

# Insights and future directions for systems and complexity-informed evaluation

Mat Walton<sup>1</sup>  | Emily F. Gates<sup>2</sup>  | Pablo Vidueira<sup>3,4,5</sup> 

<sup>1</sup> Institute of Environmental Science and Research (ESR), Porirua, New Zealand

<sup>2</sup> Boston College, Chestnut Hill, Massachusetts, USA

<sup>3</sup> Global Alliance for the Future of Food, Washington DC, USA

<sup>4</sup> Comillas Pontifical University, Madrid, Spain

<sup>5</sup> Universidad Politécnica de Madrid, Madrid, Spain

## Correspondence

Mat Walton, ESR, Porirua 50348, New Zealand.

Email: [Mathew.Walton@esr.cri.nz](mailto:Mathew.Walton@esr.cri.nz)

## Abstract

With the increasing maturity of systems thinking and complexity science (STCS) within evaluation, this issue of *NDE* provides case examples of contemporary application. This article identifies themes across case examples to identify emergent patterns and opportunities for the continued development of evaluation practices that draw upon STCS. Each article describes STCS and its applications within specific evaluation contexts. Our review across cases identified three themes: (1) the importance of setting boundaries around what is included and excluded from within the STCS field in a specific evaluation; (2) using STCS concepts and theories to enhance program and evaluation theory, and (3) broadly drawing upon STCS to support changes in systems and develop capacity within organizations. While these implications show some promise for systems and complexity informed evaluation, they, more importantly, underscore the need for deeper engagement with STCS theories and methods while simultaneously remaining accessible for evaluators.

## INTRODUCTION

Systems thinking and complexity sciences (STCS) in evaluation can no longer be considered of as new. For example, it has been 13 years since Williams and Iman edited *Systems Concepts in Evaluation* (Williams & Iman, 2006). It is over 20 years since significant texts considered the application of complexity sciences within social sciences (Byrne, 1998; Cilliers, 1998). There has been a steady number of publications since that provide case examples, reviews, and theoretical development of systems thinking and complexity sciences in areas of social science and evaluation (Gates, 2016; McGill et al., 2020; Morell, 2010; Walton, 2014, 2015, 2016; Williams & Hummelbrunner, 2011; Williams & van 't Hof, 2014). Yet, the knowledge of evaluation practitioners regarding STCS varies in depth and

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2021 The Authors. *New Directions for Evaluation* published by American Evaluation Association and Wiley Periodicals LLC.

sophistication. Walton (2016) interviewed 41 policy and evaluation practitioners and academics, all of who had published work drawing upon STCS. Within this sample, some were experienced across multiple projects and methods; many had limited depth, working on only one or two applications of STCS to date. A more recent study interviewed 30 participants in the UK, all with a role in evaluation, from commissioning through to design and conducting evaluation (Barbrook-Johnson et al., 2020). Participants describe their understanding of complexity and implications for evaluation practice. Findings suggest that most use the term “complexity” informally to mean difficult, challenging, or comprised of multiple factors, rather than conceptions of complexity grounded within STCS literature, as discussed in Chapter 1 (Gates et al., this issue).

## STCS DEVELOPMENTS WITHIN SOCIAL SCIENCE AND PUBLIC MANAGEMENT

Developments in the use of STCS within evaluation can be placed within wider developments in social theory and social research methods, such as trends towards participatory, constructivist, mixed-method, and transdisciplinary approaches creating a permissive context for STCS (Walton, 2016). Recent social science applications of complexity have considered how viewing social phenomena as emergent from complex adaptive social systems fit with other social theory lenses, such as evolutionary theory (Gerrits & Marks, 2017); ideas of hegemony developed by Gramsci (Williams, 2020); and critical realism (Byrne & Callaghan, 2014; Flaherty, 2019). Critical realism has also been used by Mingers and Brocklesby as one of several approaches considering multi-methodology within critical systems approaches (Jackson, 2019). There have been relatively fewer theoretical developments linking STCS and evaluation practices, yet this has still been a fruitful area of development (Eoyang & Holladay, 2013; Foote et al., 2021; Gates, 2018; Patton, 2011; Torres-Cuello et al., 2018).

