Don’t panic: Bringing complexity thinking to UK Government evaluation guidance

Martha Bicket
University of Surrey, UK

Dione Hills
The Tavistock Institute, UK

Helen Wilkinson
Risk Solutions, UK

Alexandra Penn
University of Surrey, UK

Abstract
Central government guidance seeks to ensure and enhance the quality of practice and decision-making across – and sometimes beyond – government. The Magenta Book, published by HM Treasury, is the key UK Government resource on policy evaluation, setting out central government guidance on how to evaluate policies, projects and programmes. The UK Centre for the Evaluation of Complexity Across the Nexus was invited to contribute its expertise to the UK Government’s 2020 update of the Magenta Book by developing an accompanying guide on policy evaluation and ‘complexity’. A small multidisciplinary team worked together to produce a set of guidance, going through multiple stages of work and drawing on a variety of sources including academic and practitioner literature and experts and stakeholders in the fields of evaluation, policy and complexity. It also drew on Centre for the Evaluation of Complexity Across the Nexus’ own work developing and testing evaluation methods for dealing with complexity in evaluation. The resulting Magenta Book 2020 Supplementary Guide: Handling Complexity in Policy Evaluation explores the implications of complexity for policy and evaluation and how evaluation can help to navigate complexity. This article, designed primarily for practitioners who might be interested in this guidance and how it was developed, describes the processes involved, particularly related to the

Corresponding author:
Martha Bicket, Centre for Research in Social Simulation (CRESS), Department of Sociology, University of Surrey, Guildford GU2 7XH, UK.
Email: m.bicket@surrey.ac.uk
interdisciplinary dialogue and consultation with other key stakeholders that this involved. It also briefly outlines the content and key messages in the guidance, with reflections on the experiences of the authors in developing the guide – including the challenges and insights that arose during the process, particularly around the challenges of communicating complexity to a broad audience of readers.

**Keywords**
complexity, evaluation, government guidance, Magenta Book, policy analysis

**Introduction**

A complex and multifaceted policy intervention put in place within an equally complex setting can pose considerable challenges for those seeking to commission its evaluation and for evaluators themselves. In 2016, the UK Government’s Cross-Government Group on Evaluation began to revise the *Magenta Book* (HM Treasury, 2011: UK’s central government guidance on how to evaluate policies, projects and programmes). Recognising these complexity challenges, as part of this process the Cross-Government group asked CECAN (Centre for the Evaluation of Complexity Across the Nexus)\(^1\) to develop a supplementary guide on complexity in policy evaluation to accompany the main book.

CECAN was set up in 2016, as a partnership of experts from the fields of policy, evaluation and complexity science, who have since that time been working together with government departments and agencies to bring innovative policy evaluation methods to real-world complex evaluation challenges. One of the themes emerging from this collaborative work is that, when faced with complexity in a policy or programme, evaluators, policy makers and evaluation commissioners can feel considerable concern – even anxiety – that the evaluation strategies most frequently available to them were just not fit for purpose or effectively for addressing the challenges posed. Providing new guidance on complexity in policy evaluation was seen as one way to help those using it to navigate these challenges with greater ease: hence the phrase often used in events that accompanied the launch of the guide, and in title of this article: ‘Don’t panic’.

This article gives an account of CECAN’s experience of developing this guidance, the *Magenta Book 2020 Supplementary Guide: Handling Complexity in Policy Evaluation* (‘the Supplementary Guide’) (Bicket et al., 2020), reflecting on the challenges that arose, the joint learning that took place and the resulting guidance itself.

The rest of this section focuses on guidance for the analysis and evaluation of policies and programmes. It looks at existing guidance in the United Kingdom and elsewhere, and explores the motivations behind new efforts to develop guidance that deals specifically with ‘complexity’ and ‘complex systems’. The subsequent sections then set out the process of developing the *Supplementary Guide* (section ‘Developing the guide: Our process’), some of its key themes and messages (section ‘Key messages from the *Supplementary Guide*’), followed by reflections on the challenges and insights that arose as the authors developed the guide (section ‘Reflections’).

**Examples of existing evaluation guidance in the United Kingdom and elsewhere**

In this article, the authors adopt a broad definition of evaluation as encompassing a wide spectrum of ‘evaluative activity’, including both ex-ante (appraisal) and ex-post analysis. Some examples of guidance for both are set out below.
Government guidance for public servants exists to justify, inform and shape practice across a range of different remits and public functions. By providing central guidance, governments aim to enhance the quality of decisions that they and their public servants make.

