

CECAN special issue of Evaluation

Special issue title: Policy Evaluation for a Complex World

Introductory essay

Title:

Policy Evaluation for a Complex World: practical methods and reflections from the UK Centre for the Evaluation of Complexity Across the Nexus

Authors:

Pete Barbrook-Johnson, Brian Castellani, Dione Hills, Alexandra Penn, Nigel Gilbert

Length:

5,277 for main text excluding refs.

Abstract

The value of complexity science and related approaches in policy evaluation have been widely discussed over the last twenty years, not least in this journal. We are now at a crossroads; this Special Issue argues that the use of complexity science in evaluation could deepen and broaden rendering evaluations more practical and rigorous. The risk is that the drive to better evaluate policies from a complexity perspective could falter. This special issue is the culmination of four years' work at this crossroads in the UK Centre for the Evaluation of Complexity Across the Nexus. It includes two papers which consider the cultural and organisational operating context for the use of complexity in evaluation, and four methodological papers on developments and applications. Together, with a strong input from practitioners, these papers aim to make complexity actionable and expand the use of complexity ideas in evaluation and policy practice.

Keywords: Complexity, complexity science, evaluation, evaluation methods.

Introduction

This introductory essay outlines the aims, contents, and contribution of this special issue of *Evaluation*. The issue, written by members of the UK Centre for the Evaluation of Complexity Across the Nexus (CECAN) and some of its collaborators, aims to bring together a set of ideas and methods for handling complexity in evaluation based on extensive practical experience in addressing these issues, from a research perspective, in UK central government evaluation. Some of these ideas and methods are novel, some are more established in research or theory. None, however, are widely applied in evaluation practice and that is key to the contribution we offer here. CECAN values innovation in practice equally to methodological and conceptual innovation. This is reflected in the content of this issue which aims to render the ideas and methods of complexity-appropriate evaluation as

practical, actionable, and rigorous. We are therefore pleased that the authors include many evaluation practitioners as well as academics.

We first outline the background to CECAN and describe the ‘turn to complexity’ in evaluation over the last twenty years, in order to frame the intended contribution of the special issue to evaluation research and practice. We then introduce the six papers in the issue, before concluding with some reflections on the CECAN experience and plotting some tentative ways forward.

What is CECAN?

CECAN has the core aim of ‘pioneering, testing, and promoting innovative evaluation methods and approaches’ (CECAN, 2018). It is made up of an interdisciplinary team of researchers and practitioners from nine universities and five practitioner organisations. It was established in early 2016 as a result of a successful bid by the project team to a group of research funders: the UK Economic and Social Research Council (ESRC), the Natural Environment Research Council (NERC), and four UK government departments and agencies: the Environment Agency, the Food Standards Agency, the Department for the Environment, Food and Rural Affairs (Defra) and the Department of Business, Energy and Industrial Strategy (BEIS). The initiative arose because of a strong feeling among some government social researchers that the methods recommended for policy evaluation in, for example, the *Magenta Book* (the UK government's official guidance to evaluation; HM Treasury, 2011), failed to do justice to the complexity of the policies and their contexts that they were being asked to evaluate. A small group of government social researchers approached the research councils to lobby for an academic centre to provide advice about complexity-appropriate evaluation approaches and methods, along the lines of the existing ‘What Works Centres’ (<https://www.gov.uk/guidance/what-works-network>). After discussion, it was decided that a stand-alone centre should be established, funded jointly by the research councils and those government departments and agencies that had a strong interest in policy evaluation. The ‘Nexus’ in CECAN refers to the food, energy, water, and environment nexus. CECAN’s initial remit focussed on these policy areas, partly because of the domains of interest of its funders (both the government departments but also an interest from UK research councils - see <https://thenexusnetwork.org/>); and that these are policy areas which exhibit the characteristics of complexity.

The bid from CECAN was somewhat unusual, in comparison with typical UK research council proposals for example, in that it described a process for setting up the Centre based on a spirit of co-design with policy and evaluation practitioners. This is, in part, why CECAN deliberately left undefined the specific evaluation arenas it would explore and only decided on them following discussion with the four government partners after the Centre had been established. This co-produced, responsive approach is reflected in the content of this issue: a set of articles that constitute an organic and evolving narrative, reflecting the priorities of the UK government and the spaces for innovation in method and practice that CECAN came to occupy.

