How do policy evaluators understand complexity?

Pete Barbrook-Johnson
University of Surrey, UK

Amy Proctor
Newcastle University, UK

Sara Giorgi
Brook Lyndhurst, UK

Jeremy Phillipson
Newcastle University, UK

Abstract
There is a well-documented interest in how insights from the study of complexity can be applied to policy evaluation. However, important questions remain as to how complexity is understood and used by policy evaluators. We present findings from semi-structured interviews with 30 UK policy evaluators working in food, energy, water and environment policy domains. We explore how they understand, use and approach complexity, and consider the implications for evaluation research and practice. Findings reveal understandings of complexity arising from contextual factors, scale-related issues and perceptions of unpredictability. The evidence indicates terminological and analogical use of complexity and its concepts by policy evaluators, but limited evidence of its literal use. Priorities for the future include framing complexity more pragmatically and as an opportunity not a cost. Communicating this up the policy hierarchy is the key to progressing complexity-appropriate evaluation – this can be enabled by strengthening links between policy evaluation and academic communities.

Keywords
complexity, evaluation, government, policy making, United Kingdom

Corresponding author:
Amy Proctor, Centre for Rural Economy, School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK.
Email: amy.proctor@ncl.ac.uk
Introduction

Over the last 20 years, there has been a growing interest in applying insights from the study of complex adaptive systems to public policy settings (Anzola et al., 2017; Byrne and Callaghan, 2014; Byrne and Uprichard, 2012). This focus complements and builds on earlier and ongoing debates concerning the difficulty in understanding, studying, and solving complex and subjective social and policy issues, or in other words ‘wicked problems’ (e.g. Head, 2008; Rittel and Webber, 1973) that often require ‘interdisciplinary’ (Lowe et al., 2013) ‘systems thinking’ (e.g. Checkland, 1981; Meadows, 2008). In this article, we focus on the use and understanding by evaluators of complexity and its associated conceptual and methodological insights. We use ‘complexity’ as a term to be inclusive and cover the wealth of work across different complexity science research areas (see Castellani, 2018 for an excellent overview of the complexity sciences).

Broadly understood, complexity focuses on systems which have characteristics that make their behaviour hard to understand and predict. Key characteristics associated with complex systems include their adaptive and dynamic nature, feedback loops, multiple scales, thresholds for change, areas of high and low stability, and open or ill-defined boundaries that can span (socio-technical) domains or areas of expertise and responsibility. Such features result in systems characterised by tipping points, non-linearity, emergent new properties, and unpredictability (Centre for the Evaluation of Complexity Across the Nexus (CECAN), 2018). These properties of complex adaptive systems are often evident in the contexts, systems and behaviours that are the focus of a particular policy intervention or suite of interventions, and in the characteristics of the interventions themselves, and thus, policy evaluation.

What insights complexity can bring to the practice of policy evaluation is increasingly of interest to stakeholders both within and outside of government, as a potential way to address the frequent gap between policy development, implementation and review (Gates, 2016; Reynolds et al., 2012). In international development and, to a lesser extent, health and public health policy fields, theory and methods from complexity have been used to show how systems can respond differently to interventions depending on local conditions and individual, institutional and environmental factors (e.g. Atkinson et al., 2018; Prashanth et al., 2014; Ramalingam, 2015; Rutter et al., 2017). However, in other significant policy domains, attention to the prospects and challenges that complexity might pose for evaluation is less well-developed. This is the case for the interdependencies, tensions and trade-offs that exist between food, water and energy security, in the wider context of environmental change (Barbrook-Johnson and Twigger-Ross, 2017; Economic and Social Research Council (ESRC), 2015; Howarth and Monasterolo, 2016; University of Cambridge Institute for Sustainability Leadership (CISL), 2016). Here, those working in government and evaluation are dealing with difficult, sometimes intractable problems, such as the energy trilemma (i.e. the trade-off between energy security, prices and emissions), loss of biodiversity, climate breakdown, and challenges to health and wellbeing, which are entangled with economic, social and environment systems.

While complexity is gaining increasing traction in policy making and evaluation settings (Howarth and Monasterolo, 2016; Ofek, 2016), the multiple meanings, interpretations and potential ambiguity around the concept have the potential to impinge upon its usefulness in practical terms (Gates, 2017; Walton, 2016a). This is problematic for governments and evaluators keen to ensure that their staff ‘have the skills and expertise they need to develop and
implement policy, using up to date tools and techniques and have clear understanding of what works in practice’ (HM Government, 2012: 16). If the use of complexity is to move from the fringes, it is important to reflect on how its concepts, and methods, are beginning to be understood, used and approached by policy evaluators themselves, and consider the implications of this in practice. Such an understanding is so far missing from the literature but can indicate whether the uptake of complexity thinking into the evaluation mainstream is present and, if so, whether it is effective, appropriate and lasting, and signpost areas where complexity can be further integrated and enhance policy evaluation. As the maturing interest in the evaluation community implies, complexity does make a difference in evaluation and its use has the potential to have a significant impact on evaluation practice, so it is vital that emerging understandings of complexity among evaluators are better understood.

Describing and exploring how policy evaluators view and respond to complexity is, therefore, the core purpose of this article. To address this issue, we present findings from qualitative research with UK policy evaluators, focusing mainly on those who are working within government who commission and conduct evaluations. This includes a range of professional groups from policy leads and evidence specialists to analysts and evaluation specialists; our definition of policy evaluator is, thus, relatively encompassing. Our aim is not to ‘judge’ participants’ views against academic definitions and debates, but to enrich our understanding of how these ideas are understood and used (or not) in practice.