Evaluation guidance has been published by local, national, supranational and intergovernmental organisations as well as other organisations of various shapes and sizes. Some, but not all, are domain-specific (e.g. the European Commission’s guidance to cost-benefit analysis for cohesion policy investment projects (2014); the United Nations Development Programme (UNDP) Handbook on planning, monitoring and evaluating for development results (2009); and the World Health Organization (WHO) Handbook on Monitoring and Evaluation of Human Resources for Health (Poz et al., 2009)). Other guidance sets out to serve as a general reference across a range of topics and policy domains. For example, in its Better Regulation Guidelines, the European Commission sets out key requirements and obligations for evaluations of all European Union (EU) policies, programmes and legislation (European Commission, 2017: 50). The guidance covers what is evaluation (defined to include fitness checks, final, ex-post and interim evaluations), why evaluate, and dictates procedure for EU officials.

In the United Kingdom, central government evaluation guidance is covered in two key and complementary references: the Green Book (HM Treasury, 2018) and the Magenta Book (HM Treasury, 2011). The Green Book provides guidance on appraisal, monitoring and evaluation, with a primary focus on supporting decision-making processes earlier in the policy cycle through guiding the development of business cases and Regulatory Impact Assessments, for example. In contrast, the Magenta Book gives more specific, detailed guidance on evaluation (particularly ex-post) and evaluation methods. Central guidance can also have a role in standardising practice and reporting, which may lead to greater consistency and comparison across outputs and appraisal of the different options available to decision-makers (e.g. Welsh Government, 2018).

Beyond central guidance, commissioners, managers and users of government evaluations may also draw upon guidance developed for and by networks of evaluation practitioners, such as the American Evaluation Association’s (AEA) Guiding Principles for Evaluators (2018), the BetterEvaluation Rainbow framework (2014) and the UK Evaluation Society’s (UKES) Guidelines for good practice in evaluation (2019).

**New efforts to develop guidance that deals with complexity and complex systems**

Despite this array of evaluation guidance, individual UK Government departments have highlighted ongoing gaps in the knowledge-base and available guidance on evaluation. For example, despite a rising number of impact evaluations conducted since 2000, a National Audit Office (2013) review of evaluation across government departments raised concerns about the lack of appropriate impact evaluations across a number of government departments. A key tension has been that, while this and other guidance (including the Green book and previous edition of the Magenta book), called for more robust, experimentally based, impact evaluations, others identified the need for guidance on how to evaluate programmes where this approach is difficult to apply.

For example, the UK Department for International Development (DFID) identified major gaps in evaluation knowledge and capacity contributing to knowledge gaps around effective interventions: ‘Most impact evaluations today are assessing relatively “simple” interventions,
consisting of simple programme theories, fewer stakeholders, clear goals, and operating in uncomplicated environments’. However, ‘most development programmes operate in uncertain and complex environments, and consist of packages of different initiatives promoted by different stakeholders; often without a clear definition of goals or the means to achieving them’ (Masset and White, 2019). DFID identified a need for the development of guidelines for, and testing of, new methods which are better suited to these complex and changing contexts where standard evaluation methods and designs may be less easily applied. Similarly, the UK Department for Environment, Food and Rural Affairs (Defra) is working to build evaluation capacity to deal with the complex nature of its policies and the contexts in which they are implemented (e.g. Defra, 2014), and commissioned a Complexity Evaluation Framework to equip evaluation commissioners with core considerations to ensure that evaluations ‘sufficiently consider the implications of complexity theory’ (Defra, 2020).

This need for complexity-appropriate methods and methodologies is well-recognised in the broader research and evaluation communities (e.g. Economic and Social Research Council (ESRC), 2015; Magro and Wilson, 2013; Vincent, 2012). It was into this space – and to address this need – that CECAN was launched in 2016 by a coalition of UK research councils and government departments and agencies.

The UK government periodically revises and updates its central guidance on analysis and evaluation. When the time came (in 2016) to update the existing 2011 edition of the Magenta Book, CECAN was approached by the Cross-Government Evaluation Group (CGEG) to produce an additional supplement on complexity and evaluation. Discussions about the challenges of dealing with complexity had already taken place between members of the CECAN partnership and individual members of CGEG, both prior to establishing the centre, and in discussions within its Advisory Board, on which several government analysts were represented. The idea of a supplement was discussed in quite general terms. Beyond agreeing on a broad outline, little specific guidance was given with regard to content, apart from requiring the need for the guidance to be compatible with, and not duplicate, material in the main Magenta book, and to be relatively short (10 pages). Both the main Magenta Book and the supplementary guidance took considerably longer than anticipated to finalise, with the final result, Magenta Book 2020 Supplementary Guide: Handling Complexity in Policy Evaluation published in March 2020 by HM Treasury at the same time as the 2020 edition of the Magenta Book was released.

