

Taking Stock of Complexity in Evaluation: A Discussion of Three Recent Publications

Lasse Gerrits & Stefan Verweij

This is the version of the article that was accepted for publication by the journal. The final, definitive article has been published in *Evaluation* by SAGE Publications Ltd, all rights reserved.

Gerrits, L.M. & Verweij, S. (2015). Taking Stock of Complexity in Evaluation: A Discussion of Three Recent Publications. *Evaluation*, 21 (4), 481-491.

Keywords

Complexity, Complexity Science, Complex Systems, Evaluation, Systems Theory

Taking Stock of Complexity in Evaluation: A Discussion of Three Recent Publications

Introduction

Arguably, the current interest in the complexity sciences has its roots in the natural sciences, often in interplay with, and enhanced by, developments in mathematics and informatics (see e.g., Mitchell, 2009). An oft-cited reason for this interest has been the increased ability of current computing systems to deal with complex mathematics and algorithms (e.g., Colander and Kupers, 2014; Waldrop, 1992). As complexity is gaining more traction in the natural sciences, so does it in the social sciences (see e.g., Castellani, 2009). It should be noted here that the complexity sciences, or ‘complexity theory’ as it is more often called, encompass a rather loosely coupled group of theories, mechanisms, and metaphors, some of which are proven and others not, and that is not always internally consistent. The language of complexity has a clear transdisciplinary appeal (cf. Simon, 1962), so its popularity across the board is not surprising.¹ For the sake of readability of the current text, we will simply refer to this group of theories as ‘complexity’.²

Naturally, complexity has also invaded the evaluation literature since the 1990s (e.g. through the work of Brenda Zimmerman), where it is increasingly discussed and applied (cf. Walton, 2014). For instance, this journal has recently published a steady number of complexity-related pieces (e.g., Barnes et al., 2003; Byrne, 2013; Callaghan, 2008; Davies, 2004; Mowles, 2014; Rogers, 2008; Sanderson, 2000; Verweij and Gerrits, 2013; Westhorp, 2012). A search within the journal on the terms “complexity theory”, “complex system” or “complexity science” yielded forty-nine articles in a growing trend (see Figure 1).³ Inquiries with Scopus into complexity and evaluation yielded similar results, as shown in Figure 2.⁴ In this article, we take stock of recent progress and discuss what complexity holds for evaluation by discussing three recent books. Before doing so, it is necessary to discuss the issue of theory transfer.

¹ Interested readers are referred to e.g., the work of David Byrne (i.a., 1998; Byrne and Callaghan, 2014) for an extended discussion about how complexity matters to the social sciences.

² It is acknowledged that this can be considered imprecise given the many ways in which ‘complexity’ is understood and discussed. However, an extensive discussion of what complexity exactly is lies beyond the purpose of this contribution. We would like to point readers to e.g., Rescher (1998) for an excellent in-depth overview.

³ The same search for the *American Journal of Evaluation* yielded twenty-three articles.

⁴ The syntax used for the results shown in Figure 2 is: TITLE-ABS-KEY ("complexity theory" OR "complex adaptive systems" OR "CAS" OR "soft system" OR "eco* system" OR "complexity science*" OR "complexity thinking" OR "complex system*") AND TITLE-ABS-KEY ("policy eval*" OR "prog* eval*" OR "policy analysis" OR "formative eval*" OR "process eval*" OR "outcome eval*" OR "impact eval*" OR "context eval*" OR "project eval*") AND (LIMIT-TO (DOCTYPE, "ar")).