The article is structured as follows: Section ‘The use and understanding of complexity in evaluation’ reviews relevant literature on the use and understanding of complexity in evaluation. Section ‘Methodology’ presents our methodological approach involving the use of in-depth qualitative interviews. Section ‘Findings’ presents our substantive findings. This is followed by a discussion in Section ‘Discussion’, before we conclude in Section ‘Conclusion’.

The use and understanding of complexity in evaluation

There are many examples of authors discussing the application of the theory and methods of complexity to public policy (see for example, Ansell and Geyer, 2017; Eppel, 2017). In the evaluation literature specifically, complexity has also received much attention. Since the early 2000s, scholars have considered the theoretical implications of complexity for evaluation (e.g. Sanderson, 2000), explored the contexts in which complexity might be used (e.g. Barnes et al., 2003), including its fit with theory-based approaches (Stame, 2004), and started to consider its empirical implications (e.g. Callaghan, 2008). More recently, reflections and reviews have been produced on the actual application of complexity in evaluation, for example its use in evaluation scholarship (Gates, 2016; Mowles, 2014; Walton, 2014), the uptake of complexity in evaluation practice (Gates, 2017; Reynolds et al., 2016; Walton, 2016a) and performance auditing (Van Der Knaap, 2011), and reflections on which elements of evaluation complexity can be most useful (Williams, 2015).

Other authors have explored, to varying degrees of practical application, how complexity can directly inform and complement existing evaluation approaches (Gates, 2017; Reynolds, 2015). They have considered, for example, how complexity may provide a mandate for going beyond limited or narrow theories of change (Garcia and Zazueta, 2015), how evaluators can develop complexity-consistent theory in relation to realist evaluation (Westhorp, 2012, 2013), how they can develop complexity-appropriate evaluation questions (Larson, 2018), or match approaches to different levels of complexity and non-linearity (Ofek, 2016) or interventions in
real-time (Ling, 2012). Gates (2017) also directly analyses how, in eight specific examples, systems approaches were used by experienced evaluators.

Scholars have also detailed how some specific approaches and methods – notably Randomised Control Trials – can falter in the face of complexity, for example because they do not adapt to emerging and unexpected impacts (Bamberger et al., 2016). There are also many descriptions of the use of specific complexity-appropriate methods or methodological approaches in evaluation, whether it be Qualitative Comparative Analysis (Befani, 2013; Blackman, 2013; Blackman et al., 2013; Byrne, 2013), Process tracing (Schmitt and Beach, 2015), case study approaches (Woolcock, 2013), causal loop diagrams and Systems Dynamics (Dyehouse et al., 2009; Fredericks et al., 2008; Grove, 2015), agent-based modelling (Morell et al., 2010), social network analysis (Drew et al., 2011; Durland and Fredericks, 2005), or co-produced or exploratory approaches to working with stakeholders (Copestake, 2014).

Together this body of practical and conceptual insights into existing approaches and practices is a key step to embedding complexity-appropriate theory and methods into evaluation. The evolution of questions being addressed in this literature – from framing and theorising complexity, to applying complexity and reflecting on application – echoes both the increasing awareness (and potentially, actual use) of complexity in evaluation, but also its distance from being a mainstream approach. However, despite this welter of work looking at the relevance of complexity and complexity-appropriate methods to policy evaluation, Gates (2016) notes that there is need for a deeper understanding of how complexity is actually being understood and used by evaluators. Anzola et al. (2017), for example, classify the use of complexity in one of three ways: terminologically, analogically, or literally. A terminological use of complexity involves reference to the term in purely linguistic terms rather than in conjunction with concepts from complexity science. An analogical use of the term is when defining characteristics of complexity are being employed or implicitly referred to, but where there is still no explicit link to complexity science or its theoretical and methodological foundations. Finally, the literal use of the term(s) ‘implies an explicit awareness of or reference to complexity science’ (Anzola et al., 2017: 227).

Gates (2017) further calls for more empirical research on both how evaluators view and experience complexity to understand the wider implications of this complexity and how systems ‘turn’ to evaluation practice as whole. Walton (2016b) has similarly argued there has been little systematic examination of actual experiences of applying complex systems thinking and approaches to evaluation, with ‘little discussion of the contextual conditions in applying complexity theory’ (p. 73). Gates (2016, 2017) and Walton (2016a) address these issues using, literature review, case studies on published evaluations (with no geographic or policy domain constraints), and semi-structured interviews (with a geographical focus on New Zealand and Australia, and health policy). In this article, we seek to complement and build on these contributions with a focus on evaluators’ understanding and use of complexity at the food-energy-water-environment policy interface.

**Methodology**

We undertook in-depth qualitative interviews with 30 participants for this study. Participants were chosen with the intention of constructing a diverse profile of evaluation settings and roles. This included a range of professional groups operating within the UK government. All were working on some aspect of evaluation from design and commissioning through to
Some interviews included two or three participants (where multiple team members were working on evaluation) and three participants were interviewed on two different occasions as part of a follow-up exercise. In total, we interviewed 11 policy leads, 8 evidence specialists, 5 policy analysts, and 4 specialist evaluators, from 5 UK Government departments and agencies: the Department for Environment, Food and Rural Affairs (Defra), Department for Business, Energy and Industrial Strategy (BEIS), Department for Communities and Local Government (DCLG), Environment Agency (EA), and Food Standards Agency (FSA). From outside government, we also interviewed two evaluation practitioners working with and on behalf of government departments (Table 1).