**Developing the guide: Our process**

Developing the Supplementary Guide involved multiple stages of work during which the authors drew on a variety of sources including academic and practitioner literature; experts and stakeholders in the fields of evaluation, policy, and complexity; and CECAN’s own work developing and testing evaluation methods for dealing with complexity.

The first stage of developing the Supplementary Guide involved a review of the existing literature. This is not an unexplored area: the challenges that complexity and complex systems pose for policy and evaluation – and strategies for dealing with them – are discussed in a growing body of literature on the topic, with particular activity in the fields of health and international development. Some notable contributions include the Medical Research Council’s (MRC) Developing and evaluating complex interventions (Craig et al., 2006); United States Agency for International Development’s (USAID) Systems and Complexity
white Paper (Global Obesity Prevention Center et al., 2016); Dealing With Complexity in Development Evaluation (Bamberger et al., 2016); Tackling Wicked Problems (Australian Public Service Commission, 2007); Exploring the science of complexity: Ideas and implications for development and humanitarian efforts (Ramalingam and Jones, 2008) by Stephens et al. (2018); and Stern (2015). The review identified literature at the intersection of complexity and evaluation and sought to explore how complexity has been approached in different fields and evaluation domains. In particular, the authors sought to gather and consider understanding on key themes such as

- What are the properties of complex systems and how have these been addressed?
- Examples of complexity in policy or evaluation.
- The challenges that complexity poses for evaluation, and examples of evaluation failures.
- What evaluation approaches and methods have been used with complexity?
- The suitability of evaluation approaches and methods when working with complex systems.

The authors also consulted with the wider CECAN team of academics, researchers and policy and evaluation practitioners to collect further real-world examples of complexity in action in nature, society and policy.

The findings from these activities were iteratively appraised and synthesised into a draft guide through a series of author discussions and workshops. The guide evolved over time as the author team reflected on additional examples and perspectives. For example, the content and scope changed over time as the authors converged on a formulation of which aspects of complexity are most relevant in terms of evaluation practice, and this fed through into recommendations on approaches and practice. Acknowledging the variety of views in the field of complexity, it was decided not to give a specific definition for the term ‘complexity’ beyond the description of key characteristics of complex systems. The overall aim was to give broad guidance and points to be held in mind when dealing with complexity, whether this was within a broad policy or specific programme, and whether the complexity was within the intervention itself, its context, or indeed, within the evaluation process and its management.

An independent advisory group was convened to review and provide feedback on the developing guide from its early stages. Members were selected for their expertise in policy, evaluation and complexity and also to represent a range of potential users of the Supplementary Guide. The advisory group included individuals from the UK CGEG, the UKES Council, government departments and agencies and devolved government. The group met face to face with the author team and they used their knowledge and experience to contribute to deliberations and further shape the development of the guide.

The guide’s structure and recommendations were refined through two further author workshops. These workshops also provided an opportunity to consider CECAN’s own latest research and development of complex evaluation methods, reflecting on insights from CECAN’s case studies, fellowships and events. For example, ongoing work around ways to build evaluation capacity for dealing with complexity (by one of the authors of this article) and on commissioning challenges that can inhibit the uptake of complex appropriate methods (Cox, 2019) contributed depth and detail to section ‘Key messages from the Supplementary Guide,’ on management and commissioning. Other issues related to finding the best way to communicate complex terms in straightforward language (see section ‘Reflections’ below), and how to avoid duplication...
between different sections of the guidance. There were also debates about how best to address the difficulties of using experimental methods in complex settings (given the dominance of these in some areas of government evaluation practice) leading to an eventual nuanced position (see section ‘Reflections’). Regular contact with the team drafting the main Magenta Book revision led to discussions of how to describe the different stages of the evaluation process (to align more closely with the main Magenta Book) and the best terminology to use regarding different evaluation approaches or methods. Another concern was how to address the challenge that much of the guidance that applies to complex evaluation is also ‘best practice’ in any evaluation activity (e.g. consultation with stakeholders), but with added emphasis when the intervention and its setting was complex. The understanding accumulated throughout the entire process so far was consolidated and distilled down into key points under each of the four sections which now make up the final version of the Supplementary Guide. At 67 pages, this was somewhat longer than the original 10 pages envisaged at the outset.