CECAN is a multi-disciplinary, multi-institutional team of academics and evaluation practitioners. It is this diversity and the close involvement of a range of stakeholders that

constitutes one of the strengths of CECAN. The Centre has also gained greatly from close contact with 'live' policy issues through working with policy analysts and makers in the UK central government and from their interest in and enthusiasm for learning about new approaches to evaluation. The initial project funding lasted until February 2019, after which ESRC has provided a further tranche at a reduced scale to partially support the Centre until 2022. The responsibilities of CECAN's original funders revolved around food, energy, water and the environment, and these gave us our initial policy domains to focus on. However, our remit has broadened in response to demand and opportunities, so we now also operate in public health, transport, and social policy areas, although with a continuing emphasis on national policies.

One of the main outcomes of our first four years has been a *Supplementary Guide: Handling Complexity in Policy Evaluation*, which is included in the new 2020 edition of the *Magenta Book* (HM Treasury, 2020; 2020a). This addendum set out to provide state-of-the-art guidance to address the original objectives of the government social researchers who were seeking support for better ways of handling complexity in policy evaluation. This special issue represents a longer form, and more academically focussed, output which further addresses and complements CECAN's objectives. The issue does not cover the entirety of CECAN's work; some has, or will be, published elsewhere in domain or methodological journals, and some is covered by non-disclosure agreements necessitated by the proximity of our work to sensitive policy topics. We recommend visiting our website – www.cecan.ac.uk - for additional information on case studies, projects, papers, blogs, and webinars.

The 'complexity turn' in policy evaluation

We do not wish to grapple with definitional issues around 'complexity' in evaluation. Others have already done this well, for example Gates (2016), Moore et al. (2019), and Walton (2014, 2016). Nonetheless, it is worth making clear that CECAN uses a broad and inclusive definition of 'complexity', which aims (albeit critically) to include ideas and methods from across the complexity sciences, including the study of complex adaptive systems and systems science and thinking, so that these various perspectives and tools can be brought to bear on the complex systems problems faced in evaluation.

Even in the early days of systems theory and systems science, there were scholars exploring the implications of complexity theory and complex systems thinking (including its emerging methods) for public policy and its evaluation (Hammond 2010). Examples include the Macy Conferences of the 1940s and 1950s, the establishment of the Tavistock Institute (circa 1946-47), the development of cybernetic theories of society and social influence (particularly through second-order and socio-cybernetics); and the massive inroads that complexity theory concepts and methods made into managerial studies, organisational psychology and systems engineering, particularly in the UK (Stacey, 2001; Stacey 2006; Klir 2001). Here one thinks also, in terms of methods, of such novel techniques as systems mapping, soft systems modelling, and simulation – all of which are still used today (Castellani and Hafferty, 2009). In the 1960s and 70s, Fred Emery, while at the Tavistock Institute made a considerable contribution to reigniting systems and complexity ideas in organisational theory (e.g. Emery 1971; Emery and Trist 1965).

Still, for most, the real possibilities of envisaging how to apply complexity theory to public policy came with the 'complexity turn' in the social sciences in the later 1990s (Byrne, 1998; Urry 2003; Byrne and Uprichard, 2012; Byrne and Callaghan, 2013; Anzola et al., 2017; Barbrook-Johnson et al. 2020), which followed and participated in the academy-wide excitement over the ideas coming out of the *Santa Fe Institute* (established 1984) and the new complexity sciences (Cairney 2012; Gerrits 2012; Geyer and Rihani 2012; Gilbert et al 2018; Haynes 2015; Koliba 2012; Morçöl 2013).

One of the fields in the social sciences which has embraced the ideas of complexity most, including in this journal, is evaluation. For example, if you do a search of *Evaluation* with the key word 'complexity' you get 726 hits between 1995 and 2020. More specifically, if you search the title word 'complex' you get forty-seven hits. Of these hits, forty-four are research articles. The first, '*Evaluation in Complex Policy Systems*' was published in 2000 by Ian Sanderson. Themes for these forty-seven articles follow the typical trajectory for the 'complexity turn' in the social sciences, starting with, as one would expect, articles seeking to introduce the ideas of complexity to evaluation (e.g. Sanderson, 2000; Barnes et al, 2003; Stame, 2004). From here, in this journal and others, things became more focused, with articles exploring empirical implications of complexity (e.g. Callaghan, 2008), and then more recently beginning to review applications (e.g. Mowles, 2014; Walton 2014; Gates, 2016; Reynolds et al. 2016; Gates, 2017; Walton, 2016a; Williams, 2015; Ofek, 2016). There is also a strong 'realist' bent to the complexity work during these years, with eight articles in this journal having the term 'realist' or 'realism' in the title (e.g. Westhorp, 2013). Articles have considered how complexity can inform existing approaches (e.g. Reynolds, 2015; Gates, 2017; Garcia and Zazueta, 2015). Others have thought about how to develop complexity-appropriate evaluation questions (e.g. Larson, 2018). Gates (2017) and Barbrook-Johnson et al. (2020) also directly consider how systems and complexity approaches are understood and used by evaluators.