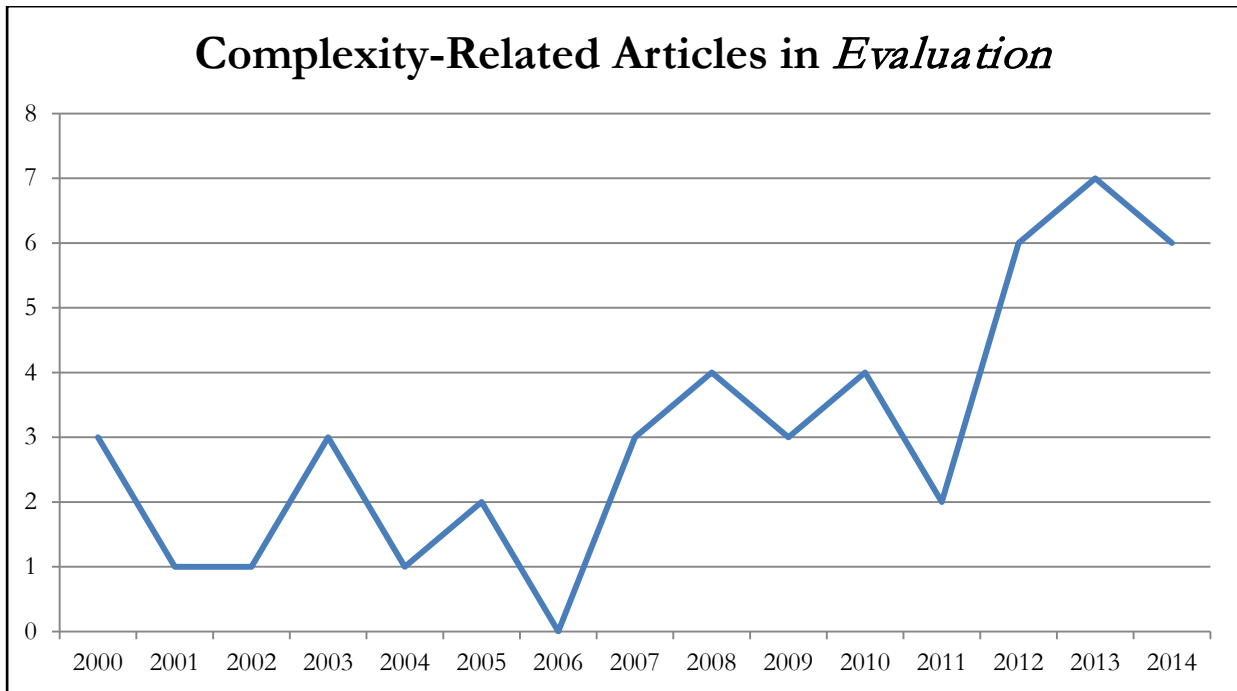


Figure 1: complexity-related articles in the journal *Evaluation* (search performed on 23-04-2015)

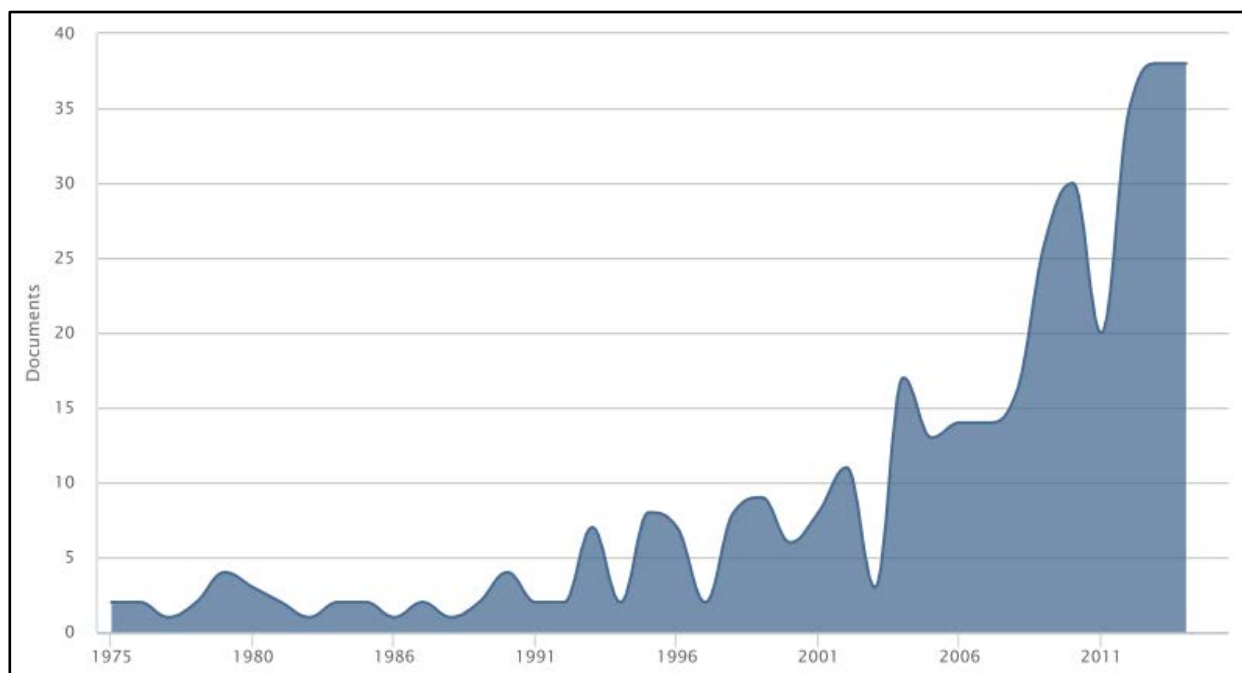


Figure 2: complexity-related publications about evaluation by year, non-cumulative (search performed on 29-06-2015)

Theory transfer

If anything, concepts from complexity enable versatile, open-ended heuristics that invite a wide range of applications and domains. The diversity of complexity's roots and domains of application brings into view the issue of theory transfer. Theory

transfer is reinforced by popular accounts that talk about complexity in generic terms and that assert that its concepts transcend disciplinary boundaries. However, it requires a leap of faith to use a concept such as Bénard instability from fluid dynamics,⁵ to name just one, to explain societal change.