In parallel, having converged on a key set of characteristics that complex systems can exhibit, the authors collaborated with a design expert to develop an accompanying set of visual images to be used in the guide. The aim of these images was to help lay people to recognise and better understand the implications of complexity. The authors sought to develop images that could be widely understood across different fields and sectors in order to facilitate the necessary conversations and decision-making between researchers, policy makers, practitioners and evaluators. The resulting designs were co-produced with contributions from the author team, other members of CECAN, as well as conference attendees at Relating Systems Thinking and Design 6 (RSD6) held in Oslo, Norway (Boehnert, 2018; Boehnert et al., 2018).

A final stage of the process involved inviting and responding to additional feedback from methods experts and the Magenta Book 2020 drafting team. This further improved the quality and coherence of the guide and allowed the authors to check alignment with the evaluation methods community and with the accompanying Magenta Book update itself. Any changes made at this stage related primarily to aligning terminology, rather than any significant change in content.

**Key messages from the Supplementary Guide**

The resulting Magenta Book 2020 Supplementary Guide: Handling Complexity in Policy Evaluation explores the implications of complexity for policy and evaluation and how evaluation can help to navigate complexity. It describes some of the challenges posed by complexity for evaluation and how these can be addressed – from key management considerations to approaches and methods that can help when evaluating in complex domains. The content is divided across four chapters: (1) why complexity matters, (2) the challenges of complexity to evaluation, (3) commissioning and managing evaluations and (4) selecting complexity-appropriate approaches. This section recaps the key points of each chapter as they appear in the Supplementary Guide.

**Why complexity matters**

The opening chapter of the Guide provides an accessible introduction to complexity and why it matters for policy-making. It describes and illustrates the properties of complex systems with the support of visual images and real-world examples. These highlight some of the common characteristics of complex systems that make their behaviour hard to predict; the fact that
they are in a continual process of change and that outcomes of policy intervention in one setting may have quite different results in another, because of the way that these interact with different historical circumstances and contexts. Policy interventions can also evolve in unpredictable ways over time as systems themselves adapt. While challenging for the evaluator, this also provides an opportunity for an appropriate evaluation strategy to support the policy implementation, helping to track changes over time, increase understanding of unexpected effects and support the review of implementation processes should things take an unexpected course.

The challenges of complexity to evaluation

The second chapter explores in more depth the challenges that complexity poses for evaluation, with relatable examples of how the challenges can manifest themselves in practice, illustrated with anecdotes from real-world evaluation failures. The point is made that, while the challenges themselves are similar to those in any evaluation, these become intensified, the more complexity is present in the system. Demonstrating causality (i.e. whether the policy led to a particular outcome) is a particular challenge in complex settings, because it makes the creation of a standardised intervention, or the isolation of a control group, harder. Evaluation activities may help to identify some of the challenges facing delivery of the intervention: such as parts of the system having a disproportionate influence, helping to mobilise or slow down change and making a system vulnerable to disruption. However, the same challenges can also significantly affect the evaluation, enabling or obstructing evaluation activities. With the system itself changing, evaluation strategies may also have to change to ensure that these remain appropriate.

Commissioning and managing evaluations

The third chapter provides guidance for those commissioning and managing an evaluation and includes a list of questions that commissioners can use as an aid at each stage of the evaluation planning process. Complex interventions may challenge both traditional notions of evaluation design, and usual practices around evaluation, and project management. Engaging key stakeholders from the outset in the design of the evaluation can bring insight into the complexity challenges which might emerge and help to ensure that an appropriate evaluation strategy is adopted. However, stakeholders may also have different levels of understanding of complexity, and may need to be alerted to the fact that the level of quantitative rigour and certainty of outcome might be limited in highly complex settings. Even when using sophisticated evaluation methods, there is a need for realism about what can be achieved. Those involved in the governance and management of evaluations will need to be flexible in responding to emergent changes to the intervention, or to system responses to the intervention, or new understanding as these emerge. Any of these may require a review, and changes to, the evaluation strategy adopted.

Selecting complexity-appropriate approaches

The final chapter provides guidance on the selection of evaluation approaches that match the complexity present in a particular policy intervention. There is a wealth of evaluation
approaches and methods available that work well with complexity: which ones are chosen will depend on the complexity characteristics of the system, evaluation purpose and the feasibility of the available approaches given the resources and expertise available.