Participants were identified and recruited through work conducted as part of a larger programme of research for the United Kingdom Centre for the Evaluation of Complexity Across the Nexus (CECAN). Some participants had specifically expressed an interest in working with the centre’s team of academics, practitioners and methods specialists to explore innovative approaches to the evaluation challenges they faced. Other participants were recruited through a snowballing process, nominated by colleagues or others who had already been interviewed. No criteria were given for finding others through the snowball process; this was left to the participants’ judgement. All interviewees were working in food, energy, water and environment policy domains. The interviews focused on evaluations of programmes, policies and interventions of varying scales, from national level to sub-national level and at different stages in the policy evaluation cycle, from those in the early stages of planning to those about to be commissioned and tendered for, to those implemented and completed.

The interviews were conducted in person, over the telephone, or by Skype (audio only) between June 2016 and November 2017, using a semi-structured interview schedule. The schedules were developed from a standard template, and were updated or adapted for specific interviewees, but all included wide-ranging questions exploring the specifics of the policy area and interventions concerned, the key challenges posed in evaluating the policy, the

| Table 1. Interviewee role profiles. |
|-------------------------------------|
| Role                              | Description                                                                 | Participants |
|------------------------------------|-----------------------------------------------------------------------------|--------------|
| Government Policy Lead             | Responsible for the design and development of policy actions to meet government priorities and ministerial objectives | 11           |
| Government Evidence Specialist     | Responsible for building and evaluating the evidence-base from both internal and external sources to inform and direct operational policy and practice. | 8            |
| Government Policy Analyst          | Responsible for using data and analysis to assess likely scenarios from policy implementation, addressing complex policy issues using strategic and analytical skills | 5            |
| Government Specialist Evaluator   | Responsible for the design and conduct of evaluation of policy within Government | 4            |
| Evaluation Practitioner            | Independent evaluation practitioner (working in the private or third sector) who is contracted to conduct an evaluation alongside or on behalf of Government | 2            |

Source: HM Government (2020).
institutional context within which they operated, and the methodological approaches to evaluation being considered or adopted (Table 2).

The interviews specifically asked respondents to reflect on their experience and understandings of complexity in relation to the policy area and its evaluation. The interview schedules were designed to take a neutral view on the utility of complexity in evaluation and to avoid any assumption that there is a ‘correct’ understanding of complexity. In addition, follow-up prompts were used to explore specific characteristics of complexity (e.g. feedback loops, emergence, self-organisation, open systems) and to consider the interviewees’ perception of whether there are differences between complicated or complex systems (though definitions of these were not given).

Interviews were recorded and transcribed verbatim. The analysis was structured according to the key themes identified in the interview schedule. These were used to generate high-level codes with additional data-driven codes emerging from the data following a grounded theory approach (Strauss and Corbin, 1990). These data were analysed manually through rounds of thematic analysis and coding by two members of the research team, working individually on transcripts at first and then collectively to ensure consistency and maintain inter-coder reliability in using the codebook (DeCuir-Gunby et al., 2011). Data

| Theme                                   | Example questions (and follow-up prompts)                                                                 |
|-----------------------------------------|---------------------------------------------------------------------------------------------------------|
| Understanding the policy characteristics | • Description of policy area – aims, objectives, budget, timeframe and geographical scale                 |
|                                         | • Point in policy cycle (issue identified, policy in development, policy appraised, policy implemented and monitored) |
|                                         | • Policy type (experimental, pilot, time limited, ongoing programme, European programme with mandatory reporting) |
|                                         | • Links to other policies                                                                             |
|                                         | • Expected outputs/impacts of policy intervention (short, medium, long term)                           |
|                                         | • Presence of programme logic for how the intervention achieves its objectives                         |
| Complexity in the policy                | • Is complexity seen in the policy? How? (Is/are there interlinked systems; multiple actors; networks; conflicts of interest; complicated governance; social/natural system interaction) |
| Evaluation context                     | • What challenges are envisaged in evaluating the policy? (difficulties in placing boundaries; difficulties in attributing impacts; time constraints; context; emergence; multi-level (upwards and downwards) causation; adaptive; path dependence; non-linearity and tipping points; linked systems (social and natural); other) |
|                                         | • What is the rationale for conducting the evaluation?                                                |
|                                         | • What stage is the evaluation at?                                                                    |
|                                         | • Will the evaluation be conducted in-house or contracted out?                                        |
|                                         | • Geographical scope of the evaluation?                                                               |
|                                         | • Will the evaluation involve inter-agency working/stakeholders?                                       |
|                                         | • If underway, has the evaluation had impact already?                                                  |
| Evaluation methods                      | • What methods are being applied or considered in evaluating the policy? (Why are you interested in that method? What do you feel it will allow you to do?) |
|                                         | • What is the potential for trialling/experimentation of evaluation methods?                         |
were analysed for each respondent, thematically across different professional roles and across the entire sample.

**Findings**

In this section, we first consider the ways in which respondents understand complexity, and then explore how evaluators are actually approaching evaluation in response, or not, to these understandings. Throughout, we avoid comparing participants’ views against academic definitions and debates, as our aim is to gain a richer understanding of their understanding and use (or not) of complexity.

**How do evaluators understand complexity?**

Participants were asked if and how complexity was seen in the policy or intervention they were trying to evaluate. Their understandings and implicit definitions of complexity encompassed a broad set of meanings, however, three key themes emerged from the data, namely complexity arising from context, from scale and from unpredictability. We examine each in turn before considering one further issue raised by some participants concerning debates around defining issues as either ‘complicated’ or ‘complex’. We conclude this section with a consideration of understandings of complexity across different professional roles.

