



Article

Towards systemic evaluation in turbulent times – Second-order practice shift

Evaluation

2020, Vol. 26(2) 205–226

© The Author(s) 2020

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/1356389020909120

journals.sagepub.com/home/evi



Barbara Schmidt-Abbey  Martin Reynolds
and Ray Ison

Open University, UK

Abstract

The need for, and possibilities of, a second-order shift in evaluation practice are explored. Second-order evaluation practice enables an evaluator to improve practice as a skilled practitioner, acknowledging her embeddedness within an evaluand. The article explores evaluation practice as experienced by professional evaluators, using ideas from developmental evaluation coupled with systemic evaluation in the tradition of systems thinking in practice. Systemic evaluation aims to capture systemic sensibilities – the bigger picture – of complex turbulent situations of change underpinning evaluands. Attributes of second-order practice with systemic evaluation are understood as being aligned with both systemic and systematic modes of evaluation praxis. Personal experiences are provided where this juxtaposing praxis has been found wanting. By example, a systems thinking in practice framework is explored as heuristic support for making a second-order practice shift. The article concludes with a discussion of some implications for developments in professionalising evaluation practice and research.

Keywords

developmental evaluation, evaluation praxis, evaluation professionalisation, isophor, second-order practice, systemic evaluation, systems thinking in practice

Introduction

Interest in systems thinking in evaluation has come and gone in cyclical waves over some decades. As noted in a 2011 Editorial in this journal, ‘Systems thinking in the social sciences waxes and wanes and now appears to be once more in the ascendancy’ (Stern, 2011: 324).

Corresponding author:

Barbara Schmidt-Abbey, Applied Systems Thinking in Practice Group, The Open University, PO Box 197, Milton Keynes MK7 6BJ, UK.

Email: Barbara.Schmidt-Abbey@open.ac.uk

Interest is sustained through conferences like the European Evaluation Society (EES, n.d.), where topics devoted to systems approaches and complexity in evaluation have featured prominently over the last decade. The American sister organisation (American Evaluation Association (AEA) (n.d.) even has an active ‘Topical Interest Group’ on Systems in Evaluation.

In theory, if not always in practice, the evaluation community is aware of the many challenges encountered by evaluation practice within turbulent, complex situations of change and uncertainty. A call for more systems thinking has been a common response to these experiences along with an increasing sense of urgency to step-up the changes required in practices to more effectively address situational turbulence (Caffrey and Munro, 2017; Catwell and Sheikh, 2009; Piirainen et al., 2012; Reynolds et al., 2016). For example, the EES’ biennial conference 2020 (EES, 2020) is devoted to ‘evaluation in an uncertain world’ and the importance of complexity, legitimacy and ethics. The previous conference (EES, 2018) was dedicated to the search for evaluation for more resilient societies.

This article is grounded in evaluator practitioner experiences where there is a gap between espoused systems thinking, or systems-thinking-in-theory and systems-thinking-in-practice (STiT cf. STiP). The argument is made that a ‘second-order shift’ in practice by evaluation practitioners is needed to foster the emergence of *systemic evaluation*. In addition to practitioner experience, this proposition draws on ideas from Developmental Evaluation (DE) and the expanding tradition of STiP (Ison, 2017). We explore the central role the evaluation practitioner ought to play at the intersection of evaluation practice, the context and situation where the evaluation practice is enacted, and the choice and use of appropriate theory, methods and frameworks for systemic evaluation.

A systemic conception of practice and the practitioner is first introduced, followed by the role second-order practice can play in evaluation. We then introduce systemic evaluation realised through STiP, as a primary means of enacting second-order practice. How evaluators can act as reflexive, systems thinking practitioners is then explored through considering their own Being, Engaging, Contextualising and Managing (BECM). Following an exploration of the implications of our arguments for professional evaluation practice, we conclude by raising questions critical to the further development of second-order practice in evaluation. Key to such a shift, we argue, is the use of a reframing within systemic evaluation practice leading to a focus on ‘small r’ research practice associated with the design and enactment of learning systems.

Second-order practice and practitioners

Evaluation literature and practice, including evaluation professionalisation discourse, share a strong focus on theoretical frameworks, methods and tools. An element that is often neglected is the evaluation practitioner, that is, knowingly or not the practitioner is abstracted out of practice. Frequently, the evaluation practitioner is literally ‘not in the picture’ in methodological discussions about evaluation. Much of the attention in the literature and professional discourse is given to methods, which concern the ‘how’ of evaluation implementation. In these discussions, evaluation guidelines, standards and ‘best’ practices are typically geared towards conducting ‘robust’ evaluations.

For example, the evidence-based policymaking approach favours certain methods over others. Randomised control trials (RCTs) and other experimental approaches are often hailed as a ‘gold standard’. While there undoubtedly are cases to be made where RCTs and experimentation can be used to good effect, the promotion of ‘gold standards’ and proliferation of ‘standards of

evidence' is the antithesis to complexity-sensitive or systemic evaluation. The evaluation and evidence-based policymaking communities have extensively debated these contentions in the recent decade: for example, Cairney (2016), Duffy (2017), French (2018) but the contestations are far from settled.¹

Frequently, the impression is created that the desired robust evidence can be produced in an objectified manner, with an assumption of reproducibility, as if dealing with a clinical experiment. Rarely is the practitioner herself given much attention in these methodology-focused discussions.

A shift from first-order (exploring the world) to second-order (reflecting about the exploration) research centrally involves the role of the practitioner engaged in such explorations. In the dominant first-order research tradition, the researcher is an independent objective observer who is outside the situation of concern, which is treated as the object of research. The individual doing the observation is not of concern and assumed to be an objective and dispassionate (and replaceable) observer and investigator. In the evaluation field, this separation of the observer from the situation of interest is even further accentuated by the emphasis put on evaluator independence and impartiality, which implies being external to the evaluand.

By contrast, the practitioner in a second-order research approach reflects upon their explorations as being integral to the situation of concern, of which he or she is part. Observer inclusion is a hallmark of second-order research approaches, such as second-order cybernetics in the traditions of Heinz Von Förster (1984, 1992) and Humberto Maturana (Maturana and Poerksen, 2004).

Drawing on the traditions of systems thinking and second-order cybernetics, we propose that a second-order shift towards systemic evaluation requires more attention to be paid to what it is that an evaluation practitioner (evaluator or evaluation commissioner) actually does, when she 'does what she does' (Ison, 2017: 5, paraphrasing Maturana).

Praxeology is the study of human action, relating to engaging in purposeful and willed behaviour. Practice is often referred to in relation to a professional practice: it is what professionals 'do when they do what they do' – this defines a practitioner's practice (based on Ison, 2017: 14). There is also praxis, which concerns a theory or philosophy becoming a practical social action.

What do evaluation practitioners actually do as their practice? The default assumption of what may be meant by an evaluation practitioner is being an evaluator, conducting evaluations. But it can involve any other practice role in relation to evaluation, such as evaluation commissioner, evaluation user or any other stakeholder involved and affected by an evaluation, or evaluation researcher. As Wadsworth (1997) noted, we all do evaluation every day.

The evaluation tradition of DE invokes the idea of the evaluator as being embedded in the evaluand (Patton, 2011, 2018). Our article shares this view, with the difference that (1) we explicitly call this second-order practice and (2) make it relevant to all evaluation practices in all evaluands – not contingent on specific circumstances of the evaluand or situation.

One of the hallmarks of becoming a skilled practitioner (in any field) is being a reflective practitioner. The key concepts introduced by Donald Schön in the 1980s (Schön, 1984, 1987) are reflection-on-action and reflection-in-action. The former is *ex post*, while the latter is enacted and thus embodied in the unfolding doing of practice. The reflective practitioner concepts are integral to second-order practice; they imply that the practitioner reflects-on and -in practice about their practice, including their language, assumptions, values, repertoires, theo-

ries and emotions. Ison (2017) explains reflexivity as a reflection-on-reflection, a second-order practice that encompasses both of Schön's distinctions.