There are many specific complexity-appropriate methods presented in the evaluation literature, including agent-based modelling (Morell et al. 2010), case study approaches (Woolcock, 2013), causal loop diagrams and Systems Dynamics (Dyehouse et al. 2009; Fredericks et al. 2008; Grove, 2015), Process tracing (Schmitt and Beach, 2015), Qualitative Comparative Analysis (Befani, 2013; Blackman, 2013; Blackman et al. 2013; Byrne, 2013), social network analysis (Durland and Fredericks, 2005; Drew et al, 2011), and co-produced or exploratory approaches to working with stakeholders (Copestake, 2014). Moving toward 2020 we get closer to some of the concerns of this special issue, as scholars begin to become more critical and self-reflexive about the value and use of these ideas and methods, particularly in terms of the realities of using them in the practice of evaluation (e.g. Gerrits and Verweij, 2015; Walton, 2016).

A renewed energy and practicality to complexity

This twenty-year journey, from ideas and proposals, through applications and methods, to reflections and critiques (with many iterations and cycles in-between), brings us to this special issue. We see this issue, and indeed CECAN's existence, as a key reflection and renewal point in this journey to understand and use complexity in evaluation. (It is worth

noting, this timeframe is actually a subplot to the ebbs and flows in popularity of complexity and system science ideas across a range of disciplines since the mid 20th century). There have been successes, but also failures and missed opportunities.

The potential for interest in, and use of, complexity to wane if its value is not advocated and realised, is very real. We see the risk of this flowing from the application of ideas and methods in ways that are not practical, pragmatic, or accessible to practitioners and 'non-converts'. This point, amongst all others, is primary. Therefore this special issue aims to contribute to a renewed excitement and energy behind the use of complexity in evaluation by innovating around its practical challenges, particularly in terms of methods. We seek to promote a renewed energy that captures the initial excitement of twenty years ago, but which is more pragmatic, actionable and accessible. Importantly, we promote an energy led by practitioners as much as by academics, and an energy which is being felt and catalysed across the public sector and the institutions and branches of government. There is a balance to be struck here: applying the ideas and methods of complexity in central government, ideas is sometimes seen to question or undermine assumptions of central control or coordination. Our position, outside government departments but with strong connections, has allowed us to tread this narrow path.

Our aims are made real in the issue in its organisation into two sections. Made up of two papers, the first section focuses on the wider cultural, organisational, and bureaucratic operating context of evaluation (Bicket, Hills, Penn and Wilkinson; Cox and Barbrook-Johnson). It addresses a lack of focus in the existing literature on these wider contexts in which the use of complexity in evaluation is actually happening. The second section presents four papers all offering methodological innovations. They are 'innovative' either in the sense they apply cutting-edge method(s), developed by CECAN and its collaborators, to evaluation for the first time (Schimpf, Barbrook-Johnson and Castellani; Barbrook-Johnson and Penn), or they combine or connect two existing methods for the first time (Befani, Elsenbroich and Badham; Wilkinson, Hills, Penn and Barbrook-Johnson). In all four papers, these are methodological developments that are underpinned by experiences of co-producing research and methods with policy analysts and makers, and which are grounded in practical concerns about how value might be added to actual evaluations.

The contribution of this special issue is thus in broadening and deepening the turn to complexity in evaluation described above, not only in terms of methods, but also in terms of application in practice, the actionable framing of complexity; and an understanding of the operating context. Its strength lies in bringing these new and existing ideas and methods to government. Reflecting CECAN's mode of working, this research has used the tools of co-production and participatory approaches to implement the use of complexity and its methods.

This practicality reflects the fact that the special issue and CECAN are primarily concerned with methods and practice and less with wider consideration of values, ethics, or politics. We adopt the view of evaluation as a technical craft which provides information for the policy making process but is somewhat removed from politics and values (in a normative sense). This position is not taken in opposition to other views, but simply reflects the origins and aims of the Centre. Our acceptance of this understanding of evaluation is in line with