The question then is whether a concept from one domain accurately describes the object of research in another domain (Knuuttila, 2011). Crucial in theory transfer is the difference between structural representation and surrogative reasoning. The first allows using a concept for drawing conclusions about the thing it represents, whilst the second allows using a concept from a domain as a surrogate in our reasoning about another domain (Swoyer, 1991). For example, the application of a certain model from the source domain to the target domain can be considered structural representation when it manages to address similar mechanisms and to generate similar types of statements. If a concept from the source domain is used as a metaphor in the target domain, it can be considered surrogative reasoning. The boundary between the two usages is not always clear; a strict application of a model cannot accommodate certain features in the target domain, rendering model modification inevitable (Bolinska, 2013). Conversely, a metaphor may in fact point to actual causal relationships that need to be fleshed out more thoroughly (Chettiparamb, 2006).

The issue is compounded by the fact that complexity seems to repeat many notions from systems theories, something already pointed out by Steven Phelan in 1999. Indeed, systems theories and associated holistic or ‘whole system’ approaches date further much further back than many of the concepts featured in complexity. François (1999) explored the emergence of systems thinking in the fields of philosophy and science and found the earliest mentions to be in the 17th century works of Descartes and Leibnitz. A more contemporary approach, i.e. one using terminology that most current scientists would feel comfortable with, can be credited to Von Bertalanffy whose General Systems Theory contained many features now rediscovered in complexity (Gerrits, 2012). And while General Systems Theory was introduced in 1968, its intellectual foundations were developed in the 1930s. Von Bertalanffy’s heritage would be further developed in the social sciences, for example by Forrester (e.g., 1971), Checkland (e.g., 1981), Senge (1990), and Flood (e.g., 1999). Although there is much utility in this fusion, it puts claims of complexity being ‘the new paradigm’ in a different light. Complexity’s transfer and application are notoriously hard and sometimes at odds with more accepted evaluation methods; however, this does not make complexity new (cf. Mowles, 2014). In fact, one could argue that evaluations were always complex. As such, we may well consider it a rediscovery.

The reason for discussing theory transfer before discussing the books is that it will help clarifying the wide diversity of approaches and applications encountered in the books. Some authors approach complexity from the daily experience that ‘things are complex’ and that, therefore, evaluations need to reflect that experience of

⁵ Prigogine and Stengers (1984) give a clear explanation of Bénard instability in Chapter 5 of their book.

complexity. Others may take the conceptual route and start from the existing corpus of complexity literature to define a revised framework for evaluation methods. Naturally, each point of entry generates entirely different types of accounts.

Of course, there are recurring themes, such as non-linearity, emergence, and far-from-equilibrium (Byrne and Callaghan, 2014). These themes stress the importance context, time, and human agency in evaluation (Verweij, 2015; Verweij and Gerrits, 2013). However, a group of related and oft-cited concepts, theorems, and ideas does not make a consistent and proven theory. This absence of a clear, universally agreed-upon causal theory of complexity for the social sciences is not necessarily problematic for complexity scientists because it can serve as template to start from in order to generate answers (cf. Byrne and Callaghan, 2014). As discussed elsewhere (Gerrits and Marks, 2014), it is inherent to maturing theories that gaps appear. The question is not whether these will appear – they will – but whether they are solved in the long run, i.e., whether signs of scientific progress are visible. What thus matters is how complexity's notions influence evaluative thinking and research, in terms of the methods used and the kind of conclusions generated.

Three books

The books discussed in this essay are Forss et al.'s (2011) edited volume *Evaluating the Complex: Attribution, Contribution, and Beyond*, Patton's (2011) *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*, and Wolf-Branigin's (2013) *Using Complexity Theory for Research and Program Evaluation*. Each of these books is explicitly positioned in the field of complexity and evaluation.⁶ Our discussion focusses on a number of questions. First, we will consider why complexity matters to the authors. Second, we will look at how complexity is defined in the books. Third, we review how the authors have transferred complexity to evaluation. Finally, we evaluate whether the authors have been able to achieve their books' aims. We conclude with a critical reflection on the future of complexity in evaluation.

Why complexity?