What is meant by a 'second-order shift'?

Grounded in empirical work with pastoralists in semi-arid Australia, Ison (2017: 278–282) makes a distinction between 'first-order' and 'second-order' research traditions (see Russell and Ison, 2000). The 'first-order research tradition' continues to dominate how science and research – and evaluation – are practised. A first-order research tradition – applied to the field of evaluation – is characterised by the dominance of well-established social science research approaches based on notions of linear causation, a systematic linear chain of causal factors. The purpose of an evaluation in first-order understanding is to systematically observe and record such chains with appropriate levels of dispassionate objectivity. Ideas of linear causation are typically expressed for instance in traditional log frame models, by targets, goals and the understanding of objects of study or evaluation as fixed entities that can be studied and measured objectively. An expression of first-order evaluation practice might be associated with the founding traditions of evaluation as a discipline through the works of Scriven (1991, 1996, 2001, 2003), features of which have been described by Patton as 'external accountability' (Patton, 1994: 318). A shift from a first- to second-order tradition involves a shift in focus from researcher/evaluator objectivity to researcher/evaluator responsibility.

Methodologically, it is regarded as good practice within the evaluation field to draw on the methods from a repertoire of well-established social science research, with a preference for multiple methods, and methodological plurality, as defined in evaluation guidelines of major evaluation professional organisations at national and international level (e.g. the AEA and the EES). Arguably most guidelines support evaluation practice in terms of first-order, external accountability.

A second-order research tradition on the other hand is characterised by experiential and relational understanding of practitioners engaging with situations they themselves are part of rather than being distanced observers. In evaluation practice, this would, for example, involve critical reflections and reflexivity about a reality that includes the evaluator. The reality is brought forth with participation and inclusion of the evaluator and other practitioners involved in the situation.

Second-order reflexivity has found its way into some forms of evaluation practice already. For example, through the incorporation of elements of action research, as practicing evaluation as an ordinary everyday part of what we do. Wadsworth (2016) introduces doing 'evaluation on the run' to non-specialist evaluators to develop a culture of ongoing evaluation as part of their normal business. This bears some common features with internal evaluation (Love, 1991), and enacts continual action-learning cycles of observation, reflection, dialogue and implementation to be applied to all our actions (a form of small r research) as cycles of continuous monitoring and evaluation (Wadsworth, 2016: 94).

Similarities might also be found through some aspects of DE (Patton, 1994, 2011). Patton comes close to describing second-order practice in relation to DE, while not using this term or overtly drawing upon its associated intellectual and methodological lineages (Patton, 2018).

Some common features within the STiP tradition suggest that DE and Blue Marble Evaluation (Patton, 2019) might provide a promising avenue towards systemic/second-order research and evaluation. However, DE to date does not make explicit claims towards

advocating second-order practice. Furthermore, DE is founded on a contingency viewpoint; suggesting that DE is only appropriate for specific evaluands, such as niche interventions associated with social innovation.

DE also considers the evaluator as an embedded and embodied constituent stakeholder along with others in the evaluand. In situations or evaluands deemed by DE practitioners as being complex, the role of the evaluator requires a sense of ethical internal responsibility where the evaluator is part of, rather than external to, the evaluand. But what is complex is not an ontological choice but an epistemic choice based on the capabilities enacted through practice; as Cook and Wagenaar (2012: 9) claim in their epistemology of practice ‘as an inquiry into the possibilities and constraints of being engaged, embodied, contextualised agents’.

Our claim in this article, in contrast, is that second-order evaluation based on a tradition of STiP has value in all evaluands, among all evaluators and for all evaluations (see Reynolds, 2015).

Features of a critically reflexive systemic evaluation that would constitute a second-order research tradition would include, for example, the following:

- Appreciating that the evaluand’s context is constantly changing;
- Questioning the terms of reference (ToR) for an evaluation (the ascribed purposes/standards used);
- Iterating on measures of success and other criteria used for evaluation;
- Adapting tools at hand rather than seeking a reified ‘best-practice’ or ‘best-fit’ tool(s);
- Regarding the evaluator as part of the evaluand rather than separate from it.

In terms of desired outcomes and benefits, it is appropriate to recall three principles for more systemic evaluation, initially proposed by Reynolds et al. (2015). We argue that through the enactment of such a second-order shift, evaluations, evaluators and the evaluands can display the following:

1. More systemic, reflexive and humble boundary conversations between values (evaluations) and unbounded reality (evaluand);
2. More empathic, ethical and response-able engagement with evaluand stakeholders based on reflexivity;
3. A more adaptive use of ‘tools’ and methods as part of evaluation praxis, while recognising the limitations and ultimate fallibility through increased epistemic awareness.

Other examples of evaluation initiatives implicitly aligned with second-order practice include, as mentioned, Wadsworth’s (2016) ‘evaluation on the run’ to develop a culture of ongoing evaluation as part of normal business, values-based evaluation (Hall et al., 2012) and Schwandt (2017) calling for more democratic professionalism in evaluation. A specific interest for systems approaches in evaluation lies in the area surrounding values and ethics, and the use of boundary critiques, cf. works by Schwandt (2015, 2017, 2018) and Schwandt and Gates (2016). Recently, Schwandt (2019) has signalled the emergence of ‘post-normal evaluation’: mirroring the now established discourse of ‘post-normal science’ (Funtowicz and Ravetz, 1993).

Vignette 1. Experience of a 'jobbing evaluator' working in first-order practice.

The lead author works as an internal evaluation practitioner in an organisation. This role can be described as a 'jobbing evaluator', whose main professional responsibility is to engage professionally with evaluations.

Such a professional evaluation role is distinct from what might be described as a crafting/bricoleur evaluator: a practitioner who embeds evaluation into other professional practices through creative application of evaluative thinking and acting.

The day-to-day experience of working as a 'jobbing evaluator' involves first-order evaluation practices primarily aiming at external accountability. In the regulatory framework and context of the organisation, evaluation of activities and programmes is a regulatory requirement. Therefore, evaluations 'must be done', and are subject to reporting and auditing. Great emphasis is put on evaluator independence, and summative and formative evaluations, although evaluation is also valued as a source of organisation learning.

Standard professional tasks of this 'jobbing' evaluator include developing an evaluation policy and programme, developing terms of reference, commissioning evaluations to be conducted by external evaluation contractors, contract management, liaising with evaluation stakeholders, and ensuring evaluation findings and recommendations are useable, used and acted upon, and feed into organisational learning and development, as well as conducting, where appropriate, some evaluations directly as an internal evaluator. These are typical tasks shared by many monitoring, evaluation and learning (MEL) practitioner roles in many organisations, involving systematic application of professional evaluation good-practices. An outcome of the systematic approach is the production of standardised guidelines, sometimes blueprints and protocols rather than the creation and recreation of a contextualised systemic evaluation performance.

Over time, I developed a more critical appreciation of the nature of the information and knowledge provision for policy and decision-making purposes in complex situations using this standard approach. I became increasingly uncomfortable with the available methodologies and the evidential claims derived from them.

Through the study and application of STiP, I learnt that there are different ways to frame, study and evaluate interventions and policies. I came to understand that providing knowledge and evidence to policymakers is steeped in positivist assumptions and favours linear knowledge transfer models which constrains opportunities to introduce a more systemic approach to evaluation in practice.

The second-order systems approach advocated here situates the practitioner as central to their own practice and thus moves the debate from what is the best method or approach to what might be the best enactment, or performance (in the sense of a choreographer or dramaturgist), of contextualised systemic evaluation. The following vignette exemplifies what is at issue.