some previous work on complexity in evaluation. Mowles (2014) is perhaps most effective in critiquing previous work from this standpoint; arguing that the adoption of complexity in evaluation scholarship has often under-achieved because it has accepted the prevailing constraints or understandings of what evaluation is and what its purpose should be. While many members of CECAN, especially those with complexity and system science backgrounds, would be sympathetic to this critique, readers will not find examples of how to address this critique. They will, however, find examples of how complexity can be used to transform the evaluation process as it is used in practice. This reflects CECAN's origins, purpose and aims and its close relationship and co-production with actors holding a variety of perspectives within government. It has been our experience that policy makers, analysts, and evaluators in the UK central government are unsure of how to make sense of the complexity they intuitively understand in their policy domains. We have sought to help primarily in that effort, which is transformative in its own right. They were less keen to explore more ambitious or radical understandings of evaluation and complexity. However, the use of co-production should not preclude this by default. Nonetheless, we acknowledge other perspectives and outline thoughts on more differentiated ways forward in the concluding section to this essay.

Despite not providing outright critiques or alternatives to our primary technical and instrumental understandings of evaluation, there is an important theme through the special issue which does expand our horizons beyond a focus on the views or practice of evaluators alone. Throughout several of the papers (most notably Bicket et al, 2021; Cox and Barbrook-Johnson, 2021; and Wilkinson et al, 2021) runs a theme of considering commissioners and users of evaluation (not just evaluators themselves), both as objects of study, and also as key stakeholders with whom to engage. The study of the operating context, wider sets of stakeholders and building of methods need to be undertaken together and this is also part of the contribution of the issue.

Overview of articles in the issue

Without further ado, let us now introduce the articles in the special issue. The first two papers form the practical pragmatic 'hull' of the issue. They allow readers to understand the social, cultural and organisational operating context of evaluation in the UK, and the use of complexity within it, which sets the ground for the methodological papers that follow. Martha Bicket, Dione Hills, Alexandra Penn and Helen Wilkinson in *'Don't panic: bringing complexity thinking to UK government evaluation guidance'* do this by reflecting on the process of developing guidance on handling complexity in evaluation for the previously mentioned UK government *Magenta Book* on evaluation (HM Treasury, 2020; 2020a). They explore the interdisciplinary dialogue and consultations that informed the development of the guidance, and consider the range of challenges and insights that arose during this process. These relate particularly to how to communicate complexity and to bridge the substantial divide between the abstractions of complexity science and the realities of real-world evaluation practice. The paper also highlights the social and intersubjective nature of framing policy domains as complex systems, making clear the role of all involved, including evaluators, as stakeholders and participants in these systems.

Next, Jayne Cox and Pete Barbrook-Johnson in *'How does the commissioning process hinder the uptake of complexity-appropriate evaluation?'*, apply a more formal approach, using findings from interviews with nineteen evaluation commissioners and contractors. They use these to explore how the commissioning process inhibits the uptake of ideas and methods from complexity, and outline a number of practical ways forward to remove or negotiate the barriers. These include using scoping studies to understand the nature of complex areas, having early and more open dialogue with contractors, using less prescriptive tenders, and operating flexible contract management to allow contractors to update and revise plans. Taken together, we hope these two papers will leave readers well-placed to advocate for the use of complexity in evaluation in practical, intelligent, and sophisticated (or put simply, non-naïve ways), thereby bypassing many of the potential stumbling blocks in applying complexity ideas and methods in evaluation.

The second set of papers all present a specific method, or combination of methods, which are in some sense complexity-appropriate, practical, usable, and useful. That is, they help develop or implement an analytical approach which allows us to understand, and act in, the systems in which we operate as complex adaptive systems while also complementing and contributing to existing evaluation approaches.

Pete Barbrook-Johnson and Alexandra Penn in *'Participatory Systems Mapping for complex energy policy evaluation'*, present the application of a new approach to the construction and analysis of system maps in two real-world evaluation case studies in the energy policy domain. The authors set out the context of a range of 'systems mapping' methods, including Theory of Change diagrams, before outlining their approach in detail, both in abstract form and through description and reflection on the two case studies. They also describe how and when to choose to use the approach or not, and other ways it can be used in evaluation. The novelty here is in how these maps are analysed using causal network analysis to create submaps, and in the nuances around map construction, which are strongly stakeholder-led.

Building on the theme of systems mapping, Helen Wilkinson, Dione Hills, Alexandra Penn and Pete Barbrook-Johnson in, *'Building a system-based Theory of Change using Participatory Systems Mapping'*, develop and test a methodology for turning large complex systems maps into more practical Theory of Change diagrams. They do this to allow evaluators to take advantage of the breadth and scale of large systems maps, while still producing outputs such as a Theory of Change map that they can work with in usual ways. This is important because large system maps can be overwhelming, with researchers, practitioners, and users not sure how to make use of them. Making the connection to Theory of Change maps clear and manageable helps unlock some of their value.