Of relevance to evaluation, there is also an ongoing debate about how STCS can inform governance design and government services delivery (Haynes, 2015; Ison & Straw, 2020; Room, 2011). From a complexity theory perspective, Hawe and colleagues have written about public health interventions as “events in systems”, which focuses attention on how services or programs work locally even if designed nationally (Hawe, 2015; Hawe et al., 2009). STCS has recently been used to examine the role of social service commissioning and delivery agents, suggesting service ecosystems’ goal as learning systems supporting local relevance and adaptation over time (Lowe & Plimmer, 2019; Oakden et al., 2020). The work of both Oakden and Lowe highlights a challenge in applying STCS. That is the need to learn and engage with at least two fields of knowledge, in these cases complexity sciences and public administration related to social services. Use of STCS within evaluation requires engagement with at least three areas of knowledge, STCS, evaluation, and the topic area(s) of the program or service being evaluated. A multi-methodology approach would also require engagement with different parts within the STCS field. This means that STCS evaluation requires considerable substantive expertise (i.e., evaluation, STCS, and evaluation specific knowledge) and varied methodological skills (i.e., multiple and mixed evaluation as well as STCS methods). This raises the question as to whether such expertise can reasonably be developed within one researcher or evaluator. Miller (2016) makes this point in relation to Developmental Evaluation:

Mastery of any of the major systems thinking traditions that inform DE—complexity theory, soft systems, and critical and complex adaptive systems—or of other schools Patton referenced in early treatments of DE (e.g., cybernetics, system dynamics) is no small feat. Each comes with particular methodology

as well as a unique perspective on how to understand a problem situation and determine its multiple boundaries. Each also brings an ideological stance on which users are important, and each suggests how local users ought to be engaged by system practitioners. (Miller, 2016: pp. 267–268).

The point is that we should be modest in our expectations that someone trained in evaluation will artfully craft innovations in STCS without exposure to the breadth (and depth) of ideas within STCS.

## GUIDES TO SYSTEMS AND COMPLEXITY INFORMED EVALUATION

Given the breadth and depth of STCS fields, numerous introductory guides to using STCS have been developed to support entry into the fields, focused for different purposes. For example, in the U.K., NPC has published a guide to systems change that defines some concepts in systems thinking and presents six principles for planning and doing systems change (Abercrombie et al., 2015). Williams and van 't Hof have produced a workbook to design systems interventions based on systems thinking concepts of inter-relationships, perspectives, and boundaries (Williams & van 't Hof, 2014). In Australia, Davidson and Morgan developed the systems change framework, focused upon public health (Davidson & Morgan, 2018).

More specific to evaluation, recently published in the U.K., the Magenta Book, which provides guidance on evaluation for the public service, was updated with a supplementary guide *Handling Complexity in Policy Evaluation* (Bicket et al., 2020). This guide lays out an argument for why complexity matters and why complexity science-informed evaluation approaches are relevant. Guidance is then provided on designing a “complex” evaluation. The AEA Systems TIG has also provided an introduction and guidance to using STCS in evaluation through four overarching principles, grouped by systems concepts of interrelationships, dynamics, perspectives, and boundaries (Systems in Evaluation Topical Interest Group of the American Evaluation Association, 2018).

Each guide to using STCS provides introductory guidance material, as well as questions that help provide a systems orientation for those designing programs, commissioning evaluation, and conducting evaluations. Each provides pointers to areas where deeper engagement in the field may be required. Each also draws the boundaries of what is STCS in a certain way, for a particular purpose. For someone new to STCS, it may be difficult to critically reflect on the boundaries drawn within these guides.

Several points stand out in taking stock of STCS developments within the social sciences, public management, and evaluation. There is considerable interest and momentum in applying systems and complexity concepts, methods, and approaches. Each application of STCS highlights a diversity of approaches and the necessity of boundaries placed around any application. The interaction with context is also highlighted, with the roles of program design, implementation, and evaluation commissioning included within guidance documents. In our view, STCS informed evaluation extends beyond the capacity of any one practitioner following a guide or set of principles. It is found in mixed roles teams, blending expertise related to evaluation, systems and complexity, the subject area of the evaluand, and other domains.

Given the increasing maturity of application of systems thinking and complexity within evaluation, this issue of *NDE* sought to collate a series of case examples of contemporary application. This final article provides a description of themes identified by the co-editors from across ten case examples. We close the article by considering the implications of themes for further advancing the application of STCS in evaluation. Our aim is to provide a modest contribution to depth in practice and support further innovation.

## METHODS

The authors, as issue co-editors, purposefully sought out articles from co-author groups that conducted formal case studies of their practices as well as practice-based case examples. Article authors were provided guidance and prompts in writing to aid comparison across articles. In conceptualising articles as cases, we drew upon the “complex realism” approach of David Byrne (Byrne, 2013; Byrne & Callaghan, 2014). Byrne in turn, draws upon Charles Ragin’s use of cases, which require a process of “casing” where boundaries and context of the case are described (Byrne & Ragin, 2009; Ragin & Amoroso, 2011). Comparison between cases, in a similar vein to realist evaluation approaches, requires consideration of cases as whole systems, where a comparison is of the whole context, mechanism, outcome configuration in each case.