As the experience in Vignette 1 shows, a jobbing evaluator's practice can be firmly – or even exclusively – rooted in first-order practice, often to the perhaps unconscious exclusion of second-order practice. A concern we have is how STiP can provide the means for creative expansion of first-order evaluation practice, towards second-order evaluation practice.

Systemic evaluation informed by STiP

Systemic evaluation (SE) can be considered as part of a wider tradition of systems-based evaluation (Reynolds et al., 2016). Systemic evaluation might be a means to correct over-systematic features of conventional first-order practice. To explore the value of systems thinking further, it is necessary to first appreciate usage of the systems idea derived revealed from

within the STiP tradition. We offer distinctions between conventional systems-based evaluation (based on STiT) and systemic evaluations (based on STiP):

- (a) Systems-based evaluation is largely systematic-oriented evaluation in a first-order tradition, which is looking at the ‘system’ understood as real, ontological devices, following a positivist worldview. This worldview may be held with awareness or unknowingly. In this understanding, a system is considered to exist as a real entity, to be systematically studied through scientific methods including modelling and characterisation. Systematic systems-based evaluations are afforded by use of the log frame; evidence-based evaluations and use of experiments and methods such as RCTs.
- (b) In contrast to this ontological understanding of systems-based evaluation is systemic evaluation in the tradition of STiP, rooted in a constructivist understanding of systems. In this understanding, systems are brought forth, or distinguished by, practitioners interested to engage with a situation and understand or change it systemically. The role of a system is to be used as an epistemological rather than ontological device, that is, as a way of knowing about a situation of concern, including an evaluand.
- (c) With systemic sensibility, systems literacy and STiP capability (Ison and Straw, 2020), the systematic and the systemic can be regarded as a duality, as combining to constitute a holistic response to a situation of concern. Reframing as a duality enables breaking away from the domination of systematic thinking and practice without abandoning it, as happens when unhelpful, self-negating pairs are constituted as dualisms (e.g. mind–body). Awareness and internalisation of these distinctions enables a productive dynamic between systemic and systematic practice – frequently switching between them as required in a dynamic, interdependent, contextual and emergent relationship.

The distinction between an ontological and epistemological understanding of systems is important here. An inherent risk of naively mapping systems in comprehensive systems maps – as ontological devices – is the temptation to confuse the ‘map’ with the ‘territory’ (Korzybski, 1933: 58). A systems map developed without STiP capability can convey the misleading illusion that ‘this is the system, and it shows it the way it is’. But it is important to consider that any systems map is only a – still partial, and biased – representation of what is perceived to be the system by stakeholder-practitioners, including modellers. A systems map is how it appears to the modellers from their respective lenses, which is influenced by many factors and subject to biases, partial perceptions and omissions. ‘All models are wrong’ (Box, 1976: 792) also applies to systems models – keeping in mind that they may still be useful, with the necessary caution and epistemic awareness of their limitations.

As we will discuss, the bringing forth of systems as epistemological devices raises ethical and design concerns and possibilities. For a system to be a system involves someone making a boundary choice – a distinction between what is in and what is out of a system-of-interest. Applied to systemic evaluation and to overcome this difficulty, we therefore offer the distinction that evaluands are *situations of interest* to be explored by the evaluation, and not ‘systems’ as such.

Over the last decade and longer, authors have advocated a need for the evaluation profession to engage with systems thinking and complexity science (STCS), for example, Westley et al., (2007), Williams and Imam (eds) (2007), Midgley (2007) Patton et al., (2011). Practitioners’ toolkits were provided by Williams and Hummelbrunner (2009), and

Reynolds and Holwell (eds) (2020) offer a guide to systems approaches to change from a STiP tradition.

Of concern to us is the extent to which these claims are becoming institutionalised within the evaluation community and whether they are being institutionalised as first, or second-order STiP praxis, or both. It is apparent that investment in greater STiP capability combined with conducive institutional innovation that allows STiP to flourish is warranted.

Three elements for systemic evaluation: Interrelationships, multiple perspectives and boundaries – Applied to evaluation practice

From epistemic awareness in STiP, it follows that three elements of how systems can be viewed and approached analytically are particularly key for evaluation (Reynolds and Holwell, 2020; Williams, 2013):

- *Interrelationships* between systems components;
- *Multiple perspectives* through which a system can be viewed, by different stakeholders and from different worldviews;
- *Boundaries*. The way judgements are made about what is ‘in’ the system-of-interest, and what is ‘out’, and a critical engagement with these boundary judgements.

These three core concepts find resonance in methods and approaches developed in the STiP tradition which enable evaluations in practice, that is, to building systemic evaluation capability. Exploring situations regarded as evaluands through the lens of these three concepts open up innovation-through-design possibilities as well as enabling ethically defensible praxis.

Fortunately, showing the *interrelationships, interdependencies and causal links* between systems components has gained interest in the evaluation field in recent times. There is growing interest to ‘show the system to itself’ through the use of visual representations, or systems maps. Undertaken with epistemological awareness, STiP practitioners realise how understanding (learning) can be enhanced by choosing to map elements in a situation-of-interest *as if* they were a system. Mapping systems visually has the intention of seeking to understand and representing system components, causal links and behaviours as comprehensively and succinctly as possible. Systems maps can consist of diagrams created through a variety of systems mapping methods (Blackmore et al., 2017). Epistemologically aware practitioners always carry at the forefront of their practice the questions: Whose system? Whose boundary judgements? Their practice can also reveal whether what might be perceived as a system actually functions as a system and what its purpose may be from the perspective of different stakeholders.

A well-known example is the obesity systems map (UK Government Office for Science, 2007). This diagram captures a multitude of factors and interrelationships in one picture. It illustrates the simultaneous strength and weakness of such diagrams: there can be a temptation to pack ‘everything in’ to be comprehensive and in trying to capture ‘the whole system’. This ambition comes at the cost of understandability and accessibility for the reader. At first sight, such a diagram can be casually described as a ‘spaghetti diagram’ and is very difficult to read. It may have the opposite effect than what is intended, as it can put a reader (e.g. policymaker) in a position of feeling overwhelmed by the complexity and lead them to disengage, rather than feeling empowered to see intervention possibilities.

In the hands of an experienced STiP practitioner, the different choreographic possibilities of their practice with other stakeholders are appreciated. A systems map done alone is not the same as one done with others (raising the question of which others). The act of mapping as an emerging creation, following multiple iterations with others, is in itself a mini-learning system.² Primary insights arise in the process itself among those participating. A final map, used for presentational purposes, creates far more limitations in comparison to an enacted learning system. What is systemic often becomes systematic, following a linear mode of communication-as-delivery.

Multiple perspectives of a system-of-interest by different stakeholders are traditionally in focus in evaluations. Many evaluation approaches have been developed to address and capture different stakeholder perspectives (e.g. participatory evaluation, empowerment evaluation and other approaches). Systems approaches can be combined and supplement these well-established evaluation approaches. For example, Soft Systems Methodology (SSM) is well suited to dig deeper into the different perspectives – not only descriptively – what different stakeholders think – but also what the underlying philosophical foundations are that influence these differences in perspectives: the underlying ‘worldviews’ (Weltanschauungen) which inform the different perspectives (Checkland and Scholes, 1990).

Making boundary choices and judgements, and critical reflections about these choices, is central to evaluation design and implementation. For example, evaluation commissioners frequently pre-determine the boundaries of what is in scope for an evaluation in the evaluation ToR. With a view to the stakeholders that are affected but not necessarily involved in an evaluation, it is a question of evaluator responsibility (ethicality) to be critically aware of the boundary choice that is implied in the scope of a ToR. It is therefore important for evaluators to have ways to ethically address these choices and their implications. As an example, Critical Systems Heuristics (CSH) (Ulrich and Reynolds, 2020) can be mentioned as one of the more popular systems approaches with application in evaluation practice, possibly for this reason. CSH has been found useful and has been used by evaluators and quite extensively written about. Examples can be found in works by Ulrich and Reynolds (2020), Gates (2018) and Stephens et al. (2018).