The chapter (and an annex) includes tables that list the strengths and weaknesses of different methods and approaches in different circumstances, the specialist skills and resources required and how best to match different methods and approaches with different evaluation questions and types of complexity challenges. In general terms, participative approaches, including system mapping, are described as particularly useful for addressing diversity of viewpoints in a complex system, as they help to bring actors together to generate deeper, shared understanding of what is happening. Developmental approaches are useful in supporting adaptive management of the policy response when interventions are highly innovative, and the system is rapidly evolving. Qualitative and theory-based approaches are useful for exploring how and whether the policy is contributing to change and in understanding the underpinning mechanisms – and accompanying complexity features – through which change is taking place. Configurational (case-based) approaches help to identify those factors, or combinations of factors, that appear necessary or sufficient, for achieving the hoped for outcome (moving away from the assumption that one cause leads to one outcome). Computational system modelling can provide a ‘virtual’ counterfactual when it is not possible to establish an experimental counterfactual and can also allow the evaluator to project forward into the future and explore what further change may happen.

Reflections

In this section, the authors reflect on the challenges and insights that arose during the process of developing the Supplementary Guide. The views represented here are those of the authors of this article and not of CECAN, HM Treasury or any other UK government department, agency or group.

Time frame

Overall, the process of writing the guidance was a long one, with the initial invitation to write the guidance issued in 2016 (in the early days of the CECAN programme) and final publication taking place in 2020. The time period was determined largely by the parallel activity taking place on the revision of the main Magenta Book, since it had been clear from the outset that the supplementary guidance should complement rather than duplicate material in the main book, and match terminology and content (such as stages in the evaluation process) as closely as possible. However, the long time frame had its advantages. By 2019 (when the final draft was written), a great deal of experience had been gained within CECAN, on the most effective way to engage policy makers and analysts in issues related to complexity (through case study work described elsewhere in this special edition), and about the evaluation approaches and methods that were of particular interest. We had also learned more about the management and commissioning processes that support complexity-appropriate evaluation activities. This experience helped to clarify and refine the points being made, and how best to communicate these. The parallel activity of developing visual representations of key feature of complexity made an important contribution, providing an additional forum in which the team could refine
our descriptions of these features, finding suitable examples and wording that could best communicate these alongside the visual representations.

Had we known at the outset that we had such a long time frame to work with, we may have had a different process. It might, for example, have been useful to have more specialists with experience of specific evaluation methodologies as part of the team, or have planned, from the outset, for more rounds of consultation. However, overall, we felt it useful to keep the main writing task to a tight team of five, with consultation with others, whether within the CECAN team, and in Government Departments, undertaken as appropriate.

**Communicating complexity**

The first challenge faced during the development of the guide was how to present complexity. From the early stages of reviewing literature and initial attempts at defining complexity for a multidisciplinary audience, a tension appeared to emerge between accuracy and faithfulness to the complexity science literature on the one hand, and accessibility of concepts and the reader’s ease of understanding, on the other. Deliberations ranged from ‘how to present complexity in a simple way?’ to even whether and to what extent it was useful to explicitly mention complexity in the guide. A way forward emerged thanks to the author team’s multidisciplinary experience and expertise – which ranged from complexity science and mathematics to evaluation practice, environmental science and policy. Without a shared technical vocabulary, the authors found themselves sometimes resorting to using paper and pen to illustrate and debate different characteristics of complexity with each other. This highlighted the potential of images to help communicate these concepts to others, and eventually led to a parallel supporting study to develop visuals to illustrate the key characteristics of complexity in the guide (see Boehnert, 2018).

One particular challenge around communicating the implications of complexity – and the consequences of overlooking it – was the paucity of examples of and knowledge-sharing about evaluation failures in both the academic and evaluation practitioner literature. The authors highly commend efforts such as Hutchinson (2018) to fill this gap and feel that the evaluation and academic communities would benefit from greater openness and transparency about evaluation failures and reflection on the reasons behind them.

**Creating a useful and usable guide**

A closely related challenge was how to inform rather than overwhelm the reader about the challenges that complexity poses to both policy and evaluation. An impactful guide on complexity and evaluation ought to equip its audience with the know-how to recognise complexity and adapt how one conducts, manages and uses evaluation accordingly. It would be a step backwards, therefore, for readers to infer from the guide that ‘complex’ means ‘too difficult’ or ‘impossible’. The authors took great care to assess the tone of drafts and recommendations at various stages in the writing process and to emphasise the benefits of, and highlight solutions for, working with complexity. This is also reflected in the content of the guide: the second half of the *Supplementary Guide* (chapters 3 and 4) – which focuses on ways to deal with complexity – is more substantive than the first (chapters 1 and 2) where the challenges are introduced.