Wolf-Branigin addresses complexity-informed evaluation for social work researchers and evaluators. His point of entry is the experience of complexity in social work regarding social workers as “followers of complexity” (2013: 5) because social workers, by definition, need to understand the complex networks in which needing people are entangled. As such, social workers are said to be intuitively familiar with how complexity operates, yet lack the vocabulary and methodological toolkit to deal with it effectively. The purpose of his book, then, is to provide an introduction to

⁶ Related books have appeared during the same period (Funnell and Rogers, 2011; Morell, 2010; Williams and Hummelbrunner, 2010), but are less explicitly positioned in the complexity sciences; they are not part of our discussion but have been reviewed elsewhere (Hargreaves and Podems, 2012).

complexity, including a series of complexity-friendly methods for social workers, and social work researchers and evaluators.

The volume edited by Forss et al. emerged out of a similar need, namely: the quest for theories and methods that could reflect the social complexity as encountered by evaluators. If the world is complex, they reason, the theories and methods of evaluation should mirror that complexity. For example, they highlight how similar interventions can lead to different outcomes in different instances, which calls for “complex strategies” (2011: 1). Their concerns are more than methodological, as they argue that public policies and accountability fail to recognize this complexity of the public sector. This lack of understanding is fueled by, and motivates, evaluation methods that do not take the complexity into account.

Of the three books discussed here, Patton’s (2011) is the most-cited one. Heavily influenced by Brenda Zimmerman (e.g. Zimmerman, Lindberg & Plsek 1998; Westley, Zimmerman & Patton, 2007), he is very explicit about his reason for using complexity. In the introductory chapters he writes: “[...] *what brings me to complexity is its utility for understanding certain evaluation challenges*. Complexity concepts can be used to identify and frame a set of intervention circumstances that are amenable to a particular situationally appropriate evaluation response, what I am calling here developmental evaluation. This makes dealing with complexity a defining characteristic of developmental evaluation’s niche. Principles for operating in complex adaptive systems inform the practice of developmental evaluation” (2011: 9 original italics).

All three books appear to be rooted in a dual concern: that the complexity of social reality is often ignored, leading to misguided evaluations and policy recommendations, and that the current methodological toolbox is not particularly well-suited to deal with that complexity. The ways in which the authors deal with this dual concern, though, are considerably different, as we will discuss below.

What complexity?

As said, there are multiple points of entry in complexity. Inevitably, any chosen point also discards other options. To Wolf-Branigin, complexity or complexity theory is a new science or paradigm for the social sciences. He follows Page (2011) in discussing various definitions of complexity, ultimately settling on Mitchell’s (2009) definition which focuses on the self-organizing aspects of complex systems, out of which collective behavior emerges. This ties in with the school of thought around emergence, not necessarily a new theme (cf. Juarrero and Rubino, 2010; Vesterby, 2008), but a very relevant and persuasive one still. Emergent behavior, in Wolf-Branigin’s view, is a process that is embedded in complex systems. Complex systems become complex *adaptive* systems when the constituent elements show mutual adaptation.

While they start from the experience of complexity, Forss et al. assert that there is a difference between complexity as an experience and complexity as precise quality of social processes and structures. Still, they somewhat circumvent the issue by giving

multiple definitions of what they would like to focus on. On the one page it concerns a system state somewhere between order and chaos, which resonates with statements from physicists such as Prigogine and Stengers (1984) and management scholars such as Stacey (e.g., 1992). Yet, elsewhere they claim to focus on the non-linear and situated nature of complex systems. These are not mutually exclusive things, but the conceptual back-and-forth makes it a bit harder to grasp where the point of entry lies, even though we fully accept the possibility of multiple starting points.

Patton, then, does something similar and offers a number of core properties upon which his narrative is built, namely: non-linearity, emergence, adaptive behavior, uncertainty, dynamics, and coevolution. As he does not need to accommodate a wide range of authors in his monography, it is easier for him to maintain a consistent narrative, something which sometimes lacks in the edited volume by Forss et al.⁷ Thus, Patton describes rather than defines complexity but it works because the argument is closely tied to the complexity core properties he describes. Later on in the book, he also bolts Holling's adaptive cycle and panarchy (e.g., Gunderson and Holling, 2002) on top of the already rich conceptual library. Although the adaptive cycle helps tying the concepts together, it also increases the conceptual density.

In some ways, there appears little conceptual and theoretical progress since the early iterations of complexity for the social sciences. All authors felt obliged to go through the canon of complexity at some point in their books. While this can be useful to novices, it is also telling that, on the whole, these overviews are nearly identical to books published a while ago. A reference to, e.g., Marion's *The Edge of Organization* (1999) would have been a more efficient way of dealing with this.⁸ Either way, this observation leads us to the question what the progress has been in the two decades since the first books on social complexity were published. The fact that the current books repeat the same canon, without explicitly discussing whether certain concepts and mechanisms have been proven in the social realm (which, incidentally, is not the case), leads one to think that complexity has come to a standstill.