CSH is well suited to conduct boundary explorations and critiques and can also be helpful in revealing interrelationships and multiple perspectives. Through the format and underlying concepts, CSH can offer a ‘framework for understanding’ a situation being evaluated. CSH contains 12 critical questions that can be asked about a situation, in two different modes: how it ‘ought to be’ (in an ideal situation), and how it ‘is’ (in reality). This contrast between normative and actual mode of application of these questions can be used dialectically to critically explore and expose the boundary choices that have been made and highlight the consequences of these boundary decisions across different dimensions.

From a STiP perspective, the task of exploring an evaluand is always to start systemically (as with Reynolds’ model of an evaluation-adaptive complex; Reynolds, 2015). This involves not (systematically) pre-judging a situation as being either simple, complicated, complex, or wicked, but rather to assume that the situation will in all likelihood have a mixture of features which vary with perspective, that is, the situation of concern will be open to the framing choices of the practitioner(s). STiP is then about making the situation more amendable to purposeful action by bounding the evaluand within a situation of concern, but always with some element of systemic awareness.

Complexity thinking in practice

Complexity thinking – which is (arguably) a sub-set of systems sciences – has come into good currency in the evaluation profession over the last decade, as can be seen by numerous publications and discursive discussions at conferences. Forss et al.'s (2011) compilation of 'evaluating the complex' was one attempt to bring complexity science concepts into view in evaluation theory and practice. Other authors like Ramalingam (2013) and Bamberger et al. (2016) have explored how dealing with complexity can be applied to the domain of development evaluation to become more complexity responsive.

Tensions between STCS advocates in the evaluation field may appear to have been resolved by detecting an emerging relationship between complexity, systems thinking and evaluation (Reynolds et al., 2012). The debates continue though as documented in special journal editions, for example, special editions of the Bulletin of the Institute of Development Studies (IDS Bulletin, 2014) dedicated to 'exploring the potential of systems ideas and complexity concepts to meet the increasingly complex challenges of an increasingly ambitious development agenda' (p. 1). The tensions exposed and debated between systems thinking on the one hand and complexity science on the other include contrasting uses of the systems idea as either an ontological device for understanding complex systems, that is, modelled as real observable entities, or alternatively as epistemological devices for understanding and engaging with situations, that is, as constructions for social learning and understanding. Within the complexity discourse 'complex adaptive systems' (CAS) is frequently used unreflexively as an ontological device rather than as a conceptual framing which can be chosen; that is, what could be gained by considering this situation *as if* it were a CAS?

Despite recent investments and some promising examples there continue to be relatively few cases of genuinely systemic evaluations in practice, although this may be changing. Kusters et al. (2019: 34) list several promising examples. With reference to monitoring and evaluating the sustainable developmental goals (SDGs), there are calls for complex systems thinking and using systemic approaches to evaluations that connect (Ofir et al., 2019). There are some examples in the areas of eco-systems (Müller and Sukhdev, 2018), and examples in the works of Stephens et al. (2018).

Despite growing calls for systemic evaluation, there continues to be a gap between what is claimed and promoted (namely, to promote whole-systems evaluations or complexity-responsive evaluations) on the one hand, and the actual implementation and practical use of systems and complexity approaches and methods in day-to-day evaluation practices, outside of the show-cased examples in the literature.

The research-for-practice-reform gap

Existence of a practice gap resonates with the lead author's own research and professional experience based on attempts to introduce systems approaches in evaluation practice 'by stealth'. This experience motivates a programme of 'small r' and 'big R' research to address this gap based on first-person inquiry and case study research, respectively. 'Stealth' attempts expose practical difficulties and tensions. Documented examples of these difficulties were reported in a 2019 workshop report by the 'Scaling Solutions toward Shifting Systems initiative' of the Rockefeller Philanthropy Advisors, devoted to assessing systems change, and building philanthropic funding organisations' capacity. Workshop participants were asked to identify what they believed were the barriers to adopting systems evaluation approaches in

their organisations and the (development) sector as a whole. A diverse set of observed barriers was elicited (Rockefeller Philanthropy Advisors, 2019: 2). This list ranges from lack of knowledge and appreciation, unclear definitions and concepts to lack of resources and organisational capacities, and the dominant use of the logic-model paradigm for evaluations which is not suited to systems-wide change.

In addition to this list, from the authors' own experiences in the field, the scarcity of practical examples of systemic evaluations and reported difficulties point to a gap between an espoused theory of what is claimed needs to be done (e.g. to conduct an evaluation that does justice to the complexities encountered in the evaluand) and what happens in practice (theory-in-use) (Argyris and Schön, 1974).

This poses the question: how can such a gap be bridged? The primary argument arising from our work is that taking a second-order approach to evaluation practice can help to overcome existing dichotomies and dualisms, to bridge the practice gap. A shift needs to happen, and yet there are many barriers along the way.

One such barrier is the epistemic contrast of applying systems thinking as a first-order (ontologically fixed) application compared to the second-order understanding of STiP as an epistemic approach for learning and exploration of a situation-of-interest. Vignette 2 illustrates an experience by the lead author where an epistemic clash arose during the attempt to design a systemic evaluation. In this example, this contrast manifested itself as a dualism – rather than a duality – with the result that the ambition for a systemic evaluation was abandoned as the differences could not be reconciled.

Distinctions between first- and second-order traditions. The example in Vignette 2 shows how different practitioners grounded in different traditions of systems practice approach their practice in very different ways. It is well known that paradigmatic and epistemological commitments differ within disciplinary fields; the challenge is to bring them into awareness and conversation.

In Vignette 2, reflections about this practice experience revealed that the SD practitioner had approached the project from the understanding of a first-order systems tradition, whereas the STiP practitioners came to it from a second-order tradition. These paradigmatic differences in understanding needed mutual recognition and a shared language and repertoire and were not recognised from the outset. Initially, all members of the team had assumed that they were all systems practitioners – albeit from different schools – and it would therefore be possible to come to a shared understanding. The differences between first- and second-order systems traditions had been unexpressed and thus underestimated by both parties. Only through the failure of this project was the depth of the gap between the two understandings and epistemological stances revealed.

The experience presented in Vignette 2 reinforced the question of how a shift between first- and second-order systems traditions can be enacted between practitioners in given situations. The role of the practitioner him/herself within the dynamic between the practitioner(s), the situation, and the methods and frameworks used seems to be crucial, as well as the practitioners' relationships between themselves. Capacity of team members to have a generative conversation and joint reflection from which something new can arise from an experience seems essential, that is, reflexivity.

In the following, we propose a way to create opportunities to bridge this existing gap, towards more systemic evaluation practices by putting evaluation systems practice and the

Vignette 2. Example of an experience of the gap between first-order application of systems thinking with second-order understanding of systems practice.

The lead author was one of two OU STiP practitioners who were approached to get involved in an evaluation project which had an explicit ambition to incorporate systems thinking approaches. The client had previous exposure to System Dynamics (SD), one of many theoretical and practical lineages within the systems field, and had a very positive experience of her applications of SD. From the perspective of this practitioner, explicit inclusion of SD into the project design was needed in order to make the evaluation more systemic. The client's expectation was that the use of the SD technique of Causal Loop Diagrams (CLDs) in particular would enhance the understanding of the complexity of the situation to be evaluated. Due to this explicit request for SD, the STiP practitioners felt they needed to secure some additional SD modelling expertise beyond their own for this project. They contacted an SD consultant with view to exploring the feasibility to form a joint project team. The members of this potential project team attempted to co-design a customised methodology for this project involving evaluation concepts such as theory-of-change, combined with elements of systems approaches consisting of Soft Systems Methodology (SSM) and Critical Systems Heuristics (CSH), and SD.