As the guide went through various iterations, the authors grappled with how best to structure the content. While writing is inherently linear, a complex evaluation is iterative and the
authors struggled with how to set out the challenges, guidance and tools in a way that reflects and supports the evaluator’s journey without needless repetition. Recognising also that the instinctive approach was to be writer-perspective-led and for the authors to document their own journey through the subject matter, they used the later author workshops to re-shape the guide to be reader-perspective-led by identifying and restructuring the guide around the key points that they wanted the reader to take away from each chapter. At the same time, these concerns had to be considered alongside the evolving direction that the new edition of the Magenta Book itself was taking as it was drafted. Working alongside the Magenta Book drafting team the authors strengthened coherence between the two guides through attention to alignment of terminology and also by referencing and building on the structure of the Magenta Book (e.g. its key stages in the commissioning and management of an evaluation: scoping, design, choosing appropriate methods, conducting the evaluation and disseminating and using the learning).

The extended time frame available for this work, described above, gave time to develop and apply this iterative process of development, allowing the authors us to stand back from the process of drafting and reflect on content, seek external feedback and review and revise.

**Complexity-appropriate methods**

Writing the Supplementary Guide, the authors were particularly aware of, and sensitive to, perceived tensions between proponents and opponents of experimental methods, and were especially cautious about how to situate themselves within this debate. After much discussion, the authors found the debate to be largely over-simplified. In the guide, the authors attempt to challenge the perceived dominance of experimental methods in a constructive manner; experimental methods such as randomised control trials can have great value in answering certain context-specific evaluation questions. However, when working with complex domains, meticulous care is needed to ensure that these methods are planned, applied and interpreted appropriately. Depending on the questions the evaluation seeks to address, an evaluation will almost always need to complement any results from experimental methods with other methods as well.

**The importance of adopting a ‘complexity mindset’**

A final challenge and reflection is around the importance of adopting a pragmatic mindset when working with complexity, and willingness to acknowledge and adapt to the uncertainty and change inherent in complex systems. One of the authors reflected in discussions that choice of evaluation approach and methods is often driven by being concerned about ‘getting the right answer’ but, when working with complex adaptive systems, the greater concern should be ‘how not to get an answer that’s very wrong’. In a complex setting, there are a number of reasons why an evaluation can result in wrong conclusions being drawn or generate findings that key stakeholders find difficult to accept. The most obvious cause is choosing an evaluation approach that fails to reflect the complexity involved, leading to overly simplistic, or misleading, conclusions being drawn. This risk is heightened when evaluation commissioners have not yet embraced the notion of complexity or have limited experience of the range of evaluation approaches available. Choosing the wrong design can lead to longer or less straightforward causal chains being overlooked, or important variations in impact occurring in
different settings ignored. Through a complexity lens, such variations can be seen as arising from multiple causes, such as different kinds of self-organisation and adaption taking place in different localities, different prior conditions (path dependency) or settings being impacted differently by their wider context (given the ‘open system’ nature of complex systems). Anticipating variations of this kind can help ensure that the choice of evaluation approach is one that is able to recognise and take such variations into account.

However, as several of the examples in Hutchinson’s (2018) book on ‘Evaluation Failures’ demonstrates, a complexity-appropriate evaluation design may be chosen, but then implemented, or managed, in a way that does not take into account the ‘emergent’ character of complex adaptive systems. Examples are given in which stakeholders did not have the implications of the evaluation approach fully explained to them, in terms of the kind of data that will be collected, or the kind of findings that the evaluation approach is able to generate. Messages about unpredictability or uncertainly can be particularly difficult to convey to those working in organisations, or sectors, where a high value is placed on simplicity and certainty, and less value accorded to exploration and learning.

Problems can also arise when the management style prevents evaluation designs from being reviewed – and potentially changed – in the event of an intervention evolving in an unexpected way, or being impacted by changes taking place in the wider environment in ways that have important implications for outcomes. When dealing with complexity, there is no one champion approach or method that can be relied on to give a definitively right answer. Instead, the Supplementary Guide recommends a more closely joined-up continuous process of learning and adjustment over the course of an intervention, and a willingness to adapt an evaluation and its design over time in response to new information and any changes.

Conclusion

The Magenta Book 2020 Supplementary Guide: Handling Complexity in Policy Evaluation explores the implications of complexity for policy and evaluation and provides a range of management, design and methods considerations that can help when evaluating in complex domains.

