief nets (also known as dependency modelling) can be used to structure expert judgement and stakeholder opinion, explicitly capturing the uncertainty implicit in both judgements and real-world contexts and enabling groups to explore and clarify understanding of processes at work.</p>	<p>Strengths: In some cases evaluation techniques such as matching, regression and natural experiments can be applied successfully in cases where variation in exposure to the evaluand is considered 'naturally occurring' (variation in exposure is a given and treated as if it were independent of the evaluator and the policymaker).</p> <p>Approaches such as dependency modelling can provide a rapid way of exploring a complex setting and framing ongoing work. They can provide a framework within which different types of information, of different quality, can be brought together and a safe space where different perspectives and the impact of uncertainty can be explored by groups.</p> <p>Weaknesses: Requires data on sufficient cases (beneficiaries, households, businesses etc.) to be available to inform the model. Has difficulty in accounting for the effects of other, unexpected, contemporaneous factors, not captured in the data. Constructing valid comparison groups in such cases where the factors driving exposure are less clear is essentially more challenging than is the case with quasi-experimental. In general, weaknesses are similar to those facing quasi-experimental designs but potentially more severe. Where data is sparse, methods such as dependency modelling can use expert judgement to explore complex environments.</p>

Conceptual approach	Description	Strengths and weaknesses with respect to complexity
<i>Statistical association approaches</i>		
<p>Statistical association e.g. Longitudinal studies, Econometric studies,</p> <p>Statistical classification approaches e.g. Cluster analysis, Artificial intelligence (neural networks trained on case data).</p>	<p>Depend on the frequency of association between cause and effect (often called regularity) to establish causation. Seeks correlation between cause and effect or between variables to establish the influence of (usually) isolatable multiple causes on a single effect controlling for 'confounders'.</p> <p>No a priori mechanism model or theory of change is used to shape the analysis – however any correlations that do emerge from the statistics may be explainable by thinking about mechanisms.</p>	<p>Strengths: Can provide quantification of impacts where sufficient cases (beneficiaries, households, businesses etc.) are available for statistical analysis.</p> <p>Weaknesses: Not good at dealing with contextualisation – taking account of cultural, institutional, historical and economic. Work best when causal factors are independent of each other, but not if various causal interact with each other.</p>
<p>Synthesis designs e.g. Realist synthesis Bayesian synthesis 'Restatements' Systematic review</p>	<p>Combination provides structured ways for combining results from several different evaluations e.g. across different contexts Synthesis approaches can also be used as a method of synthesising different types of information within a single evaluation</p>	<p>Strengths: Explore the impact of context. Make use of available data and therefore can be relatively rapid and less resource intensive than other methods. Provide robust, systematic ways of combining information from different sources and different levels of quality. Can provide quantitative measures of the confidence that can be placed in findings and help focus further evaluative activity.</p> <p>Weaknesses: Reliant on available information which can be of variable quality, have gaps, or may have been gathered for different purposes. Techniques such as Bayesian and realist synthesis can help address these.</p>

Table 8: Highlighting some complexity-appropriate approaches and methods

Approach or method	Description	Strengths and weaknesses
Participatory systems mapping and fuzzy cognitive modelling	<p>Qualitative, participatory mapping approaches to draw out multiple perspectives and generate deeper shared understanding of the system, causal mechanisms and the process of change. Maps can be analysed using network analysis tools to help identify aspects of the system such as: key influences, trade-offs, feedbacks, and vulnerabilities.</p>	<p>Strengths: Allows a deeper shared understanding of interdependence and non-linearity in complex policy environments to be generated. Allows the structure of the system, and its influence on system behaviour to be explored. Supports continual learning and adaptive management of complex issues. Support construction of theories of change that reflects the influence of the wider system, presence of feedbacks etc.</p> <hr/> <p>Weaknesses: Depends critically on getting the 'right people in the room'. Will not pick up on the unknown, unknowns – emergent properties of systems that have not been observed, or can be postulated, by the people in the room – e.g. long timescale factors. May not provide simple definitive answers that are easy to communicate. Participants will get more value from the exercise than those who have not participated.</p>