The experience of co-designing and negotiating this methodology turned out to be very difficult due to very different understandings of the systems traditions which were exposed during the exploration phase. Over the course of the exploration, it emerged that the assumption of the SD consultant was that the situation to be modelled is a fixed system, which needed to be modelled as a whole using a range of interconnected CLDs. This understanding can be described as 'this is the system which I need to understand in order to engineer and model it'. The SD consultant understood their role as an external objective observer and expert modeller, who needed to objectively model 'the' system, in order to mirror it back to the stakeholders in the situation. (This understanding is similar to that of mainstream consultants, business analysts as well as some evaluator practitioners).

The STiP practitioners on the other hand had a very different understanding of the situation arising from their exposure to a different set of systems traditions: the 'system' concerned in the situation to be evaluated does not exist as such in a pre-conceived form. It is not a system 'out there'. It can depend on the different stakeholders involved who have very different experiences of the situation being evaluated (multiple perspectives), and for whom the evaluand may have very different purposes. Therefore, there can be confusion, contestation and complexity. The boundaries of the evaluand represented as a system (what is considered to be part of the system or outside of it) also differ and can change – different boundaries can be drawn for different purposes and by different actors.

Rather than modelling some pre-conceived system, understanding the evaluand means to organise an exploration (or evaluation) of it as a learning system. The role of the STiP practitioner is then to organise a systemic inquiry for an exploration of the situation for the purpose of learning and transforming the situation of concern.

How did the experience end?

In this case example, the very different assumptions and understandings of systems that were exposed through this feasibility exploration were experienced as profound, with the result that the project was not feasible to be implemented in this combination, and leading to abandonment.

practitioner at the centre (enacting small r research/learning) and proposing a research agenda (which can be understood more as big R research).

Evaluators as systems thinking practitioners?

To make the desired shift towards second-order evaluation practice, evaluation practitioners can benefit from approaching their practice by drawing on heuristics that reveal key

choreographic, or performative, relational dynamics. Drawing on the repertoire from the STiP tradition, a possible vehicle for enacting second-order practice that has been found useful within Open University (United Kingdom) STiP education is that of systems practice as comparable to a juggling act (Ison and Blackmore, 2014).

Systems practice as a juggling act for evaluation practitioners – A social dynamic

Ison (2017) conceives systems thinking and practice as an active social dynamic. To bring this to life, he introduces the isophor³ of a systems practitioner as a juggler, keeping several balls in the air as part of the juggling act (p. 60). We here briefly apply the image (isophor) of this juggling act to the practice of enacting an evaluation.

The evaluation practitioner is the key player in the ‘performance’ of an evaluation. Similar to the juggler in a juggling performance, the practitioner is invited to ‘juggle’ four different ‘balls’ that need to be played and kept in the air during an evaluation (performance): these ‘balls’ are Being, Engaging, Contextualising and Managing (BECM):

- *The ‘B’ ball* concerns the ‘Being’ of an evaluation practitioner, with awareness of his or her tradition of understanding that informs his or her evaluation practice, and the ethical responsibility he or she needs to take in enacting the evaluation. Being a reflective/reflexive practitioner is a key capability to nurture this capacity, to take ethical responsibility for our actions as evaluators. This concern extends to those that are potentially affected by the consequences of the evaluation, or the situation being evaluated. As evaluation practitioners, we need to be constantly aware of such consequences which instil our ethical responsibility.
- *The ‘E’ ball* is for ‘Engaging’ with the situation the evaluation is concerned about and engaging with the evaluand. Juggling the E-ball requires awareness and agency to make the choices available to the evaluation practitioner of how to engage with the real-life situation of an evaluation. How situations are perceived has important implications for the frames chosen for it and choices made for their evaluation. For example, a situation may be perceived as well-defined (normal), or uncertain and complex. Critically exploring and reframing our perceptions of the situations we engage with in evaluation and developing our appreciative settings (Vickers, 1970) is then key to hone the capacity to engage reflexively in situations we are evaluating (e.g. Ison, 2018).
- *The ‘C’ ball* is about ‘Contextualising’. Applied to evaluation, this concerns the choices of methods, techniques and tools to be used for an evaluation. When evaluation practitioners decide on which methods or tools they will use for an evaluation, they contextualise their evaluation practice to a specific evaluation situation, for example, for a specific evaluation assignment. In the STiP tradition of understanding, the distinction is made between tools, techniques, methods and methodologies. Ison (2017) describes methodology as the ‘conscious braiding together of theory and practice in a given situation, as a context specific enactment’ (p. 167). It requires a broad awareness of concepts, knowledge of techniques and tools, and methods. A methodology involves the design of adapted methods (‘bricolage’) customised to fit for the specific situation, in a way that feels systemically ‘right’.

Contextualising may also involve the exploration of the evaluation purpose from the perspective of different stakeholders and stakeholdings which may differ from the overtly stated purpose in the ToR provided by the evaluation commissioners. A boundary critique could explore the contestations of the purpose of the system-of-interest, and the interests involved. An adaptation of the CSH 12 question framework (Ulrich, 1983, 1996), for example, can lend themselves as bridging practice (between Contextualising and Engaging) to explore who is/ought to be the systems' client, or whose interests are/ought to be served by it (Ison, 2017: 166).

- Finally, 'Managing' involves simultaneous looking both outside to interaction with others, as well as looking inwards. In evaluation practice, this concerns how the evaluation practitioner manages his or her involvement in the evaluation: the performance and the relationship with the evaluation commissioners and other stakeholders, and indeed the overall context. Effective management of an evaluation (juggling) performance crucially involves nurturing and maintaining of meaningful and engaging relationships through the flux of time (Vickers, 1978: 71–72).

Centrally, the juggler herself is the person who keeps all these balls in the air, through their practice. Juggling involves the whole body and mind. The juggler uses her own body throughout the performance and thus brings it into being: by throwing the balls into the air and catching them again, balancing her body through contact with the floor in response to the motion of the different balls. It is a co-evolving and adaptive practice, an interactive social dynamic, in which the juggler is coupled in relation to the four balls.

These BECM dimensions can unfold their full power and relevance by relating them to the core theme of STiP of juxtaposing systematic (or first-order) and systemic (or second-order) practice, applied to evaluation practice.

- 'Being':
 - Expressed systematically can mean to apply the evaluation 'tools of the trade' instrumentally (to get the job done). The evaluator acts like a tradesperson.
 - In systemic mode, 'Being' involves ethics, by acknowledging wider consequences of the work done (evaluator as craft artisan/bricoleur).
- 'Engaging':
 - In systematic mode, the evaluand is often framed as a tame (solvable) difficulty or problem. The evaluator acts with confidence among stakeholders, including commissioners and intended beneficiaries and co-evaluators.
 - In systemic mode, the evaluand is better framed as a (potentially) wicked problem situation.
- 'Contextualising':
 - In a systematic approach, this means fulfilling contractual obligations, and keeping commissioners assured.
 - Understanding contextualising in systemic mode, it can mean to (courageously) question the ToR from the commissioner.
- 'Managing':
 - When viewed from a systematic framing, managing an evaluation focuses on maintaining immediate task-oriented relationships with stakeholders.

- Approached from a systemic framing, managing an evaluation involves a concern for forging longer term and wider relations, beyond the immediate task and concern.
- When enacted as duality (i.e. both systematically and systemically), an evaluation can function effectively (i.e. be managed) by generating learning through feedback, and thus adaptation and change.