The key message that the authors would like readers to take away from the guide is that complexity can have significant and wide-reaching implications for both policy and evaluation – ignore it at your peril. Nevertheless, the message is ‘Don’t panic’ (as indicated in the title of this article); there is a wealth of evaluation management best practice and complexity-appropriate evaluation approaches that can help to navigate and work with complexity – these are outlined in the guide. Above all, the guide emphasises that complexity-appropriate evaluation requires an adaptive management approach and mindset, where change and uncertainty are anticipated and planned for. When working in complex policy environments, evaluations can be more resource intensive; getting your head around complexity takes time (as was poignantly well reflected in the authors’ experience of developing the guide), and there is no one universally best approach or method – complexity-appropriate evaluation must be tailored to its specific needs and context, and will likely need to be adapted over time. Much of this advice can be argued as being applicable to and relevant to the success of any evaluation. The authors do not disagree. However, when working with complexity, these considerations become even more critical to the validity and usefulness of any results.
The authors recognise that written guidance is not enough on its own to build capacity. The advice discussed here needs to be accompanied by a range of other activities, including training and development and organisational cultural change. CECAN’s current work is continuing to refine complexity-appropriate evaluation methods and apply them to public health, well-being, the local natural environment and enterprise domains in partnership with national, regional and local government, non-government funders and the third sector. This is supported by the provision of a range of training programmes, specialist workshops, briefing papers, webinars and seminars. Work in the area of complexity and evaluation is also now being taken forward by other organisations in the United Kingdom, such as the Centre of Excellence for Development Impact and Learning (CEDIL), and a revised version of the UK MRC guidance on developing and evaluating complex interventions.

Acknowledgements

The authors would like to thank ESRC for funding provided, and everyone involved in the development of the Magenta Book 2020 Supplementary Guide: Handling Complexity in Policy Evaluation, including co-authors Professor Nigel Gilbert and Ian Christie who provided valuable advice and contributions throughout the process, and those who provided feedback and guidance on various aspects of the Supplementary Guide: Elizabeth Robin, Beverley Bishop; Siobhan Campbell, Steven Finch, Simon Maxwell, Wendy Vander Neut, Mike Daly, Candice Howarth, Ben Shaw, Ben Fagan Watson, Barbara Befani, Professor Stephen Morris, Chris Rees and Emma Uprichard. All errors and omissions are the authors’ own.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship and/or publication of this article: This work was supported by the ESRC (grant numbers ES/N012550/1 and ES/S007024/1).

ORCID iDs

Martha Bicket https://orcid.org/0000-0001-9904-4182
Dione Hills https://orcid.org/0000-0002-0090-8751

Note

1. Centre for the Evaluation of Complexity Across the Nexus (CECAN) is an Economic and Social Research Council (ESRC) research centre. It was established in 2016 with the support of the ESRC, the Natural Environment Research Council (NERC), the Department for Environment, Food and Rural Affairs (Defra), the Department of Energy and Climate Change (DECC) (now the Department for Business, Energy and Industrial Strategy (BEIS)), the Environment Agency (EA) and the Food Standards Agency (FSA).

References

American Evaluation Association (AEA) (2018) American Evaluation Association guiding principles for evaluators: 2018 updated guiding principles. Available at: https://www.eval.org/p/cm/ld/fid=51
Australian Public Service Commission (2007) Tackling wicked problems. Tackling Wicked Problems: A Public Policy Perspective. Available at: https://www.taylorfrancis.com/books/e/9781849776530
Bamberger M, Vaessen J and Raimondo E (2016) Dealing with Complexity in Development Evaluation: A Practical Approach. Thousand Oaks, CA: SAGE.
Evaluation 27(1)

BetterEvaluation (2014) Rainbow framework. Available at: https://www.betterevaluation.org/en/rainbow_framework (accessed 5 December 2019).

Bicket M, Christie I, Gilbert N, et al. (2020) Magenta Book 2020 Supplementary Guide: Handling Complexity in Policy Evaluation. London: HM Treasury.

Boehnert J (2018) The visual representation of complexity: Sixteen key characteristics of complex systems. In: Proceedings of RSD7, relating systems thinking and design, 23–26 October, Turin, Italy. Available at: http://openresearch.ocadu.ca/id/eprint/2737/

Boehnert J, Penn AS, Barbrook-Johnson P, et al. (2018) The Visual Representation of Complexity: Definitions, Examples and Learning Points. Available at: https://www.cecan.ac.uk/sites/default/files/2018-06/The%20Visual%20Communication%20of%20Complexity%20-%20May2018%20-%20EcoLabs.pdf

Cox J (2019) How does the commissioning process inhibit the uptake of complexity-appropriate evaluation? CECAN. Available at: https://www.cecan.ac.uk/sites/default/files/2019-09/Jayne%C2%B0Cox%20Report%20-%20final%29.pdf

Craig P, Dieppe P, Macintyre S, et al. (2006) Developing and evaluating complex interventions. Available at: www.mrc.ac.uk/complexinterventionsguidance

Defra (2014) Making the Most of Our Evidence: A Strategy for Defra and Its Network. London: Defra. Available at: https://www.gov.uk/government/publications/evidence-strategy-for-defra-and-its-network

Defra (2020) Science search: Science and research projects: Complexity evaluation framework. Available at: http://sciencesearch.defra.gov.uk/Default.aspx?Module=More&ProjectID=20401 (accessed 12 February 2020).