Approach or method	Description	Strengths and weaknesses
Most significant change	<p>A form of participatory monitoring and evaluation. Significant change stories emanating from the field level, are periodically collected and the most significant selected by panels of designated stakeholders or staff. Once changes have been captured, groups sit down together, read the stories aloud and have regular and often in-depth discussions about why the stories are considered significant and in particular, to learn about the similarities and differences in what different groups and individuals' value. It provides some information about impact and unintended impact but is primarily about clarifying the values held by different stakeholders.</p>	<p>Strengths: Focusses the attention of teams on program impact. Enables meaningful conversations about impact and how it is generated even where there is a wide diversity and even conflicting viewpoints. It can be very helpful in explaining how change comes about (processes and causal mechanisms) and when (in what situations and contexts). It can therefore be useful to support the development of programme theory (theory of change, logic models).</p> <p>Weaknesses: It takes time and an appropriate project infrastructure and must be repeated through a number of cycles. It can be challenging to get engagement of the different groups involved in the process and to maintain their interest. Works best in combination with other options for gathering, analysing and reporting data. It doesn't provide comprehensive information about the impacts produced by an intervention.</p>
Outcome harvesting	<p>Outcome harvesting is a participatory methodology that collects ('harvests') evidence of what has changed ('outcomes') and, then, working backwards, determines whether and how an intervention has contributed to these changes. Various data collection methods such as interviews and surveys (face-to-face, by telephone, by e-mail), workshops and document review can be employed. It is repeated as often as necessary to understand what the intervention is achieving.</p>	<p>Strengths: Useful when it is not possible to define concretely most of what an intervention aims to achieve, or what specific actions will be taken over a multi-year period. Helps ensure unintended outcomes of interventions are identified. Is common-sense, accessible and engages informants quite easily.</p> <p>Weaknesses: Can be resource intensive. Only those outcomes that informants are aware of, are captured. Starting with the outcomes and working backwards represents a new way of thinking about change for some participants.</p>

Approach or method	Description	Strengths and weaknesses
Narrative methods Story telling	<p>Tacit, experience-based knowledge and emotional responses emerge more easily in stories than in more structured interviews. Storytelling encourages people to make sense of their own experiences, reflect on the array of contextual factors that influence outcomes, and articulate possible futures. Personal stories provide a human face to evaluation data and so can strengthen communication of key messages about a policy, including the impact on people and communities, their emotions and perspectives in ways that descriptions cannot.</p>	<p>Strengths: Stories can help evaluators engage stakeholders in evaluations, understand complexity, uncover the logic and emotion behind numbers and make sense out of complex data.</p>
Process tracing with Bayesian updating	<p>Process tracing with Bayesian updating is a recent development of process tracing. Process tracing is a case-based approach to causal inference used to test the strength of given qualitative and semi-quantitative observations in relation to specific theories. Central to process tracing is the idea that some pieces of evidence provide higher inferential power than others, and that the probative value of a piece of evidence for a specific causal mechanism is context-dependent and not necessarily related to sample size or the availability of a control group.</p> <p>Process tracing with Bayesian updating extends process tracing by allowing the evaluator to quantify the level of confidence that can be placed in the contribution claim. Quantification is carried out in a workshop setting using a rigorous technique known as Bayesian updating [for a plain English explanation see this link].</p>	<p>Weaknesses: Care must be taken to ensure that attention is given to issues of representativeness, confidentiality and verification, through the use of systematic processes to collect, analyse, select and communicate stories (disciplined inquiry).</p>
		<p>Strengths: Process tracing used in combination with contribution analysis has been shown to have greater power than using either approach alone. It improves the conceptual precision, clarity and scientific quality of theories of change and introduces additional rigour where quantitative methods are not possible.</p>
		<p>Weaknesses: While it should improve the efficiency of evaluation by focusing data collection efforts on evidence with highly probative data, the mathematical technique involved requires specialist input and it can be resource intensive. It should only be used once plausible and well-defined causal chains (for both the scheme and alternative explanations) have been developed.</p>