Adding capacity building for second-order evaluation practice to the evaluation professionalisation agenda

Enacting the juggling of these four balls can deliver both systematic effectiveness and systemic effectiveness. Its use shows how it is possible to re-frame the engagement and for doing systemic evaluation by switching between the two modes, as appropriate to the situation. Von Förster (1992) understood this as the essence of an ethical performance because more choices are offered to those within the situation.

The STiP focus on the evaluator as practitioner and their central role in enacting responsible and systemically desirable evaluations could be seen as a response to the call for an enhanced evaluation ethos within evaluation professionalisation discourse. Schwandt (2015, 2017) criticises the dominant focus on normative technical evaluation knowledge and competencies in the advancement of the evaluation professionalisation agenda and notes a lack of vigorous discussion of developing a professional ethos for evaluation. Professionalising evaluation as an occupation (and supply-and-demand commodity and service) with a focus on credentials and certification falls short of advancing evaluation as a public good of societal value. According to Schwandt (2017: 548), the professional ethos refers to the ‘sum of a professional group’s moral principles, core values, epistemic and aesthetic dispositions, and aspirations that each member of the group takes into consideration in interacting with others in a professional context’. Evaluation as a public work, as advanced by Schwandt and Reynolds: ‘. . . combines insights [. . .] with ideas from critical system heuristics and the literature on knowledge utilization’ (Reynolds and Schwandt, 2017; Schwandt, 2017: 550).

If evaluation as a profession really wants to live up to the proclaimed goal of contributing worldwide to ‘a transformed global community characterized by transparency, accountability, and progress towards the common good’ (EvalPartners, 2016: 3), then, as Schwandt (2017: 552) concludes, ‘we need a much more public and energetic discussion of the professional ethos of evaluation, [. . .] what actually comprises its shared understanding of moral principles, values, aspirations, and ways of behaving’.

From the heritage of STiP, a focus on the evaluation practitioner in bringing about a second-order shift in evaluation practice can be one way to contribute to a professional ethos of evaluation practice. Second-order evaluation practice can be in support of enabling post-normal evaluation practice which is needed to respond to the post-normal characteristics of our time and the near future. Schwandt (2019) argues that the conventional understanding and practices of ‘normal’ evaluation are no longer sufficient. It is time for ‘post-normal’ evaluation to come to the fore to more adequately deal with complex situations of change and uncertainty, he claims.

Further professionalisation and capacity building for ethically reflexive, second-order systemic evaluation practice might also contribute to advance the still weak institutionalisation of

systems thinking in organisations as well as rules and norms of conceptualising, commissioning, conducting and using evaluations. There are already opportunities in evaluation professionalisation and capacity building efforts which can be built upon, for example, as mentioned below:

- The Global Evaluation Agenda 2016–2020 (EvalPartners, 2016) includes a chapter dedicated to the ‘strengthening of individual capacities for evaluation’ (Chapter A.3: 21–28), including ethical dispositions, professional autonomy, expertise and credentials (p. 21).⁴
- The innovative Voluntary Evaluator Peer Review (VEPR) pilot projects which have been implemented by the EES and UKES evaluation societies as a professional development service for their members. These are grounded in reflective practice principles, which as a central position in these professional societies’ evaluation capability frameworks, lay a fertile foundation for further innovation/professionalisation.

The next turn in this positive development may become the nurturing of second-order evaluation practice, by enhancing and expanding reflexive systemic evaluation, that is, second-order, systemic evaluation practice principles incorporated into capability frameworks. This requires extension of contemporary evaluation research agendas.

An ongoing research agenda

There is an ongoing research imperative to effect the means and substance of a transformed, second-order, systemic evaluation praxis. The authors’ own research agenda is directed at this imperative.

Research is currently under way to address and answer research questions that seek to elucidate the constraints and possibilities for transforming towards second-order practices by evaluation practitioners:

- How do evaluation practitioners engage with complex situations of change and uncertainty?
- How do evaluators reflect on the choices for approaches and methods in these situations?
- What opportunities exist for evaluators to make a ‘second-order’ shift in these situations?

The research process is designed as a set of nested learning systems grounded within and framed by situated practices. Using systemic action research principles, this research is firmly rooted in the experiential practice of the lead author (illustrated in Vignettes 1 and 2), and is conducted through three modes, or cycles, of inquiry and with differing researcher positionalities: first-, second- and third-person inquiry (Torbert, 2001):

- The practitioner/researcher as ‘evaluation and systems practitioner-conceptualizer’ and researcher embedded within this ecosystem and praxis field (through continuous first-person inquiry; ‘learning for me’);
- Practitioners engaged in the area of practice of evaluation (through second-person inquiry; ‘learning with others’ as co-inquiry);

- Actors – and practices – in the wider situation of interest of developing (complexity-sensitive) public policy knowledge and systemic evaluation practice (third-person inquiry; ‘engaging in learning with a wider community’).

This research is designed to ‘walk the talk’ of systemic practice by applying it to research and evaluation as a practice. The inquiry speaks to the broader developments and interests in the fields of public policy, science technology studies and evaluation that are concerned with the question of how policies and decision-making can become more systemic in order to respond to accelerating complexity in a world that is now increasingly ‘beyond the stable state’ (Schön, 1971). Where traditional evaluation practices are becoming less effective in complex situations of change and uncertainty, evaluation practitioners must become more epistemologically and ontologically aware and better equipped to effect change that is systemically desirable as well as culturally feasible (Checkland and Scholes, 1990).

The proposed research is designed to avoid entrapment in a rigid systematic practice that assumes the linear transfer of knowledge. By moving towards third-person action inquiry, the opportunity exists to use prior learning from these other modes as input into possible designs for learning systems enacted with other stakeholders, that is, knowing-in-action. In turn this creates the possibilities for emergent, contextualised, transformation instead of the mainstream focus on delivery and adoption, that is, knowledge transfer. Knowing-in-action can also be a means to enact that important systems concept – feedback – through participation in a learning system.

Learning from this research is expected of value for the theoretical advancement and contribution to knowledge of the evaluation discipline, but also of practical value to evaluators, to enhance their personal capacities to be, and engage, in complex situations of change and uncertainty, acting with systemically aware responsibility when doing evaluations.

Conclusion: Opportunities for second-order practice shifts in evaluation practice – Towards systemic evaluation practice and practitioners

Duffy (2017: 149) argues that it ‘is only through remaining open to potential, yet unknown emergent transformations that the disciplinary and controlling effects of knowledge production processes can be unsettled’. We conclude that a second-order shift in evaluation praxis is not only justified, but necessary, as part of an unsettling project relevant to our human circumstances in the Anthropocene (Ison, 2016).

A second-order shift helps to generate greater epistemological awareness in approaching evaluation and engaging evaluands as situations of interest that can be understood, utilised and transformed systemically. In contrast to traditional first-order evaluation practice, reality is brought forth relationally and experientially with participation and inclusion of others involved in a situation. ‘Evaluation is all about moving beyond where we are now’ (Duffy, 2017: 150).

A second-order shift involves assuming at the outset that all situations/evaluands have elements of complicatedness, complexity and conflict. Key to such a shift is the role of the practitioner – the evaluator herself. In this understanding, evaluators or researchers are themselves part of the situation they seek to understand, change, evaluate, transform using systems as ways of knowing, inquiring or doing. A second-order shift brings awareness that an ‘as if’ position is possible – to see situations as systems to learn in a particular way. A non-reflexive

commitment to first-order, systematic practice limits change possibilities and the focus of praxis to that of external observer of a system that is ‘out there’ (as with a mainstream external accountability perspective of evaluation).

Evaluation practitioners are at the heart of systemic evaluation as understood in the tradition of STiP. Evaluators can develop their own praxis capacity and capability through reflexive use of the juggler isophor: their own Being, Engaging with the situations of interest, Contextualising their systemic evaluation performances and Managing their overall evaluation performance (BECM). Developing these capacities for systemic evaluation practice can be integrated into evaluation professionalisation and capacity building efforts, building further on reflective practitioner elements.