ESRC (2015) Centre for evaluating complexity across the energy-environment-food nexus: Call specification – Full stage. Available at: http://www.esrc.ac.uk/files/funding/funding-opportunities/centre-for-evaluating-complexity/centre-for-evaluating-complexity-full-stage/

European Commission (2014) Guide to cost-benefit analysis of investment projects. Available at: http://ec.europa.eu/regional_policy/sources/docgener/studies/pdf/cba_guide.pdf

European Commission (2017) Better regulations guidelines [Commission staff working document]. No. SWD(2017)350. Available at: https://ec.europa.eu/info/law/law-making-process/planning-and-proposing-law/better-regulation-why-and-how/better-regulation-guidelines-and-toolbox_en

Global Obesity Prevention Center (GOPC), Global Knowledge Initiative (GKI), LINC, & ResilientAfrica Network (RAN) (2016) SPACES MERL Systems and Complexity White Paper. Retrieved from https://pdf.usaid.gov/pdf_docs/PA00M7QZ.pdf

HM Treasury (2011) The magenta book. Available at: https://s3.eu-west-2.amazonaws.com/golab.prod/documents/magenta_book_combined.pdf

HM Treasury (2018) The green book. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/685903/The_Green_Book.pdf

Hutchinson K (ed.) (2018) Evaluation Failures: 22 Tales of Mistakes Made and Lessons Learned. Thousand Oaks, CA: SAGE.

Magro E and Wilson JR (2013) Complex innovation policy systems: Towards an evaluation mix. Research Policy 42: 1647–1656.

Masset E and White H (2019) To Boldly Go Where No One Has Gone before: The CEDIL Evaluation Agenda. London: CEDIL.

National Audit Office (NAO) (2013) Evaluation in government. Available at: https://www.nao.org.uk/report/evaluation-government/

Poz MRD, Gupta N, Quain E, et al. (eds) (2009) Handbook on monitoring and evaluation of human resources for health: With special applications for low- and middle-income countries. Available at: http://www.euro.who.int/__data/assets/pdf_file/0011/200009/Handbook-on-monitoring-and-evaluation-of-human-resources-Eng.pdf
Ramalingam B and Jones H (2008) Exploring the science of complexity: Ideas and implications for development and humanitarian efforts. Overseas Development Institute. Available at: https://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/833.pdf

Stephens A, Lewis ED and Reddy SM (2018) Inclusive Systemic Evaluation for Gender Equality, Environments and Marginalized Voices (ISE4GEMS): A New Approach for the SDG Era. New York: UN Women.

Stern E (2015) Impact Evaluation: A Guide for Commissioners and Managers. London: Bond.

UK Evaluation Society (UKES) (2019) Guidelines for good practice in evaluation. Available at: https://www.evaluation.org.uk/app/uploads/2019/04/UK-Evaluation-Society-Guidelines-for-Good-Practice-in-Evaluation.pdf

United Nations Development Programme (UNDP) (2009) Handbook on planning, monitoring and evaluating for development results. UNDP. Available at: http://web.undp.org/evaluation/handbook/documents/english/pme-handbook.pdf

Vincent R (2012) Insights from Complexity Theory for Evaluation of Development Action: Recognising the Two Faces of Complexity. London: PANOS/IKM Emergent Research Programme.

Welsh Government (2018) Evaluation guidance for co-operation and supply chain development scheme projects. Available at: https://gov.wales/sites/default/files/publications/2018-09/co-operation-and-supply-chain-development-scheme-project-evaluation-guidance.pdf

Martha Bicket is a Senior Research Fellow at the University of Surrey and CECAN, working closely with policy makers and analysts towards the goal of better-informed decision-making and policy.

Dione Hills is a Principal Consultant Researcher with many years’ evaluation experience at the Tavistock Institute, recently also contributing to CECAN’s capacity building work related to complex evaluation.

Helen Wilkinson has over 30 years’ consulting and evaluation experience. She is a member of the CECAN team that developed the recent supplement to the Magenta Book on handling complexity.

Alexandra Penn is a Senior Research Fellow at the University of Surrey and member of CECAN whose work includes developing participatory complexity science methodologies for decision-makers.