Approach or method	Description	Strengths and weaknesses
Dynamic Pattern Synthesis	<p>Dynamic Pattern Synthesis (DPS) is a method that assumes cases (like people, businesses, organisations, countries) are entities that have dynamic and relational connections with each other, in addition to shared external pressures. Over time, they therefore need to be considered for their changing similarities and dissimilarities in relation to each other.</p> <p>Current developments with DPS use a relatively small number of cases, with comparable multivariate data scores on ordinal and/or scale variables for these cases. Data needs to be available from at least two time periods (ideally more). DPS then analyses changing case patterns over time and explores the extent to which case patterns can then be explained, or not, by variable patterns.</p> <p>Similar to other case-based methods, the main strength of the approach is its ability to treat cases as real and discrete entities, rather than defining them as ideal types based on average or typical case scores.</p>	<p>Strengths:</p> <p>Similar to other case-based methods, the main strength of the approach is its ability to treat cases as real and discrete entities, rather than defining them as ideal types based on average or typical case scores.</p> <p>Another strength of DPS is its ability to combined current quantitative case-based methods such as Cluster Analysis (CA) and Qualitative Comparative Analysis (QCA), so as to minimise the reliability issues when these methods are used alone and to increase their combined validity.</p> <hr/> <p>Weaknesses:</p> <p>DPS is a new method that is still in its relative infancy. There is still much to be explored, such as: the optimum range of case sample size; the optimum number of variables for inclusion in any single model; advancing methods for the graphical presentation of DPS results; the extent to which Boolean algebra is a useful theoretical tool for summarising case-variable relationships; the ability of DPS to include further statistical and mathematical approaches and tools beyond the current combination of CA and QCA; ad how best to use qualitative methods to further validate DPS results.</p>

Appendix 2: Resources

The following books and resources will be helpful for those wishing to find out more about complexity and evaluation – see also Appendix 1 for links to information on specific evaluation designs and approaches

General

CECAN: Centre for Evaluation of Complexity across the Nexus

- **Centre for Evaluation of Complexity across the Nexus (CECAN)** was established to integrate complexity into policy evaluation and to enable more effective policy-making <http://www.cecan.ac.uk/>. The Centre is producing regular guidance (evaluation and policy practice notes), workshops and other publications to support the adoption of complexity-appropriate evaluation methods and approaches.

Systems innovation

- **Systems innovation** is an online platform for complex systems providing a wide variety of users with, information, learning and media content. Available at: <http://systemsinnovation.io/> [Accessed 8th November 2019]

Better evaluation

- An international collaboration to improve evaluation practice and theory by sharing and generating information about options (methods or processes) and approaches. Available at: <https://www.betterevaluation.org/>. [Accessed 8th November 2019]

Health

- Medical Research Council (2006) *Developing and evaluating complex interventions: new guidance*. [pdf]. Available at: <https://www.mrc.ac.uk/documents/pdf/complex-interventions-guidance/> [Accessed 8th November 2019]

Public health

Discussions about complexity and evaluation have been ongoing for number of years in the public health sector, and they now have their own UKCRC funded centre:

- **DECIPHer**: Centre for the Development and Evaluation of Complex Interventions for Public Health Improvement. Available at: <http://decipher.uk.net> . They have already produced guidance, examples and training in complex evaluation methods, such as its 2016 guidance on
- ‘Process evaluation of complex interventions’ published by the Medical Research Council. Available at: <https://www.mrc.ac.uk/documents/pdf/mrc->

International development

Some of the most developed work related to the issue of complexity and evaluation has come out of this sector. Examples include:

- **The Centre of Excellence for Development Impact and Learning (CEDIL)** (Available at: <https://cedilprogramme.org/>) develops and tests innovative methods for evaluation and evidence synthesis.
- Bamberger M, Vaessen J, Raimondo E. (2016). *Dealing With Complexity in Development Evaluation: A Practical Approach* (2016). SAGE Publications. Available at: <https://uk.sagepub.com/en-gb/eur/dealing-with-complexity-in-development-evaluation/book242113> [Accessed 8th November 2019]
- Ben Ramalingam and Harry Jones (2008) *Exploring the science of complexity: Ideas and implications for development and humanitarian efforts*. [pdf]. Overseas Development Institute. Available at: <https://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/833.pdf> [Accessed 8th November 2019]
- INTRAC. (2012). Praxis Paper 26: *Dealing with complexity through Planning, Monitoring & Evaluation (PME) Mid-term results of a collective action research process*. [pdf]. Available at: <http://www.intrac.org/data/files/resources/736/Praxis-Paper-26-Dealing-with-complexity-through-PME.pdf> [Accessed 8th November 2019]
- Elliot Stern. (2015). *Impact Evaluation, A Guide for Commissioners and Managers*. [pdf]. Big Lottery Fund, Bond, Comic Relief and the Department for International Development. Available at: https://www.bond.org.uk/data/files/Impact_Evaluation_Guide_0515.pdf [Accessed 8th November 2019]