To facilitate this approach to cases and case-comparison, article co-authors were asked to consider the following

- Cite literature to locate article within research and theory context
- Identify key systems/complexity traditions, theories, methods, perspectives, assumptions, and so forth that the article (and evaluation practice) draws from
- Provide context of evaluation practice, locate the context of case example
- Share critical reflections on how taking a STCS evaluation fits within or challenges your boundaries around evaluation
- Comment on utility or value of STCS approaches to evaluation

Editor comments on draft papers sought to clarify which concepts, theories, methods, or approaches co-authors drew on from STCS and why, evaluation and program context of each article, as well as identify critical reflections. All three co-editors reviewed each article and provided feedback to co-authors. To generate themes across articles, co-editors each shared their reflections on themes across articles. Initial themes were then revised through regular discussions during the process of bringing together the journal issue. The identified themes are limited to the articles within this *NDE* issue, which in turn are limited to word count restrictions of the journal. We do not suggest the themes represent a “state of play” regarding STCS use within evaluation. Rather, we intend that the themes and discussion provide ideas to support reflection on the use of STCS to encourage debate and development within the field.

## FINDINGS

We identified three themes across the ten articles provided in this issue. The themes each speak to the practice of applying STCS within the evaluation. Reflection on the themes identifies opportunities for further refinement and innovation in STCS evaluation practices.

### Theme 1: Setting boundaries and positioning within STCS

In each article, authors have defined what they mean by STCS to set up a discussion of its application within their case examples. Within some articles, this was explicit, defining STCS concepts used in description and reflection on evaluation practice (Richardson and Patton, Chapter 10; Vang, Moore & Nicklin, Chapter 7; Wilkins, Kossover-Smith, Hogan, Espinosa & Wilson, Chapter 5; Bustamante, Vidueira & Baker, Chapter 6). In

other chapters, definitions are more implicit, often citing instead particular methodology within which boundaries of STCS can be located (Reeders & Brown, Chapter 2; Hepi, Foote, Ahuriri-Driscoll, Rogers-Koroheke, Taimona & Clark Chapter 4). Whether implicit or explicit, authors set boundaries within the breadth and variety of STCS traditions from which specific ideas have been selected to apply within their evaluation contexts. Considering how boundaries of STCS were set revealed two ideas we explore further conceptual versus methodological applications and aligning STCS use to evaluation purpose and context.

## Conceptual versus methodological applications

Across the articles, there appears to be a difference between authors who use STCS as sensitizing concepts and those that use STCS as methodology. We interpret the implications of this difference to suggest the value of moving beyond a conceptual application of STCS to more deeply applying theory, methodologies, or approaches.

As sensitizing concepts, articles such as Richardson and Patton (Chapter 10) and Hargreaves (Chapter 8) illustrate the use of STCS principles to reflect upon how evaluation practice seeks to understand perspectives, consider boundaries, and support networks as part of program delivery. The use of STCS as sensitizing concepts in these examples is at a high level, rather than a more detailed critical reflection to support innovation at the level of evaluation design and method. The article by Bustamante, Vidueira, and Baker (Chapter 6) developed a framework that could be used to critically reflect on evaluation by highlighting which STCS concepts were and were not included within an evaluation framework. Bustamante and colleagues describe going beyond the four STCS principles of boundaries, interrelationships, perspectives, and dynamics. To support critique, the authors identified a need for specificity not present at a principles level, as well as adding concepts regarding system dynamics from complexity sciences.

In contrast, Reeders and Brown (Chapter 2) and Hepi, Foote, Ahuriri-Driscoll, Rogers-Koroheke, Taimona, and Clark (Chapter 4) both apply established systems approaches (group model building/causal loop diagramming and boundary critique, respectively). Here there was not a requirement to explicitly define key concepts of STCS because the methodology already has a lineage within STCS. Both articles were able to use the strengths and weaknesses of the chosen method, including underpinning theory, to reflect upon the use of method and refinements required to match specific needs. Such a reflection supported methodological innovation and appeared to support critical reflection on the boundaries of STCS to a greater extent than the use of STCS as sensitizing concepts.

We consider these two approaches to defining boundaries of STCS, as broader sensitizing concepts and as particular methodologies or approaches, each support different types of innovation. As sensitizing concepts, reflection on the use of STCS appeared to support innovation in considering the role and approach of evaluation more broadly within programs and systems change. Reflection on particular STCS methodologies appeared to support innovation in specifics of program theory and of evaluation methods and methodology.