Next Barbara Befani, Corinna Elsenbroich and Jennifer Badham in *'Diagnostic Evaluation with Simulated Probabilities'* describe the use of agent-based modelling to support evaluation that utilises a diagnostic approach and Bayesian Updating. These approaches often rely on the presence of empirical data (which is not always available) or subjective expert opinion, to ground estimates of outcomes and contributions of evaluands. Agent-based modelling provides an additional source for these estimations which can complement or replace empirical data and/or expert opinion. This is a novel way of thinking about how simulation can support and complement evaluative activities.

Finally, Corey Schimpf, Pete Barbrook-Johnson and Brian Castellani in '*Cased-based modelling and scenario simulation for ex-post evaluation*', present a case-based approach (i.e. using single cases or sets of cases as a key driver of analysis, rather than focussing on variables across multiple cases) to data analysis which can be harnessed in evaluation. They outline a free and open source software tool - [COMPLEX-IT](#) - which supports users in undertaking such analysis, and describe four detailed examples of how the approach could be used. The tool has been developed across a range of different activities and projects, and the potential value to evaluation is described in detail using published evaluations, with a focus on exploring how their approach can complement the analyses already being used. The value is in making it easy for evaluators to take a fresh, potentially more complexity-appropriate perspective, on the data they are collecting, and what additional data might be useful.

Taken as a whole these four papers give a sense of some of the frontiers of, and the breadth of, complexity-appropriate methods and their application. We hope readers find value not just in learning about the methods and how they can be used, but in the rich and practical reflections on how to use these methods in constrained real projects.

Conclusion

In this introductory essay we have attempted to frame and bring together the set of six varied papers which make up this issue. The issue presents a suite of ideas and methods for handling complexity in evaluation, considers their operating context and application, and explores how CECAN has developed and put them into practice with government partners. The issue demonstrates CECAN's emphasis on innovation in practice as well as intellectual innovation. As a whole, we hope the issue makes some of the ideas and methods of complexity-appropriate evaluation more practical and contributes to the articulation of their value to evaluators, commissioners, and users of evaluation.

Reflections on the 'CECAN experience'

As a way of further stitching together the six papers, we thought it useful to summarise what we see as the most important cross-cutting insights from our experiences in CECAN. Over four years of conducting research and interacting with a range of stakeholders actively involved with evaluation, we have developed perspectives reflected throughout this special issue, including the following.

How policy and analyst teams respond to complexity and the reasons for this

Government interventions in social, environmental, and public health domains are acting on complex adaptive systems. Leading policy teams, analysts and evaluators in government (and other practitioners), acknowledge this, and act appropriately. However there are still many instances where the case for complexity still needs to be made forcefully and where the push for simple certain answers remains strong. This is most commonly driven by a desire for simple narratives to deliver to senior staff and ministers, as well as through lack of time for detailed analysis or reflection. Nonetheless, we found many evaluators, analysts, and policy teams crave the time and

space to develop capacity for complexity-appropriate evaluation but are constrained by available resources.

Policy teams need a clear picture of evaluation options in complex settings

Being aware of practicalities, makes it easier for policy-teams to articulate clearly their complexity-related needs. Evaluators and advocates for complexity approaches can help by giving a strong and creative lead to those in government on what successful complexity-appropriate evaluation looks like. Equally, we found most evaluation practitioners instinctively want to evaluate in a complexity-appropriate way but are constrained by practicalities, time and capacity. Government departments can help unleash these instincts by being clear in their demand for complexity-appropriate evaluation from internal evaluation specialists and evaluation contractors.

Embedding evaluation across the policy cycle

There are many potential benefits embedding considerations of evaluation throughout the policy process - i.e. when identifying priorities, designing policies and as part of ex-ante appraisal. This improves the quality of evaluations and increases their ability to deal with complexity; by improving, for example, data availability and collection, and policy design for learning. The best policy teams and evaluators are skilled in integrating evaluative thinking in this way, but there are often timing and bureaucratic barriers which mean evaluation is not embedded. To help realise the benefits of embedding evaluation across the policy cycle, evaluation users have to be willing to accept and use preliminary or provisional findings, which evaluators also need to be willing to provide.

Complexity highlights important decisions about the scope of an evaluation

Artificially separating questions about the intervention itself; shorter-term outcomes or longer-term effects; and the institutional and wider context, can be counterproductive. From a complexity perspective these need to be seen as a whole even if not all aspects are necessarily addressed at the same time. On the other hand highlighting inappropriate structures and objectives can be threatening to those working in government if only because this can be perceived as difficult, expensive, or time consuming to resolve.