Yet, when it comes to defining complexity, there are subtle differences between these books and the crop of publications from the mid-1990s. One favorable change, in our view, is that the uncritical claims of 'new science' and 'break-through paradigm shifts', so universally presented in those early works, are now much more toned-down.⁹ Wolf-Branigin still does it, but given his well-informed overview of complexity and understanding of existing literature, one suspects that these statements serve, first and foremost, to grab readers' attention and not because he genuinely believes that this is new science. Second, all authors warn against a too generous use of the

⁷ Cf. Patton (2015).

⁸ In fact, there are many possible references here. Marion's is an interesting one because it gives a useful overview of all available mechanisms from the complexity sciences to be used in social scientific research.

⁹ See for example the books by Ralph Stacey. Stacey, in the appendix of his 2012 book, wrote an admirably critical self-reflection, in which he says: "[...] the incorporation of notions from the complexity sciences led to some additional points but did not represent any significant shift in how I was thinking about organizations" (2012: 150).

term 'complexity'. Yes, some situations are complex but others are not, and what works in one context may not work in another context (which, in itself, is a trace of complexity). For example, the second chapter in Forss et al., authored by Rogers, deals with the differences between simple, complicated, and complex. Whereas complicated interventions are said to "[...] have multiple components, which need to be brought together to achieve clearly specified outcomes", complex interventions "[...] are more usefully conceptualized as having complex aspects where they are inherently dynamic and emergent" (2011: 35). As such, whereas complicated interventions can be evaluated by asking "what works for whom in what contexts" (2011: 35), in complex programs "it is not possible to report on these in terms of 'what works' [...] "because what 'it' is constantly changes" (2011: 36).

Such a distinction is less apparent in Wolf-Branigin's book but it definitely underlies his argument. Patton, then, also discusses the distinction between simple, complicated, and complex, but differently so from Rogers' account. He understands this tripartite as a relative scale (2011: 86–90). Complicated situations are characterized by either a high degree of conflict between stakeholders or a high degree of uncertainty and unpredictability about how to solve problems. Complex situations feature both high degrees of social conflict and unpredictability. It is not surprising then that the authors disagree about the ways in which these distinctions are made; Patton, in his review of Forss et al., argues that their book concerns complicatedness instead of complexity (Patton, 2015). This indicates that the narratives are more nuanced than they were twenty years ago, but convergence towards one unambiguous framework is still far off.

How are complexity and evaluation related?

The next question then is: how do the authors relate complexity to evaluation? Unsurprisingly, given the focus of the book on social work and workers, complexity is applied by Wolf-Branigin to conceptualize social service interventions as complex systems (Chapter 2). This requires a complexity-friendly set of evaluation methods: qualitative and quantitative methods need to be blended. Evaluators are told to engage in both inductive and deductive reasoning. To these ends, he offers a suite of methods that could be used for evaluations: agent-based modelling, participatory action research, narrative research, developmental evaluation, social networks, cluster analysis, and spatial analysis. This suite serves to carefully highlight both the metaphorical as well as the mathematical aspects of complexity. We contend that, although some methods are well-suited for evaluation research, the utility of others, such as agent-based modeling, could be contested because there can (and will) be big gaps between the model's outcome and reality. However, we agree with, e.g. Epstein (2008), that there are many reasons why the *process* of modeling, model outcomes aside, will be very informative for ex-ante evaluations. The inclusion of such methods is therefore entirely justifiable.

Less than an overview of what is available and more of a coherent, yet sprawling, offering is Patton's proposal for developmental evaluation. Complexity's

role in evaluation, in his view, is first and foremost to serve as a heuristic and sense-making device. From that understanding follows an evaluation approach that (apparently in contrast to more common approaches) favors dynamics over stability, uncertainty over certainty, equifinality and multi-finality over general laws, etcetera. At heart, developmental evaluation is a dynamic kind of evaluation that does not only seek to identify causal relationships and to serve accountability, but that also offers an approach that interacts with the programs it evaluates, preferably feeding results back into the program on the fly, so as to develop it (hence the name). Given these aims, Patton calls upon evaluators to have a complexity-friendly mindset. The aspects discussed in the book should help building that mindset. Questions, how-to's, checklists, and examples are provided throughout the book, although it would be too far of a stretch to say that it offers a quick and easy guide on how to do developmental evaluation.