**Complexity and context.** Many interviewees pointed to uncertainty in the policy context as a key source of complexity, which also had implications for creating uncertainty and complexity for the evaluation process itself. Brexit was a recurring example given; the UK referendum on membership of the European Union in 2016 and its outcome was a backdrop to many of the interviews. Consequently, evaluators recognised an additional layer of complexity arising from the uncertainty of how policies may evolve and be evaluated post-Brexit.

Complexity was also identified as arising from, for example, how particular policies might have links to, or be influenced by, wider public interest or debates in the media, and the wider supply chain for an industry or technology. These contexts, whether they be social, political, or geographical, introduce complexity, for example,

I guess the main complexity is the relationship with the supply chain. There’s no bounds on the supply chain that we’re influencing. *(Government Analyst)*

There was a [social media post] from [government department/agency] . . . just reflecting [on an NGO campaign], I wouldn’t say it was campaigning on behalf of it or promoting it but . . . that created some disquiet in the [industry] lobby . . . So, it’s a complex, international, highly valuable sector, politically sensitive initiative. *(Government Policy Lead)*

These quotes provide examples of how complexity can be perceived to arise from a wide range of contextual issues, which may not be immediately or obviously connected to the policy intervention itself.

**Complexity and scale.** Issues of scale presented particular challenges for evaluators. Participants discussed the size or breadth of the programmes or interventions being evaluated, that
often-included sub-schemes and nested policies, which presented problems of attribution and causality. As one policy lead noted,

\[\ldots\] at a programme level, [approaching complexity] might be easier because you can start to think about the attribution of impact but when you start to think of it at a scheme level [i.e. at larger scale], there’s quite a lot of complexity in those relationships and who is delivering what, what impacts are a result of what. (Government Policy Lead)

National or regional policies can be difficult to evaluate since local context often plays a crucial part in determining outcomes. Complexity was also said to arise in dealing with the scaling-up and multiplier effects of interventions. For example,

Where it gets complex is the jump which is what this [policy phase] is trying to do, it’s not really just doing [a small number of] projects . . . What it’s trying to do is doing that straightforward bit and then through demonstrating that that singular project is being viable, profitable, replicable, get it repeated at scale by others without the government money. And that’s where there’s this kind of leap and it almost feels like a leap of faith at the moment . . . So that’s where it’s very complex because there’s not an obvious channel by which those effects are going to be multiplied up. (Evaluation Practitioner)

Participants also viewed complexity as something which can be brought in or out of focus depending on the temporal and spatial scales at which an evaluation is being framed. Immediate outcomes at the local level may be different from the scaled-up longer term impacts and this can introduce an analytical tension within evaluation, as one participant describes,

This is where I think there are almost two levels to this evaluation which are quite distinct in my mind . . . that immediate effects thing, money in, technology employed, outputs. I think it is quite linear. It’s after that when you’re looking to jump up to the broader scale impacts. It’s a little bit of a challenge . . . a kind of analytical struggle . . . are we supposed to focus just on this linear path . . . or are we trying to put in the building blocks and baseline now for [broader impacts] happening . . . it’s like kind of hard to work out. (Evaluation Practitioner)

**Complexity and unpredictability.** Another common aspect of complexity respondents encountered was the unbounded or unpredictable nature of the natural environment. This was particular to the cases we were considering, all broadly located within the food-energy-water-environment nexus. This quote for example, is illustrative of a wider view across the sample:

what does make it complex is the very slow and slightly unpredictable nature of the environmental system itself. (Government Analyst)

Complexity was associated with the unpredictability of interacting variables or factors. Some respondents referred to the presence of multiple variables requiring evaluation and the problems of complex attribution and causality. This challenge related to separating out and understanding the unique contribution of interventions developed in combination or sequence where there might be resultant synergies and complementarities. One participant, reflecting on approaching complexity in their evaluation process noted,
There’s a whole range of variables that you need to consider . . . different sectors, different environmental circumstances, different things that you’re trying to prevent or do . . . So, there’s a lot to take into account and it all needs to be considered. (Government Evaluator)

Moreover, in describing the systems they were trying to evaluate, participants referred to the interactions between social and natural systems and how these intersected with broader political and economic factors, for example,

When you are looking at something that is necessarily that complex and covers most aspects of the economy, the environment and lots of social issues, and you are trying to measure the impact of the plans against social, economic and environmental factors then you get something that is very complex. (Government Evidence Specialist)

Complex or complicated? Finally, distinctions between complicated and complex systems, interventions and evaluations were also apparent in our findings. This reflects the ongoing debates among complexity and systems scholars about the nature of and differences between simple, complicated and complex systems (see Byrne and Callaghan, 2014 and Ramalingam, 2015 for some summary of these). Some scholars have questioned the value of a distinction between complicated and complex and, in particular, the concept of ‘complicated’ has been challenged. When we look at the findings, it becomes apparent that these debates do not inform the evaluators discourse around complexity. One participant for example, when asked whether a policy area is complex or complicated, introduced a distinction between the policy area as complex, and the evaluation as complicated:

Yes, in the sense that I think it is very very hard to, because it is so varied . . . it’s complex in that regard. I suppose it is a very simple concept that you [encourage people to perform a certain behaviour through the policy]. On that front it is very very simple. But, actually trying to identify what works is incredibly complicated . . . I suppose the [wider policy] picture is complex, and the evaluation to identify what works is quite complicated. (Government Policy Lead)

Another participant makes the distinction in the reverse. When asked if different aspects of complexity apply, they suggest the evaluation is complex, whereas the policy is not:

I think so yes. Does that make it complex though!? It is amazing when you think about it on that level, some of these things you tune into and others you don’t, but when you read them all out you think they all did apply . . . I can see [it is] complex from an evaluation perspective, but actually the concept from a policy perspective probably isn’t complex. Interesting. (Government Policy Lead)