Academic articles

The following include reflection and an overview of the way in which the concept of complexity is being taken up in evaluation theory and practice:

- Lasse Gerrits and Stefan Verweij (2015) Taking stock of complexity in evaluation: A discussion of three recent publications *Evaluation, October 2015; vol. 21, 4: pp. 481-491*.
- Chris Mowles (2014) Complex, but not quite complex enough: The turn to the complexity sciences in evaluation scholarship *Evaluation, April 2014; vol. 20, 2: pp. 160-175*.
- Gill Westhorp (2012) Using complexity-consistent theory for evaluating complex systems, *Evaluation, October 2012; vol. 18, 4: pp. 405-420*.

Practitioner guides

These practitioner guides outline the key characteristics of complex systems, the challenges these pose for policy makers and evaluators, and how these can be addressed.

- Australian public service. (2007). *Tackling wicked problems, a public policy perspective*. [online]. Available at: <http://www.apsc.gov.au/publications-and-media/archive/publications-archive/tackling-wicked-problems>. [Accessed 8th November 2019]
- Befani, B. (2017). *Choosing appropriate evaluation methods*. [online]. Bond.org. Available at: <https://www.bond.org.uk/resources/evaluation-methods-tool> [Accessed 8th November 2019]
- EU FUSIONS (2016). *Policy Evaluation Framework*. [pdf]. Available at <http://www.eu-fusions.org/index.php/download?download=255:policy-evaluation-framework> [Accessed 8th November 2019]
- HM Treasury. (2012) *Quality in policy impact evaluation: understanding the effects of policy from other influences*. [pdf] Crown Copyright. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/190984/Magenta_Book_quality_in_policy_impact_evaluation_QPIE_.pdf [Accessed 8th November 2019]
- Kelly, T. (2010) *Five Simple Rules for Evaluating Complex Community Initiatives*. [pdf]. The Anne E Casey Foundation. Available at: http://www.frbsf.org/community-development/files/T_Kelly.pdf [Accessed 8th November 2019]
- Preskill, H. and Gopal, S. (2014) *Evaluating complexity: propositions for improving practice* FSG. [pdf]. Available at: <http://www.fsg.org/publications/evaluating-complexity> [Accessed 8th November 2019]
- Public Health England (2018). *Guidance on process evaluation*. [online]. Gov.uk. Available at <https://www.gov.uk/government/publications/evaluation-in-health-and-well-being-overview/process-evaluation> [Accessed 8th November 2019]
- Stephens, A, Lewis, D, and Reddy, S. (2018). *Inclusive systemic evaluation for gender equality, environments and marginalized voices ise4gems: a new approach for the SDG era* [pdf]. Available at: <https://www.unwomen.org/en/digital-library/publications/2018/9/ise4gems-a-new-approach-for-the-sdg-era> [Accessed 8th November 2019]
- Quinn Patton, M. (2011) *Developmental evaluation: applying complexity concepts to enhance innovation and use*. New York: The Guilford Press. In addition to the main book, there are several associated 'user friendly' guides which spell out the key principles of complex interventions and their challenges. E.g. The J.W. McConnell Family Foundation (2010): *A Practitioner's Guide to Developmental Evaluation*. [pdf] . Available at http://www.dmeforpeace.org/sites/default/files/Dozios%20et%20al_Practitioners%20Guide%20to%20Developmental%20Evaluation.pdf [Accessed 8th November 2019]

- Quinn Patton, M. (2014). Evaluation flash cards: Embedding Evaluative Thinking in Organizational Culture. [pdf]Utilization-Focused Evaluation. St. Paul, Minnesota. Otto Bremer Foundation. Available at <https://iprc.iu.edu/spf/docs/Evaluation%20Flash%20Cards.pdf> [Accessed 8th November 2019]

HM Treasury contacts

This document can be downloaded from
www.gov.uk

If you require this information in an alternative format or have
general enquiries about HM Treasury and its work, contact:

Correspondence Team
HM Treasury
1 Horse Guards Road
London
SW1A 2HQ

Tel: 020 7270 5000

Email: public.enquiries@hmtreasury.gov.uk