The volume by Forrs et al. is the most difficult to assess because, being an edited volume, there are several ways in which complexity and evaluation are understood and discussed. After the introduction, seven case studies are presented by different authors. The studies start from the experience of complexity, i.e. a certain policy or a program is deemed complex because of e.g. conflicting stakeholder interests, which then necessitates an alternative, more complexity-friendly evaluation method. Naturally, these methods differ. Some are built on realist evaluation, others rely on contribution analysis. In some cases, and the interesting case material notwithstanding, we felt that the complexity framework has been added in hindsight or has been developed during the evaluation process. Patton himself remarked that the contributions are not clear about whether the cases are complex or complicated, which is more than just a semantic difference. Marra, in the last chapter, draws attention to the complexity of the evaluation *experience* and *process* instead of the complexity of the case. This is a tell-tale sign of the ambiguity surrounding the authors' understandings of complexity.

Have the books delivered?

Wolf-Branigin intended to provide an agenda-setting, introductory piece about complexity in the field of social work and evaluation research. By implication, it has resulted in a book that portrays what he believes 'is out there' in the complexity sciences – a pocket guide, if you will, for social work researchers and evaluators. For more in-depth readings about the subjects covered in the book, Wolf-Branigin refers to other works that we have also found very useful. To evaluators with prior knowledge of complexity, this book is less appealing as it favors breadth over depth. The trade-off of this overview is that the book does little to extend the existing complexity canon. But then again, that was never the purpose. Laudably, he stresses the importance of using both qualitative and quantitative methods for studying complexity, even though he veers towards computational approaches over qualitative ones. Naturally, actually combining qualitative and quantitative methods is

notoriously hard, and we would have liked to see clear instructions about this for concrete research projects.

Judging by the discussions, online and in journal articles, Patton's developmental evaluation has hit a nerve with parts of the evaluation community. We think that his book features useful reminders of the pitfalls of more common evaluation approaches, and offers avenues of thought for adaptive approaches that acknowledge social complexity. The book is somewhat more antagonistic than Wolf-Branigin's that just demonstrates how current methods can be deployed in complexity-friendly ways. Indeed, Patton offers an *alternative* approach, not just an overview of existing methods, and that necessitates him to contrast it to other approaches. Given the citations of the book, Patton seems to have addressed a pressing issue. A more in-depth look into these citations, however, reveals many citations that do little else than saying things such as 'according to Patton, complexity matters in evaluation.' This is undoubtedly true but it does not do justice to Patton's argument. And that is exactly our critical note: while we agree with the call for a different mindset, we wonder whether Patton's book will help the struggling evaluators to develop comprehensive, complexity-friendly evaluations.

Of the three books, Forss et al. seems to deviate most from its original definitions of complexity given in the introduction. This is inherent to edited volumes with a large number of authors but that is not really an argument for lowering expectations. However, we should also note that the book has multiple goals. The authors also wish to encourage debate and to point at interesting questions for the future of evaluations. This type of goal allows for more variety than a strictly research or theory-driven book. Perhaps the book is more about the match (or lack thereof) between evaluation methods and the object of evaluation. The reasoning is: because the object of evaluation is complex (i.e. changes over time etc.), it challenges evaluation methods that do not account for that complexity. In that sense, they are in the same room as Patton. Ultimately, the authors contend that they struggle with the complexity sciences, which, in their view, may be more a set of ideas than a theoretical framework. They also acknowledge that complexity features a language that is relatively foreign to evaluators and that is difficult to operationalize.

Conclusions and reflections

The books can be assessed by different measures. Do they serve as an introduction to complexity for evaluation purposes? Yes, all three of them give accounts of what complexity is and why it ought to matter to evaluation. The overviews seem to reflect the common canon of complexity. This brings us to the second measure: do the books add anything to the canon? No, while replicating faithfully, they do not add novelties beyond semantic suggestions. We are aware that we are walking a thin line here because one could argue that a 'new' way of understanding evaluation is a valuable addition itself. Our counter-argument is that, if anything, it is an extension of what has already been said in many niches within the social sciences. Do the books account for theory transfer? Yes, to a great extent and with much care. Most authors are quite

aware of how they utilize the concepts. Nevertheless, one should be wary of claims such as, e.g., ‘complex systems all coevolve’, in referring to the mechanism of coevolution between species and their environment as analyzed by e.g. Ehrlich and Raven (1964). What systems are we exactly talking about? Does it concern social systems as well? Has that been proven? If so, what evidence of coevolution do we have? How does reciprocal selection work? Often, such questions are not addressed directly. If proof is offered, it is highly circumstantial. Alternatively, authors (in particular Patton) chicane around such issues by saying that complexity is a sensitizing concept.