Future professionalism and the promise of efficacy of a second-order shift to systemic evaluation will need to deal with those systemic issues that plague the evaluation professional. Wadsworth (2010: 271) listed these as: ‘(i) changes that are wanted but don’t happen; (ii) changes that happen that people (stakeholders) don’t want; (iii) decisions and politics that seem unresponsive, insensitive or prematurely pragmatic; (iv) inaccurate assumptions, over-generalisations, and an inability to see and hear what people are really saying; (v) preoccupation with fixing things that are going wrong with little or no time spent to make things right in the first place, or (vi) solutions becoming new problems . . .’. To this we could add evaluations that merely tick a box or where the report sits on the shelf, where there is no feedback, learning and change. STiP-informed research and evaluation practice can support practitioners in knowing how to build supportive contexts for second-order systemic evaluation praxis – overcoming the constraints as well as institutionalising the enablers.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This study was independently funded by the authors. The article is informed by research done for the first author’s PhD in progress.

ORCID iD

Barbara Schmidt-Abbey  <https://orcid.org/0000-0003-3125-1163>

Notes

1. The readership of this journal is well informed about the content and contentions of this methodological discourse, which goes beyond the scope of this article.
2. A ‘learning system’ has ‘built-in capacity for reflection on experience and to recognize change or learning when it occurs’ (Ison and Russell, 2000: 208).
3. Juggling in this case is not a metaphor (which takes one concept into a new domain) but an isophor in which the same dynamics are apparent in different domains. This follows the understanding of Maturana, where ‘juggling is an isophor of the vision one wants to have to claim that one understands, for example, a biological or a cultural happening, such as effective systems practice’ (Ison, 2017: 61).
4. The document also recognises the importance of ‘reflective practice’ for the individual evaluator’s capacity building. The evaluation community has also developed a number of evaluators’ capabilities frameworks focused on knowledge, practice and dispositions.

References

- American Evaluation Association (AEA) (n.d.) Available at: <https://www.eval.org/> (accessed 5 January 2020).
- Argyris C and Schön D (1974) *Theory in Practice: Increasing Professional Effectiveness*. San Francisco, CA: Jossey-Bass.
- Bamberger M, Vaessen J and Raimondo E (2016) *Dealing with Complexity in Development Evaluation, a Practical Approach*. Thousand Oaks, CA: SAGE.
- Blackmore C, Foster N, Collins K, et al. (2017) Understanding and developing communities of practice through diagramming. In: Oreszczyń S and Lane A (eds) *Mapping Environmental Sustainability: Reflecting on Systemic Practices for Participatory Research*. London: Policy Press, 155–182.
- Box G (1976) Science and statistics. *Journal of the American Statistical Association* 71(356): 791–9.
- Caffrey L and Munro E (2017) A systems approach to policy evaluation. *Evaluation* 23(4): 463–78.
- Cairney P (2016) *The Politics of Evidence-Based Policy Making*. London: Palgrave Macmillan Pivot.
- Catwell L and Sheikh A (2009) Evaluating eHealth interventions: The need for continuous systemic evaluation. *PLoS Medicine* 6(8): e1000126.
- Checkland P and Scholes J (1990) *Soft Systems Methodology in Action*. Chichester: Wiley.
- Cook N and Wagenaar H (2012) Navigating the eternally unfolding present: Toward an epistemology of practice. *The American Review of Public Administration* 42(1): 3–38.
- Duffy DN (2017) Evaluation and governing in the 21st century: Disciplinary measures, transformative possibilities. In: *Palgrave Studies in Science, Knowledge and Policy*. London: Palgrave Macmillan Pivot.
- European Evaluation Society (EES) (n.d.) Voluntary evaluator peer review (VEPR). Available at: <http://europeanevaluation.org/community/thematic-working-groups/twg4/voluntary-evaluator-peer-review> (accessed 5 January 2020).
- European Evaluation Society (EES) (2018) 13th European Evaluation Society Biennial Conference, ‘Evaluation for more resilient societies’, Thessaloniki, 1–5 October. Available at: <http://www.ees2018.eu/> (accessed 5 January 2020).
- European Evaluation Society (EES) (2020) 14th European Evaluation Society Biennial Conference, ‘Evaluation in an uncertain world: complexity, legitimacy and ethics’, Copenhagen, 21–25 September. Available at: <http://www.ees2020.eu/> (accessed 24 February 2020).
- EvalPartners (2016) EvalAgenda 2020, Global Evaluation Agenda 2016–2020. Available at: <https://evalpartners.org/sites/default/files/files/Evalagenda2020.pdf> (accessed 5 January 2020).
- Forss K, Marra M and Schwartz R (eds) (2011) *Evaluating the Complex: Attribution, Contribution and Beyond (Comparative Policy Evaluation Series 18)*. New Brunswick, NJ: Transaction Publishers.
- French RD (2018) Is it time to give up on evidence-based policy? Four answers. *Policy & Politics* 47(1): 151–68.
- Funtowicz S and Ravetz J (1993) Science for the post-normal age. *Futures* 31: 735–55.
- Gates E (2018) Towards valuing with critical systems heuristics. *American Journal of Evaluation* 39: 201–20.
- Hall JN, Ahn J and Greene JC (2012) Values engagement in evaluation: Ideas, illustrations, and implications. *American Journal of Evaluation* 33(2): 195–207.
- Ison RL (2016) Governing in the Anthropocene: What future systems thinking in practice? *Systems Research and Behavioral Science* 33(5): 595–613.
- Institute of Development Studies (2014) *IDS bulletin*: 45(6) November.
- Ison RL (2017) *Systems Practice: How to Act, in Situations of Uncertainty and Complexity in a Climate-Change World*, 2nd edn. London: Springer.
- Ison RL (2018) Governing the human-environment relationship: Systemic practice. *Current Opinion in Environmental Sustainability* 33: 114–23. Available at: <https://doi.org/10.1016/j.cosust.2018.05.009> (accessed 30 January 2020).