We interpret these conversations to demonstrate the varied ways in which complexity can be viewed. Indeed, one participant took a much broader view of complexity, linked to their concerns about the burden that a policy might place on those it impacts (i.e. how much bureaucracy or cost it imposes on people or organisations):

The sort of policies that I generally am aware of and deal with relate to cost systems or incentive systems or things like that, where complexity is an integral part so we naturally would consider that because we see complexity as a burden on people . . . The guidance that deals with this is the Green
Book guidance and there isn’t a section in there that talks about complexity, they will talk about burdens and things like that . . . So I am talking about complexity in the sense of how many steps do you need to take in order to comply, or something, and if you are making it overly burdensome and complex, I would talk about that in terms of the complexity of the programme you are setting up. But I can see how other people wouldn’t necessarily think that is classic example of complexity. *(Government Policy Lead)*

This quote reflects wider concerns about the ‘passing on’ of complexity ‘down’ a policy hierarchy. This dynamic of pushing complexity down to people and organisations may be an important entry point, in the wider context of aspirations for better regulation and less onerous policy appraisal and implementation, for those wishing to convince policy makers and evaluators of the need to take a complexity-appropriate approach to policy design and evaluation.

*Understandings of complexity across different professional roles.* There was no qualitative difference in understandings of complexity across the different roles and contexts of the respondents outlined in Table 1. While we might have expected patterns to emerge across different categories, for example, those in defined evaluation roles (evaluation specialists) to have a more sophisticated understanding of complexity, this was not the case. Insights were not defined by role but instead were found to be cross-cutting and systemic. One explanation for this is that many of the participants had a declared interest in complex evaluation prior to interview and this emerges strongly in the breadth of understanding revealed. Indeed, participants recognised that the level of understanding they possessed was not necessarily shared by colleagues they worked closely with:

Part of the challenge is making sure that everybody has a shared understanding of evaluation, which is itself a fairly complex thing . . . especially with people coming from very different backgrounds. *(Government Analyst)*

The people who are doing it (evaluation within the organisation) can be very different. *(Government Evaluator)*

Many of the participants were, therefore, performing an important role as brokers within Government for increased understanding of complex policy evaluation.

*How are evaluators responding to complexity?*

In this section, we consider the ways in which evaluators are approaching evaluation in response to their understandings of complexity outlined in Section ‘How do evaluators understand complexity?’ We examine how evaluators might be embracing or working with complexity, or choosing to give it less attention.

The interviewees revealed a range of reactions to complexity in their respective policy domains. Some felt motivated and reacted pro-actively to it, others did not. As apparent as some participants’ willingness to name and identify complexity, was others’ readiness to deny or negate its importance. Some participants felt that despite a complex context or policy setting, a simple policy could be effective, or a standard evaluation would be sufficient. Others accepted that components or parts of complexity (such as feedback loops or multiple scales)
were present in their areas of work, but that these did not ‘sum up’ to overall complexity. These views have in common an acceptance of complexity, but less so in terms of implications for the approaches and methods used in policy evaluation.

Participants were asked whether policies and their evaluation should be assessed on their complexity to help inform design, implementation and evaluation approaches. One participant felt this was unnecessary:

That doesn’t happen. When we’re sitting down to think about how we can realise this outcome, like you know there’s awareness and discussions about things, there’s certainly an awareness that things are complicated and complex and interdependent and things like you’re pulling in a different direction and what have you. . . . [but] I’m not sure there’s a massive need for it [assessing complexity and using it in decisions] because as I say the complexity is just drawn out in the way things are tackled and talked about. (Government Analyst)

In this example, existing approaches and activities are deemed sufficient in accounting for complexity. Similarly, another participant felt existing processes were enough and that complexity could be dealt with informally:

I don’t think we do that massively . . . it might be done informally, something might be recognised as difficult or complicated or complex, where other things might be more straightforward . . . I think it is useful to understand what . . . are the challenges you would have to overcome. (Government Policy Lead)

This participant suggested this informal recognition of complexity would be done as a policy is assessed on its costs, burden and stakeholder support needs.

In contrast, other interviewees felt complexity rendered some traditional or typically used approaches and methods unusable, for example,

It would be very useful for evaluation, wouldn’t it, if you diagnose that your policy intervention is genuinely complex you may as well give up evaluating it in any kind of traditional way and relying on the logic map type approach and looking at change of causality and doing that kind of evaluation. So, from an evaluator’s point of view yes I think so. Also I guess from the policy makers’ point of view it becomes much more risk, there is a much higher chance that it won’t work for some kind of random complex reason that you can’t actually predict . . . Although what you might find is that everything, because everything happens, in our world everything is done in a social context and that is what creates the complexity because people aren’t all the same and households aren’t all the same, that you might end up diagnosing everything as complex which wouldn’t be particularly helpful. (Government Evaluator)

Here, the participant sees flaws in some traditional evaluation approaches, but also a diminishing value to asserting that everything is complex.