The complex nature of policy making has implications for evaluation design

The policy evaluation landscape is shifting; government is evaluating more varied policies, often made up of multiple projects, delivered in collaboration with many stakeholders, at multiple governance levels – central, regional and local. This means effective evaluations have to devote more effort to evaluation design and consideration of engagement and co-production with stakeholders. This requires more evaluation time and resources as well as the deployment of participatory evaluation methods and skills, able to support stakeholder engagement.

Repurposing complexity-appropriate evaluation collaboratively

We have observed many misconceptions about what complexity means for evaluation. Complexity sometimes poses new approach and methods challenges for evaluation. But as often as not, complexity intensifies the challenges evaluation and policy practitioners

already tussle with. Complexity-appropriate evaluation methods do not have to be sophisticated or highly technical. Often the best strategy is to co-produce and customise together with users and stakeholders a combination or hybrid of existing methods, that are adaptable, iterative and appropriate. Many methods can be repurposed for complex policies and contexts. The innovation is getting them in the right place, in the right hands, and using them in the right combination, at the right time.

We hope that the papers in this issue together with this essay begins to clarify why communicating and working on complexity-appropriate evaluation can be difficult. There are often disagreements, misunderstandings, and false starts. When we have been fortunate enough to succeed, the key has been finding the sweet spots at the junctures of accessibility and practicality; the push for innovation and exploration in methods; and underlying ideas and philosophy. Even when such sweet spots are found, what is often hidden from view is the 'one step forward, two steps back' nature of much of this work. Just as opportunities to apply complexity-appropriate ideas and methods are found and realised, organisational priorities, objectives or personnel can change, or promising explorations can simply run out of steam or hit a wall. What becomes more important than individual projects are the enabling conditions for exploration on multiple fronts. CECAN was able to provide those conditions for involved researchers and practitioners and their collaborators. Despite the lessons and achievements documented in articles in this special issue, it is still an open question as to how far those conditions can be preserved in the longer term.

Future directions

The application of the ideas and methods of complexity in evaluation is still not widespread in practice nor a default option in all situations where this would be appropriate. It therefore continues to be important to apply and document complexity-appropriate methods and approaches in evaluation, to articulate their value, and to make them accessible and practical. As this issue has highlighted, attention also needs to be directed towards the social, cultural, political and bureaucratic operating context. Understand how such operating contexts help and/or hinder complexity-appropriate evaluation will make it easier to reframe existing evaluation approaches and for these approaches to be more effectively advocated and applied.

This pragmatic and practical emphasis should not impede the potential for genuine innovation in ideas, approaches and methods: more exploratory efforts are also needed. An obvious way to pursue this would be to reconnect the literature and ideas in evaluation scholarship on complexity, with the state-of-the-art in applied complexity and systems science. Cross-fertilisation between the niche that has emerged in evaluation and the broader literature on applied complexity is likely to be difficult but fruitful. In CECAN we have made some of these connections but have not explored them fully because of our initial aims and purpose. Applied complexity science can also be a dogmatic domain at times, with an insistence on formalism which excludes some of evaluation's 'softer' and more practical work. A related direction of travel we have considered but not yet fully developed is to connect more closely to the growing interest in data science and artificial intelligence for government and policy. Many of the frameworks and lessons; and the practice we have developed, will be applicable here too. The strong grounding in theory and explainable, usable insights that evaluation provides also have the potential to complement

and inform data science and artificial intelligence approaches. The emerging field of explainable artificial intelligence (xAI) is ripe for input from evaluators.

We believe there are compelling reasons to embed complexity approaches and tools across policy making, rather than just limiting it to ex-post evaluation as it is often understood in UK government. In particular, embedding complexity approaches at the earliest stages of policy formulation has the potential to be transformative and far reaching. This entails both greater use of complexity-appropriate tools in ex-ante appraisal, but also more widely throughout policy design, ultimately allowing policy to exploit rather than simply to mitigate complexity. Moreover, taking a complexity-appropriate approach to policy making in general would reduce the emphasis on different policy stages or cycles. Instead the aim would be creating constant and agile iteration and feedback between policy design, policy implementation and evaluation, adaptively responding to complex societal systems as they reveal new dynamics or change in response to interventions.