- Ison RL and Blackmore C (2014) Designing and developing a reflexive learning system for managing systemic change. *Systems* 2(2): 119–36.
- Ison RL and Russell D (eds) (2000) *Agricultural Extension and Rural Development: Breaking out of Traditions*. Cambridge: Cambridge University Press.
- Ison RL and Straw E (2020) *The Hidden Power of Systems Thinking: Governance in a Climate Emergency*. London: Routledge.
- Korzybski A (1933) *Science and Sanity: An Introduction to Non-Aristotelian Systems and General Semantics*. Englewood, NJ: The International Non-Aristotelian Library Publishing Company.
- Kusters C, et al (2019) Conference report: Monitoring and evaluation for inclusive and sustainable food systems. Report WCDI-19-066, Wageningen Centre for Development Innovation, Wageningen University & Research, Wageningen, 3–4 April. Available at: <https://edepot.wur.nl/506604> (accessed 5 January 2020).
- Love AJ (1991) The process of internal evaluation. In: Love AJ (ed.) *Applied Social Research Methods: Internal Evaluation*. Newbury Park, CA: SAGE, 36–62.
- Maturana H and Poerksen B (2004) *From Being to Doing: The Origins of the Biology of Cognition*. Heidelberg: Carl-Auer.
- Midgley G (2007) Systems thinking for evaluation. In: Williams B and Imam I (eds) *Systems Concepts in Evaluation, an Expert Anthology*. Point Reyes, CA: American Evaluation Association, 11–34.
- Müller A and Sukhdev P (2018) *Measuring what matters in agriculture and food systems*. A synthesis of the results and recommendations of TEEB for Agriculture and Food's Scientific and Economic Foundations report. Available at: <https://pdfs.semanticscholar.org/f245/eaf4032352bb6a93e677251ab985cb8dce95.pdf> (accessed 6 March 2020).
- Ofir Z, Singh G, Beauchamp E, et al. (2019) From monitoring goals to systems-informed evaluation: Insights from SDG14. IEED Briefing, March. Available at: <https://pubs.iied.org/17706IIED/> (accessed 5 January 2020).
- Patton M (1994) Developmental evaluation. *Evaluation Practice* 15(3): 311–20.
- Patton M (2011) *Developmental Evaluation Applying Complexity Concepts to Enhance Innovation and Use*. New York: Guilford Press.
- Patton M (2018) *Principles-Focused Evaluation*. London; New York: Guilford Press.
- Patton M (2019) Blue marble evaluation. Available at: <https://www.utilization-focusedevaluation.org/blue-marble-evaluation> (accessed 5 January 2020).
- Piirainen KA, Gonzalez RA and Bragge J (2012) A systemic evaluation framework for futures research. *Futures* 44(5): 464–74.
- Ramalingam B (2013) *Aid on the Edge of Chaos: Rethinking International Cooperation in a Complex World*. Oxford: Oxford University Press.
- Reynolds M (2015) (Breaking) The iron triangle of evaluation. *IDS Bulletin* 46: 71–86.
- Reynolds M and Holwell S (eds) (2020) *Systems Approaches to Making Change: A Practical Guide*, 2nd edn. London: Springer.
- Reynolds M and Schwandt T (2017) Evaluation as public work: An ethos for professional evaluation praxis. In: *UK evaluation society annual conference: The use and usability of evaluation: Demonstrating and improving the usefulness of evaluation*, Evaluation Society, London, 10–11 May.
- Reynolds M, Gates E, Hummelbrunner R, et al. (2016) Towards systemic evaluation. *Systems Research and Behavioral Science* 33: 662–73.
- Reynolds M, Fross K, Hummelbrunner R, et al. (2012) Complexity, systems thinking and evaluation – An emerging relationship? *Evaluation Connections Newsletter of the European Evaluation Society*, 7–9.
- Rockefeller Philanthropy Advisors (2019) Scaling solutions toward shifting systems initiative: Assessing systems change: A funders' workshop report. Available at: <https://www.rockpa.org/wp-content/uploads/2019/10/Assessing-Systems-Change-A-Funders-Workshop-Report-Rockefeller-Philanthropy-Advisors-August-2019.pdf> (accessed 5 January 2020).

- Russell DB and Ison RL (2000) The research-development relationship in rural communities: An opportunity for contextual science. In: Ison RL and Russell DB (eds) *Agricultural Extension and Rural Development: Breaking Out of Traditions*. Cambridge: Cambridge University Press, 10–31.
- Schön D (1971) *Beyond the Stable State: Public and Private Learning in a Changing Society*. New York: Random House.
- Schön D (1984) *The Reflective Practitioner: How Professionals Think in Action*. Aldershot: Ashgate.
- Schön D (1987) *Educating the Reflective Practitioner: Towards a New Design for Teaching and Learning in the Professions*. San Francisco, CA: Jossey-Bass.
- Schwandt T (2015) Reconstructing professional ethics and responsibility: Implications of critical systems thinking. *Evaluation* 21(4): 462–6.
- Schwandt T (2017) Professionalization, ethics, and fidelity to an evaluation ethos. *American Journal of Evaluation* 38(4): 546–53.
- Schwandt T (2018) Evaluative thinking as a collaborative social practice: The case of boundary judgment making. *New Directions for Evaluation* (158): 125–137.
- Schwandt T (2019) Post-normal evaluation? *Evaluation* 25(3): 317–29.
- Schwandt T and Gates E (2016) What can evaluation do? An agenda for evaluation in service of an equitable society. *Evaluation for an Equitable Society*: 67–81.
- Scriven M (1991) *Evaluation Thesaurus*, 4th edn. Thousand Oaks, CA: SAGE.
- Scriven M (1996) Types of evaluation and types of evaluator. *Evaluation Practice* 17(2): 151–61.
- Scriven M (2001) Evaluation: Future tense. *American Journal of Evaluation* 22(3): 301–7.
- Scriven M (2003) Evaluation theory and metatheory. In: Kellaghan T and Stufflebeam DL (eds) *International Handbook of Educational Evaluation. Kluwer International Handbooks of Education*, vol. 9. Dordrecht: Springer, 15–30.
- Stephens A, Lewis E and Reddy S (2018) Towards an inclusive systemic evaluation of the SDGs: Gender, equality, environments and marginalised voices (GEMs). *Evaluation* 24(2): 220–36.
- Stern E (2011) Editorial. *Evaluation* 17(4): 323–25.
- Torbert W (2001) The practice of action inquiry. In: Reason P and Bradbury H (eds) *Handbook of Action Research, Participative Inquiry and Practice*. London: SAGE, 250–60.
- Ulrich W (1983) *Critical Heuristics of Social Planning: A New Approach to Practical Philosophy*. Bern; Stuttgart: P. Haupt (Paperback reprint version. Chichester: Wiley, 1994).
- Ulrich W (1996) *A Primer to Critical Systems Heuristics for Action Researchers*. Hull: University of Hull.
- Ulrich W and Reynolds M (2020) Critical systems heuristics: The ideas and practice of boundary critique. In: Reynolds M and Holwell S (eds) *Systems Approaches to Making Change: A Practical Guide*. 2nd ed. London: Springer, 255–306.
- UK Government Office for Science (2007) Foresight: Tackling obesity: Future choices. Qualitative modelling of policy options. (Online). https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/296290/obesity-map-full-hi-res.pdf (accessed 5 January 2020).
- Vickers G (1970) *Value Systems and Social Processes*. London: Penguin Books.
- Vickers G (1978) *Responsibility – Its Sources and Limits* (Systems Inquiry Series). Seaside, CA: Intersystems Publications.
- Von Förster H (1992) Ethics and second-order cybernetics. *Cybernetics & Human Knowing* 1: 9–20.
- Von Förster H (1984) *Observing Systems*. Salinas, CA: Systems Publications.
- Wadsworth Y (1997) *Everyday Evaluation on the Run*, 2nd edn. Sydney, NSW, Australia: Allen & Unwin.
- Wadsworth Y (2010) *Building in Research and Evaluation. Human Inquiry for Living Systems*. Sydney, NSW, Australia: Allen & Unwin.
- Wadsworth Y (2016) *Everyday Evaluation on the Run: The User-Friendly Introductory Guide to Effective Evaluation*, 3rd edn. Abingdon: Routledge

- Westley F et al. (2007) *Getting to Maybe: How the World is Changed*. Toronto: Random House of Canada.
- Williams B (2013) Three core concepts: Inter-relationships, perspectives, boundaries. In: *Evaluation connections: Newsletter of the European evaluation society*, June 2013, 7–8.
- Williams B and Hummelbrunner R (2009) *Systems Concepts in Action, a Practitioner's Toolkit*. Stanford, CA: Stanford Business Books.
- Williams B and Imam I (eds) (2007) *Systems Concepts in Evaluation, an Expert Anthology*. Point Reyes, CA: American Evaluation Association.

Barbara Schmidt-Abbey is a part-time PhD Researcher at the Open University. In her other fulltime professional role, she is a Monitoring & Evaluation professional at an EU agency.

Martin Reynolds joined the Open University in 2000 and is Senior Lecturer in Systems Thinking and Qualifications Lead for the postgraduate programme in Systems Thinking in Practice (STiP).

Ray Ison is a Professor of Systems at the Open University since 1994, with research and scholarship spanning the biophysical and social disciplines, and primarily interdisciplinary and collaborative in nature.