Some participants went on to suggest that there were differences in the implications of complexity depending on stages of the policy cycle and placed particular importance on the responsibilities of policy designers to address complexity. Others alluded to skills and capacity needs if complexity is to be effectively taken on board:

Not much, but it should! . . . We recognise that actually our ability to evaluate these sorts of things is not particularly great, and I think that we are hoping for insights into how to do it better. (Government Policy Lead)
In one case, where an evaluation team was small and recently established, a participant noted that because the team was relatively young within the organisation, they were on a ‘learning journey’ and only just beginning to broach the idea of complexity in evaluation. Relatedly, an analyst working on evaluation within a large government department noted,

We don’t have a huge amount of internal capacity for evaluation. We have very few people who are expert in evaluation or even really versed in evaluation. (Government Evidence Specialist)

Here, it is a capacity or skills constraint which stops the implications of complexity being addressed in an evaluation. When faced with conducting evaluation in complex settings, some evaluators were able to seek help from outside their immediate teams, for example, by drawing on department-wide, or government-wide evaluation networks which acted as a sounding board when designing and commissioning evaluations. This helped provide guidance and challenge from more experienced evaluation personnel from across the wider civil service. Others were optimistic about complexity being seen as an opportunity if framed in the right way, rather than simply implying greater costs or difficulty:

I think it [complexity] would be an interesting lens, whether it would be useful or not would need to be tested out under does introducing that type of language add something useful to an impact assessment? I think if you start to talk about ‘It’s all very difficult, it’s all very complex’ it suddenly becomes actually it sounds quite negative and it sounds as if we are trying to say ‘We can’t really do this it is all too difficult’ and that is not a message at all. So framing that issue of complexity in where are the opportunities to do this and what benefits does it bring to the policy delivery in understanding that complexity and just the very sheer understanding that this is very very complex and there are lots of different strands which inter-relate and turn into a plate of spaghetti, does us knowing that actually help, does it make a difference? (Government Analyst)

Lack of institutional structures, capacity and expertise can, therefore, leave evaluators struggling to find the tools, methods, or frameworks to evaluate in complex settings. Equally, in institutions that have well-developed evaluation structures and capacity, evaluators can find themselves restricted by these, and unable to use new or untested complexity-appropriate methods.

Discussion

We have explored the broad set of understandings of complexity revealed in our interviews and seen how these are often framed round context, scales and unpredictability. While these features could be considered common features of any policy environment (and indeed routine characteristics of complexity thinking), they are significant in the context of this research as they are core to how policy evaluators themselves identify and articulate complexity. The main perceived sources of complexity in their policy and evaluation settings related to context such as Brexit, political sensitivities and engagement in wider industries, markets and supply chains. Issues around temporal and spatial scales, and the size of programmes and policies (i.e. policies being portfolios of interventions) were another key source of complexity faced and described. Particular aspects of complexity that were most salient for participants and used to articulate complexity, included the challenge of evaluating multiple variables and attribution, the interaction of social and natural systems, and difficulties around defining boundaries. These data also
reveal contrasting definitions and understandings among participants of complicated and complex systems in relation to evaluation. Moreover, the analysis shows contrasting responses to complexity among policy evaluators with some dismissing or negating the need to address it in evaluation, and others describing how it can drive improvements in practice if skills and capacity needs can be addressed and/or if it can be integrated in existing systems.

Returning to Anzola et al.’s framework which refers to the terminological, analogical or literal use of complexity, these data also highlight how the UK policy evaluators are understanding and using the language of complexity. Each of these types of use has potential implications for whether they might drive or hinder potential changes in individuals’ or institutions’ evaluation practice towards being more complexity-appropriate. Where terms are used terminologically, there is strong potential for this to impede potential change; the chance of misunderstandings, cross-talk and lack of examination of assumptions are high as definitions and their implications are loosely used or not discussed. Discussion and collaborations between commissioners, evaluators and methodologists are likely to be frustrating for all. Where use is analogical, there is the potential for this to either impede or promote changes in practice; analogical use could be the first stepping stone towards taking action in the face of complexity, or it could again lead to misunderstandings around ill-defined terms. Finally, using complexity terms in a literal way is likely to best show change in practice already occurring, or soon to happen.

From our participants, we observed terminological and analogical use of complexity and its concepts regularly, but few examples of wholly literal use. This finding suggests one of the key barriers to the application of complexity in evaluation – multiple definitions of complexity (Walton, 2016a) – is present among the UK policy evaluators we interviewed. Considering our sampling strategy, and the likelihood that we interviewed evaluators already relatively interested in complexity, it does not seem unreasonable to assume the wider evaluation community is equally, if not more so, undecided or unclear on understandings of complexity. While it would be unfair to expect evaluators to be well-versed in complexity when they are not widely trained in it, we had expected to find more evidence of its literal use given the interest in complexity in the evaluation field.

Reflecting on the extent to which the evaluation community is embracing complexity, Gates (2017) outlines a range of transformations in evaluation practice that would signify a shift towards a more complexity-appropriate approach. These include, re-structuring evaluation itself as an intervention, re-defining the object of evaluations, re-figuring relationships with evaluation commissioners, re-purposing and expanding methods, revisiting the making of value judgements and renewing the emphasis on the instrumental use of evaluations. Similarly, Larson (2018) describes a set of evaluation questions which evaluations might use to refocus them in a more complexity-appropriate way. These include asking if a policy was grounded in history and current priorities, informed by beneficiaries and their dynamic relationship with implementers, effective in affecting those relationships, responsive to external shocks, accommodating of diversity of actors, iteratively monitoring, reviewing and updating, aware of self-organising and emergent behaviours, and considering what happens when it finished (Larson, 2018).

Our empirical analysis revealed little evidence of such transformations or the routine use of similar evaluation questions (though appreciating that interviews were conducted in 2017 so progress may have been made subsequently). This suggests a need for continued emphasis and energy on putting in place the more foundational aspects described by Gates (2017) and
Larson (2018) and in locating a focus on new methodological tools within these wider considerations of evaluation practice (i.e. what is evaluated, how evaluations are designed and commissioned, and how evaluations are used). This finding could be viewed as a criticism of the evaluation community or evaluation commissioners for not taking enough action in the face of the growing awareness of the importance of complexity. The cause for this emphasis may equally be found in the tendency of methodologists and applied complexity scientists to offer methodological innovation without engaging in the wider theory and foundational practice of evaluation.