We would like to return to the issue touched upon in the first part of this essay: why is it necessary to launch a new discrete branch of evaluation studies around complex systems? This question is particularly pressing given that there is already a long tradition of systems’ theories, and because the methods mentioned by the authors discussed here are not new (cf. Mowles’ 2014 similar argument). One could even argue that complexity does little more than gluing together some existing methods and combining it with ‘a different mindset’ in these books. So, why then risk “becoming a fad” (cf. McKelvey, 1999)? There are a number of possible explanations. It could be that these and other authors are genuinely convinced that there is something completely novel in complexity. If so, this calls for a more thorough comparison with existing approaches that lurk in the same corner of science and a more developed argumentation about the added value, beyond the many somewhat generic statements. One could also argue that scientists and practitioners alike have a stake in launching a new term and claiming it theirs. That can be considered a normal side-effect of competitive science. But if that is the case, it would *still* require authors to work vigorously on generating evidence for their claims.

Following Bartels (1987), we agree that there is an element of randomness or chance in the development of social sciences. Old theories can be rediscovered; disregarded theories can become valuable again. The selection of a particular theory is not necessarily a process that has scientific rigor. However, we feel that it is truly necessary to make the next step. If complexity is to be taken seriously, it will have to move on and start producing concrete results that are markedly better – and we use that term here on purpose, given the large claims – than results produced using ‘standard’ methods. We would like to end, however, with a positive note. The fact that these and other books are being published and are getting attention means that they appeal to the evaluation domain. All three books offer a massive range of challenges. These are perfect starting points for complexity-informed evaluation research, or putting it more strongly: there are no longer reasons to post-pone empirical testing. Hopefully, we will have a whole range of new books with inspiring results to review in a few years from now.

References

- Barnes M, Matka E and Sullivan H (2003) Evidence, understanding and complexity: Evaluation in non-linear systems. *Evaluation* 9(3): 265–84.
- Bartels J (1987) *Kennis, geschiedenis, objectiviteit. Een filosofische reflectie op enkele ontwikkelingen in de wetenschapstheorie* [Knowledge, history, objectivism: a philosophical reflection on certain developments in the theory of sciences]. Groningen: Konstapel.
- Bolinska A (2013) Epistemic representation, informativeness and the aim of faithful representation. *Synthese* 190(2): 219–34.
- Byrne DS (1998) *Complexity Theory and the Social Sciences: An Introduction*. Abingdon: Routledge.
- Byrne DS (2013) Evaluating complex social interventions in a complex world. *Evaluation* 19(3): 217–28.
- Byrne DS and Callaghan G (2014) *Complexity Theory and the Social Sciences: The State of the Art*. London: Routledge.
- Callaghan G (2008) Evaluation and negotiated order: Developing the application of complexity theory. *Evaluation* 14(4): 399–411.
- Castellani B (2009) Map of the complexity sciences. *Art & Science Factory*. Available at: http://www.art-sciencefactory.com/complexity-map_febo9.html.
- Checkland P (1981) *Systems Thinking, Systems Practice*. New York: Wiley.
- Chettiparamb A (2006) Metaphors in complexity theory and planning. *Planning Theory* 5(1): 71–91.
- Colander D and Kupers R (2014) *Complexity and the Art of Public Policy: Solving Society's Problems from the Bottom Up*. Princeton, NJ: Princeton University Press.
- Davies R (2004) Scale, complexity and the representation of theories of change. *Evaluation* 10(1): 101–21.
- Ehrlich PR and Raven PH (1964) Butterflies and plants: A study in coevolution. *Evolution* 18(4): 586–608.
- Epstein JM (2008) Why model? *Journal of Artificial Societies and Social Simulation* 11(4): 12. Available at: <http://jasss.soc.surrey.ac.uk/11/4/12.html>.
- Flood RL (1999) *Rethinking the Fifth Discipline: Learning Within the Unknowable*. London: Routledge.
- Forrester JW (1971) *World Dynamics*. Cambridge: Wright-Allen Press.
- Forss K, Marra M and Schwartz R (eds) (2011) *Evaluating the Complex: Attribution, Contribution, and Beyond*. New Brunswick, NJ: Transaction Publishers.
- François C (1999) Systemics and cybernetics in a historical perspective. *Systems Research and Behavioral Science* 16(3): 203–19.
- Funnell SC and Rogers PJ (2011) *Purposeful Program Theory: Effective Use of Theories of Change and Logic Models*. San Francisco, CA: Jossey-Bass.
- Gerrits LM (2012) *Punching Clouds: An Introduction to the Complexity of Public Decision-Making*. Litchfield Park, AZ: Emergent Publications.