Finally, we believe it is also worthwhile to follow Mowles' (2014) call for more radical and ambitious interpretations of what complexity means for evaluation scholarship and practice. Fundamental questions such as, 'What should we evaluate?' and 'What is the role of evaluation?' are part of this, but so too is an active campaign to promote the push for evaluation to be more complexity focused and grounded. Users, commissioners, and sometimes evaluators themselves, are very good at defending or entrenching the status quo, and this is perhaps one of the influences that has constrained CECAN's work within the scope of a 'traditional' understanding of evaluation. Given the successes of the last twenty years, part of the continuing journey for the application of complexity in evaluation will be to continue to push for better, rather than just more. This will involve 'better' more nuanced methods, but will likely also mean more acceptance of the uncertainty and unknowability implied by understanding the world as a complex adaptive system, and reconciliation of this with the needs of policy makers for clarity. Complex systems are fundamentally unpredictable. Our work does not 'solve' this dilemma. It does however begin to expand the repertoire of approaches and tools able to explore uncertainty with greater rigour in real-world policy settings. Hopefully this also provides policy makers with some of the tools they need to ask the right questions, fill some gaps in understanding; and make informed decisions within the context of complex systems. As such, we believe a renewed energy and excitement around complexity-appropriate evaluation is a great opportunity to improve policy evaluation; reducing the risk of inadequate evaluations whilst increasing rigour and impact. The choice we face is not whether to do a 'complex' or a 'simple' evaluation, but whether to do an effective or ineffective one.

References

- Anzola D, Barbrook-Johnson P and Cano JI (2017) Self-organization and social science. *Computational and Mathematical Organization Theory* 23(2): 221–257.
- Barbrook-Johnson P, Proctor A, Giorgi S and Philipson J (2020) How do policy evaluators understand complexity? *Evaluation* 26(3): 315-332.

- Barnes M, Matka E and Sullivan H (2003) Evidence, Understanding and Complexity. *Evaluation* 9(3): 265–284.
- Befani B (2013) Between complexity and generalization: Addressing evaluation challenges with QCA. *Evaluation* 19(3): 269–283.
- Blackman T (2013) Rethinking policy-related research: Charting a path using qualitative comparative analysis and complexity theory. *Contemporary Social Science* 8(3): 333-345.
- Blackman T, Wistow J and Byrne D (2013) Using Qualitative Comparative Analysis to understand complex policy problems. *Evaluation* 19(2): 126–140.
- Byrne D (1998) Complexity theory and the social sciences: An introduction. Psychology Press.
- Byrne D (2013) Evaluating complex social interventions in a complex world. *Evaluation* 19(3): 217–228.
- Byrne D and Callaghan G (2013) Complexity Theory and the Social Sciences: The state of the art. London: Routledge.
- Byrne D and Uprichard E (2012) Useful complex causality. In H. Kinkard, (ed) *Oxford Handbook of Philosophy of Social Science*. Oxford: Oxford University Press, pp. 109-129.
- Cairney P (2012) Complexity theory in political science and public policy. *Political Studies Review* 10(3): 346-358.
- Castellani B and Hafferty FW (2009) *Sociology and Complexity Science*. Springer.
- CECAN (2018) Policy Evaluation for a Complex Worlds. CECAN report available at www.cecan.ac.uk/resources.
- Copestake J (2014) Credible impact evaluation in complex contexts: Confirmatory and exploratory approaches. *Evaluation* 20(4): 412-427.
- Drew R, Aggleton P, Chalmers H et al. (2011) Using social network analysis to evaluate a complex policy network. *Evaluation*, 17(4), 383–394.
- Durland MM and Fredericks KA (2005) An introduction to social network analysis. *New Directions for Evaluation* 107: 5–13.
- Dyehouse M, Bennett D, Harbor J et al. (2009) A comparison of linear and systems thinking approaches for program evaluation illustrated using the Indiana Interdisciplinary GK-12. *Evaluation and Program Planning* 32(3): 187-196.
- Emery F (Ed.) (1971) *Systems thinking*. Penguin.