Finally, the findings point towards variable institutional capacity for, and experience in, evaluation across different government departments and agencies, which is having a bearing on the level of engagement with complexity. Different institutional histories and cultures mean that some are only just beginning to broach complexity while other departments have a stronger understanding and awareness of what it means and are more able to respond to complexity than others. It was clear that some of the participants interviewed were in the early stages of grappling with what complexity meant to them in terms of the challenges it poses to evaluation and the potential for thinking about new approaches to dealing with it. We found in some cases that departments or agencies with formalised and well-established evaluation practice can struggle to adapt. Those starting anew may have less ‘lock-in’ to previous ways of working, and in some cases may be more willing or able to innovate in response to complexity but may lack capacities and resources to enact change. Finally, given that insights on complexity were found not to be defined by professional role, our findings suggest the importance of a cross-profession response to any future learning and training in this area.

Conclusion

Considering how complexity is understood and acted upon by policy evaluators is invaluable in assessing progress towards complexity-appropriate evaluation. The evidence indicates terminological and analogical use of complexity and its concepts by policy evaluators, but limited evidence of its literal use (Anzola et al., 2017). The article has also considered how understandings of complexity shape its ultimate use and value in evaluation. Indeed, the findings show that one of the key barriers to the application of complexity in evaluation found by Walton (2016a) – multiple definitions of complexity – is a key factor among the policy evaluators interviewed. Understanding what issues may be undermining the use of complexity, and reflecting on how further improvements may emerge is critical.

Our findings offer several key pointers for future directions. First, complexity must be framed as an opportunity, not as a cost. Its value proposition to evaluators and policy makers must be more clearly made and articulated. Second, it should be framed more pragmatically. The common focus on ‘sophisticated’ techniques and methods needs to be complemented by pragmatic consideration of what is possible and what is not, pros and cons, what others are doing, costs and risks. These descriptions are invaluable to policy makers and evaluation commissioners; without them, their hands are tied. Third, the question of how to communicate complexity ‘up’ the hierarchies of public policy and government, and ultimately to the public, must be considered. Without the concepts, skills and tools to communicate complexity evaluators’ efforts to apply complexity will be frustrated. The responsibility for these framing issues lies with evaluators, methodological experts and applied complexity scientists, who should work together to make these more practical framings happen. Finally, it should be a priority
for evaluation researchers to develop practical guidance on how foundational practice can be changed in the face of complexity and how to apply complexity-appropriate methods in circumstances that may be resource and capacity constrained. This is no small task and will require the input of experts from across all of the evaluation and policy making community. Strengthening the relationship between policy, practitioner and academic communities through improved engagement and knowledge exchange is essential.

Acknowledgements
The authors would like to thank the CECAN team for their intellectual support, for comments and feedback on early drafts of this paper, and for helping to inform the wider research programme which allowed them to conduct the interviews in this study. They also thank Francis Naab at Newcastle University for his assistance with the French translation.

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 Economic and Social Research Council (grant numbers ES/N012550/1 and ES/S007024/1).

ORCID iDs
Pete Barbrook-Johnson https://orcid.org/0000-0002-7757-9132
Amy Proctor https://orcid.org/0000-0003-2308-7930

References
Ansell C and Geyer R (2017) ‘Pragmatic complexity’ a new foundation for moving beyond ‘evidence-based policy making’? Policy Studies 38(2): 149–76.
Anzola D, Barbrook-Johnson P and Cano JI (2017) Self-organization and social science. Computational and Mathematical Organization Theory 23(2): 221–57.
Atkinson J, Page A, Prodan A, et al. (2018) Systems modelling tools to support policy and planning. The Lancet 391(10126): 1158–59.
Bamberger M, Tarsilla M and Hesse-Biber S (2016) Why so many ‘rigorous’ evaluations fail to identify unintended consequences of development programs: How mixed methods can contribute. Evaluation and Program Planning 55: 155–62.
Barbrook-Johnson P and Twigger-Ross C (2017) CECAN’s progress and insights on evaluating complex Nexus policies. Presentation made at 6th European Environmental Evaluators Network Forum. Available at: https://www.eea.europa.eu/themes/policy/events/eeen2017/eeen-2017-presentations
Befani B (2013) Between complexity and generalization: Addressing evaluation challenges with QCA. Evaluation 19(3): 269–83.
Blackman T (2013) Rethinking policy-related research: Charting a path using qualitative comparative analysis and complexity theory. Contemporary Social Science 8(3): 333–45.
Blackman T, Wistow J and Byrne D (2013) Using qualitative comparative analysis to understand complex policy problems. Evaluation 19(2): 126–40.
Byrne D (2013) Evaluating complex social interventions in a complex world. Evaluation 19(3): 217–28.
Byrne D and Callaghan G (2014) Complexity Theory and the Social Sciences: The State of the Art. London: Routledge.
Byrne D and Uprichard E (2012) Useful complex causality. In: Kinkard H (ed.) Oxford Handbook of Philosophy of Social Science. Oxford: Oxford University Press, pp. 109–29.
Callaghan G (2008) Evaluation and negotiated order: Developing the application of complexity theory. Evaluation 14(4): 399–411.

Castellani B (2018) Map of the complexity sciences. Art and Science Factory. Available at: https://www.art-sciencefactory.com/complexity-map_feb09.html

Centre for the Evaluation of Complexity Across the Nexus (CECAN) (2018) Policy evaluation for a complex world. April 2018, Version 2.0. Available at: www.cecan.ac.uk

Checkland P (1981) Systems Thinking. Systems Practice. Chichester: John Wiley; Ann Arbor, MI: University of Michigan.