- Gerrits LM and Marks PK (2014) How fitness landscapes help further the social and behavioral sciences. *Emergence: Complexity & Organization* 16(3): 1–17.
- Gunderson LH and Holling CS (eds) (2002) *Panarchy: Understanding Transformations in Human and Natural Systems*. Washington, DC: Island Press.
- Hargreaves MB and Podems D (2012) Advancing systems thinking in evaluation: A review of four publications. *American Journal of Evaluation* 33(3): 462–70.
- Juarrero A and Rubino CA (eds) (2010) *Emergence, Complexity, and Self-organization: Precursors and Prototypes*. Litchfield Park, AZ: Emergent Publications.
- Knuuttila T (2011) Modelling and representing: An artefactual approach to model-based representation. *Studies in History and Philosophy of Science Part A* 42(2): 262–71.
- McKelvey B (1999) Complexity theory in organization science: Seizing the promise or becoming a fad? *Emergence* 1(1): 5–32.
- Marion R (1999) *The Edge of Organization: Chaos and Complexity Theories of Formal Social Systems*. Thousand Oaks, CA: SAGE.
- Mitchell M (2009) *Complexity: A Guided Tour*. Oxford: Oxford University Press.
- Morell JA (2010) *Evaluation in the Face of Uncertainty: Anticipating Surprise and Responding to the Inevitable*. New York: The Guilford Press.
- Mowles C (2014) Complex, but not quite complex enough: The turn to the complexity sciences in evaluation scholarship. *Evaluation* 20(2): 160–75.
- Page SE (2011) *Diversity and Complexity*. Princeton, NJ: Princeton University Press.
- Patton MQ (2011) *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*. New York: The Guilford Press.
- Patton MQ (2015) Book review: Evaluating the complex: attribution, contribution, and beyond. *American Journal of Evaluation* 36: 419–29.
- Phelan SE (1999) A note on the correspondence between complexity and systems theory. *Systemic Practice and Action Research* 12(3): 237–45.
- Prigogine I and Stengers I (1984) *Order Out of Chaos: Man's New Dialogue with Nature*. New York: Bantam.
- Rescher N (1998) *Complexity: A Philosophical Overview*. New Brunswick, NJ: Transaction Publishers.
- Rogers PJ (2008) Using programme theory to evaluate complicated and complex aspects of interventions. *Evaluation* 14(1): 29–48.
- Rogers PJ (2011) Implications of complicated and complex characteristics for key tasks in evaluation. In: Forss K, Marra M and Schwartz R (eds), *Evaluating the Complex: Attribution, Contribution, and Beyond*. New Brunswick, NJ: Transaction Publishers, 33–52.
- Sanderson I (2000) Evaluation in complex policy systems. *Evaluation* 6(4): 433–54.
- Senge PM (1990) *The Fifth Discipline: The Art and Practice of the Learning Organization*. New York: Doubleday Currency.
- Simon HA (1962) The architecture of complexity. *Proceedings of the American Philosophical Society* 106(6): 467–82.

- Stacey RD (1992) *Managing the Unknowable: Strategic Boundaries between Order and Chaos in Organizations*. San Francisco, CA: Jossey-Bass.
- Stacey RD (2012) *Tools and Techniques of Leadership and Management: Meeting the Challenge of Complexity*. London: Routledge.
- Swoyer C (1991) Structural representation and surrogate reasoning. *Synthese* 87(3): 449–508.
- Verweij S (2015) *Once the Shovel Hits the Ground: Evaluating the Management of Complex Implementation Processes of Public-Private Partnership Infrastructure Projects with Qualitative Comparative Analysis*. Rotterdam: Erasmus University Rotterdam.
- Verweij S and Gerrits LM (2013) Understanding and researching complexity with qualitative comparative analysis: Evaluating transportation infrastructure projects. *Evaluation* 19(1): 40–55.
- Vesterby V (2008) *Origins of Self-organization: Emergence and Cause*. Goodyear, AZ: ISCE Publishing.
- Von Bertalanffy L (1968) *General System Theory. Foundations, Development, Applications*. New York: George Braziller.
- Waldrop MM (1992) *Complexity: The Emerging Science at the Edge of Order and Chaos*. New York: Simon & Schuster Paperbacks.
- Walton M (2014) Applying complexity theory: A review to inform evaluation design. *Evaluation and Program Planning* 45: 119–26.
- Westhorp G (2012) Using complexity-consistent theory for evaluating complex systems. *Evaluation* 18(4): 405–20.
- Westley F, Zimmerman B and Patton M (2007) *Getting to Maybe: How the World Is Changed*. Toronto: Vintage Canada
- Williams B and Hummelbrunner R (2010) *Systems Concepts in Action: A Practitioner's Toolkit*. Stanford, CA: Stanford University Press.
- Wolf-Branigin M (2013) *Using Complexity Theory for Research and Program Evaluation*. Oxford: Oxford University Press.
- Zimmerman B, Lindberg C and Plsek P (1998) *Edgework: Lessons from Complexity Science for Health Care Leaders*. London: V H A, Incorporated.