- Emery F and Trist E (1965) Causal Texture of Organisational Environments. *Human Relations*, 18(1): 21-32.
- Fredericks K, Deegan M and Carman J (2008) Using System Dynamics as an Evaluation Tool: Experience from a Demonstration Program. *American Journal of Evaluation* 29(3): 251-267.
- Garcia JR and Zazueta A (2015) Going beyond mixed methods to mixed approaches: A systems perspective for asking the right questions. *IDS Bulletin*, 46(1): 30-43.
- Gates EF (2016) Making sense of the emerging conversation in evaluation about systems thinking and complexity science. *Evaluation and Program Planning* 59: 62-73.
- Gates EF (2017) Learning from seasoned evaluators: Implications of systems approaches for evaluation practice. *Evaluation* 23(2): 152-171.
- Gerrits L (2012) *Punching clouds: an introduction to the complexity of public decision-making*. Emergent Publications.
- Gerrits L and Verweij S (2015) Taking stock of complexity in evaluation: A discussion of three recent publications. *Evaluation* 21(4): 481-491.
- Geyer R and Rihani S (2012) *Complexity and Public Policy: A New Approach to 21st Century Politics, Policy And Society*. Routledge.
- Gilbert N, Ahrweiler P, Barbrook-Johnson P, Narasimhan KP and Wilkinson H (2018) Computational Modelling of Public Policy: Reflections on Practice. *Journal of Artificial Societies and Social Simulation* 21(1): 14.
- Grove JT (2015) Aiming for utility in “systems-based evaluation”: A research-based framework for practitioners. *IDS Bulletin*, 46(1): 58-70.
- Hammond D (2010) *The science of synthesis: Exploring the social implications of general systems theory*. University Press of Colorado.
- Haynes P (2015) *Managing complexity in the public services*. Routledge.
- Klir G (2001) *Facets of systems science*. Springer.
- Koliba C and Zia A (2012) 'Complexity Friendly' Meso-level Frameworks for Modelling Complex Governance Systems. In *COMPACT 1: Public Administration in Complexity*, Eds. L Gerrits and P Marks, 119–139. Emergent Publications.
- Larson A (2018) Evaluation amidst complexity: Eight evaluation questions to explain how complex adaptive systems affect program impact. *Evaluation* 24(3): 353-362.
- HM Treasury (2011) *The Magenta Book*. Available at <https://www.gov.uk/government/publications/the-magenta-book>

HM Treasury (2020) The Magenta Book. Available at <https://www.gov.uk/government/publications/the-magenta-book>

HM Treasury (2020a) Magenta Book Supplementary Guide: Handling Complexity in Policy Evaluation. Available at <https://www.gov.uk/government/publications/the-magenta-book>

Moore GF, Evans RE, Hawkins J, Littlecott H, Melendez-Torres GJ, Bonell C and Murphy S (2019) From complex social interventions to interventions in complex social systems: Future directions and unresolved questions for intervention development and evaluation. *Evaluation* 25(1): 23-45.

Morcol G (2012) *A Complexity Theory for Public Policy*. Routledge.

Morell JA, Hilscher R, Magura S et al. (2010) Integrating Evaluation and Agent-Based Modeling: Rationale and an Example for Adopting Evidence-Based Practices. *Journal of MultiDisciplinary Evaluation* 6(14): 32–57.

Mowles C (2014) Complex, but not quite complex enough: The turn to the complexity sciences in evaluation scholarship. *Evaluation* 20(2): 160–175.

Ofek Y (2016) Matching Evaluation Approaches to Levels of Complexity. *Evaluation Review* 40(1): 61–84.

Reynolds M (2015) (Breaking) The iron triangle of evaluation. *IDS Bulletin*, 46(1): 71-86.

Reynolds M, Gates EF, Hummelbrunner R, Marra M and Williams B (2016) Towards Systemic Evaluation. *Systems Research and Behavioral Science* 33(5): 662-673.

Sanderson I (2000) Evaluation in Complex Policy Systems. *Evaluation* 6(4): 433–454.

Schmitt J and Beach D (2015) The contribution of process tracing to theory-based evaluations of complex aid instruments. *Evaluation* 21(4): 429-447.

Stacey R (2001) *Complex Responsive Processes in Organizations: Learning and Knowledge Creation*. Routledge, London, UK.

Stacey R (2006) *Complexity and Creativity in Organizations*. Berrett – Koehler, San Francisco.

Stame N (2004) Theory-Based Evaluation and Types of Complexity. *Evaluation* 10(1): 58–76.

Urry J (2003) *Global complexity*. Wiley.

Walton M (2014) Applying complexity theory: A review to inform evaluation design. *Evaluation and Program Planning* 45: 119–126.

Walton M (2016) Expert views on applying complexity theory in evaluation: Opportunities and barriers. *Evaluation* 22(4): 410–423.

Walton M (2016a) Setting the context for using complexity theory in evaluation: boundaries, governance and utilisation. *Evidence & Policy: A Journal of Research, Debate and Practice* 12(1): 73–89.

Westthorp G (2013) Developing complexity-consistent theory in a realist investigation. *Evaluation* 19(4): 364–382.

Williams B (2015) Prosaic or profound? The adoption of systems ideas by impact evaluation. *IDS Bulletin*, 46(1): 7-16.

Woolcock M. (2013) Using case studies to explore the external validity of “complex” development interventions. *Evaluation* 19(3): 229-248.