Copestake J (2014) Credible impact evaluation in complex contexts: Confirmatory and exploratory approaches. Evaluation 20(4): 412–27.

DeCuir-Gunby J, Marshall P and McCulloch A (2011) Developing and using a codebook for the analysis of interview data: An example from a professional development research project. Field Methods 23(2): 136–55.

Drew R, Aggleton P, Chalmers H, et al. (2011) Using social network analysis to evaluate a complex policy network. Evaluation 17(4): 383–94.

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): 187–96.

Economic and Social Research Council (ESRC) (2015) Centre for evaluating complexity across the energy-environment-food nexus Call specification – Full stage. Available at: https://esrc.ukri.org/files/funding-opportunitiescentre-for-evaluating-complexity/ceccall-specification-full-stage/

Eppel E (2017) Complexity thinking in public administration’s theories-in-use. Public Management Review 19(6): 845–61.

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): 251–67.

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): 152–71.

Grove JT (2015) Aiming for utility in ‘systems-based evaluation’: A research-based framework for practitioners. IDS Bulletin 46(1): 58–70.

Head B (2008) Wicked problems in public policy. Public Policy 3(2): 101–118.

HM Government (2012) The Civil Service Reform Plan. London: HM Government.

HM Government (2020) Government Analysis Career framework. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/877248/AF_CF_March20.pdf

Howarth C and Monasterolo I (2016) Understanding barriers to decision making in the UK energy-food-water nexus: The added value of interdisciplinary approaches. Environmental Science & Policy 61: 53–60.

Larson A (2018) Evaluation amidst complexity: Eight evaluation questions to explain how complex adaptive systems affect program impact. Evaluation 24(3): 353–62.

Ling T (2012) Evaluating complex and unfolding interventions in real time. Evaluation 18(1): 79–91.

Lowe P, Phillipson J and Wilkinson K (2013) Why social scientists should engage with natural scientists. Contemporary Social Science 8(3): 207–22.

Meadows DH (2008) Thinking in systems: A primer. In: Wright D (ed.). White River Junction, VT: Chelsea Green Publishing.
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): 160–75.

Ofek Y (2016) Matching evaluation approaches to levels of complexity. *Evaluation Review* 40(1): 61–84.

Prashanth NS, Marchal B, Devadasan N, et al. (2014) Advancing the application of systems thinking in health: A realist evaluation of a capacity building programme for district managers in Tumkur, India. *Health Research Policy and Systems* 12: 42.

Ramalingam B (2015) *Aid on the Edge of Chaos*. Oxford: Oxford University Press.

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

Reynolds M, Forss K, Hummelbrunner R, et al. (2012) Complexity, systems thinking and evaluation. *Evaluation Connections: Newsletter of the European Evaluation Society*, (December), 7–9.

Reynolds M, Gates EF, Hummelbrunner R, et al. (2016) Towards systemic evaluation. *Systems Research and Behavioral Science* 33(5): 662–73.

Rittel HWJ and Webber MM (1973) Dilemmas in a general theory of planning. *Policy Sciences* 4(2): 155–69.

Rutter H, Savona N, Glonti K, et al. (2017) The need for a complex systems model of evidence for public health. *The Lancet* 390(10112): 2602–4.

Sanderson I (2000) Evaluation in complex policy systems. *Evaluation* 6(4): 433–54.

Schmitt J and Beach D (2015) The contribution of process tracing to theory-based evaluations of complex aid instruments. *Evaluation* 21(4): 429–47.

Stame N (2004) Theory-based evaluation and types of complexity. *Evaluation* 10(1): 58–76.

Strauss A and Corbin J (1990) *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. Newbury Park, CA: SAGE.

University of Cambridge Institute for Sustainability Leadership (CISL) (2016) *Nexus2020: The Most Important Research Questions for Business Sustainability*. Cambridge: Cambridge Institute for Sustainability Leadership. Available at: https://www.cisl.cam.ac.uk/resources/publication-pdfs/Nexus2020

Van Der Knaap P (2011) Sense and complexity: Initiatives in responsive performance audits. *Evaluation* 17(4): 351–63.

Walton M (2014) Applying complexity theory: A review to inform evaluation design. *Evaluation and Program Planning* 45: 119–26.

Walton M (2016a) Expert views on applying complexity theory in evaluation: Opportunities and barriers. *Evaluation* 22(4): 410–23.

Walton M (2016b) 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.

Westhorp G (2012) Using complexity-consistent theory for evaluating complex systems. *Evaluation* 18(4): 405–20.

Westhorp G (2013) Developing complexity-consistent theory in a realist investigation. *Evaluation* 19(4): 364–82.

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): 229–48.

**Pete Barbrook-Johnson** is a Senior Research Fellow in the Department of Sociology at the University of Surrey and a member of the Centre for the Evaluation of Complexity Across the Nexus (CECAN).
Amy Proctor is a Research Associate at the Centre for Rural Economy at Newcastle University and a member of the Centre for the Evaluation of Complexity Across the Nexus (CECAN).

Sara Giorgi is an Associate at Brook Lyndhurst, a sustainability and behaviours research company. She is also an independent researcher and evaluation practitioner specialising in the complex field of resources and waste, namely, attitudes, behaviours, services and policy.

Jeremy Phillipson is a Professor of Rural Development at the Centre for Rural Economy at Newcastle University with research interests on the development needs of rural economies and fishing communities. He is Director of the National Innovation Centre for Rural Enterprise.