## Aligning STCS use to evaluation purpose and context

Several articles highlighted the matching of STCS applications to purpose and context. To understand the appropriateness of boundaries placed around STCS within the evaluation, consideration of the context of application and purpose is important. For example,

Richardson and Patton (Chapter 10) describe the creation of a new inter-organizational coalition focused on a global change to food systems. There are few existing institutional arrangements that need to be accommodated, and likely a less rigid and more inclusive conceptualization of STCS suits the adaptive and dynamic way of working. Not being bound by specific methodological constraints, but able to move within a general STCS orientation. In contrast, Wilkins et al. (Chapter 5) describes efforts to introduce STCS practice within a range of existing institutional structures, utilizing existing monitoring/reporting structures. Here the focus is not on systems change but movement in practices, with the boundaries of STCS drawn to fit the circumstances. Matching boundaries of STCS practice with institutional boundaries appears an important part of a fit for purpose. Supporting this match of boundaries requires knowledge on two fronts: first, some depth of knowledge about a range of STCS approaches that could be utilized; and second understanding of the context. Of course, there are a range of STCS approaches for understanding a problem situation, suggesting that STCS methods can be used to support this process of matching. However, articles in this issue did not explicitly discuss how or why they selected and aligned specific STCS approaches to particular contexts.

## **Theme 2: Theory scaffolding—using STCS to enhance program and evaluation theory**

A second theme across multiple articles considers the relationship between theory(ies) within systems and complexity informed evaluation. Westhorp (2012), in discussing the use of both realist and complexity theory within the evaluation, talks about theory scaffolding. She suggests complexity theory can be used as a base layer by considering how a change in systems occurs over time. Additional layers of theory related to the topic under study get put on top. Within such a realist framework, program theory, ideas of how an intervention will lead to change should be guided by both broader theories of how systems change over time and more specific topic theories (Shearn et al., 2017). Donaldson and Lipsey (2006) make the distinction between social theory, program theory and evaluation theory. They describe evaluation theory as ideas about what evaluation should be and how it should be conducted. Across the articles, in this issue we identify ways that authors have sought to integrate STCS with both program theory and evaluation theory through two sub-themes. The first theme relates to the blending of theory embedded within STCS approaches with topics to consider program theories. The second theme relates to building evaluation theory by combining STCS with other methodologies within evaluation.

### **Program theory**

Two articles describe the use of STCS approaches to articulate program theory that combines activities within a program with the wider contexts within which the program is implemented. The article by Reed, Harwood, Jackson, and Gebreyesus (Chapter 3) used ideas from systems thinking in articulating program theory for the U.S. Department of Defense capacity building programs. Specifically, ideas of boundaries were used to identify interacting levels (individual, organizational, institutional, environment/community). Program activities were identified at each level, as well as monitoring and evaluation indicators. The task of identifying boundaries was supported by ideas of boundary critique from Midgley and use of Critical System Heuristics.



Reeders and Brown (Chapter 2), describe using theories specific to the field within which the evaluand is located to help translation from causal loop diagrams to program theory, which was then used to inform a monitoring and evaluation framework. Focused upon peer and community led programs in HIV and viral hepatitis prevention, Reeders and Brown considered existing program theory too heavily focused upon individuals informed by social psychological theories. To widen the perspective out from individuals, Reeders and Brown conducted group model building processes with those involved in delivering peer programs to identify program theory. Like Reed and colleagues (Chapter 3), the program theory articulated through causal loop diagrams highlights the interaction between elements of the program and the context within which the program is delivered. However, Reeders and Brown needed to go beyond the method of Causal Loop Diagrams to articulate program theory, combining insights from the diagram with other theory to generate middle range theory.

Articulating program theory for achieving systems change can be seen in several articles. A common theme is combining ideas from STCS with theory within the topic area relevant to what intervention or change processes are being evaluated. For example, Gates and Fils-Aime (Chapter 9) describe ReThink Health, where ideas of change in systems are combined with ideas of stewardship and promotion of health and wellbeing. Richardson and Patton (Chapter 10) combine systems thinking principles with knowledge and theory within agri-food systems to describe the approach to systems change within the Global Alliance for the Future of Food. Norman (Chapter 11) describes combining methods from design thinking with STCS to articulate processes to support systems change. We suggest two interpretations of systems change program theory from these three articles (Chapters 9, 10, and 11). First, that STCS provides a framework for considering systems change, but details within this framework need to be context specific, and therefore supplied by theories within the substantive field of focus. The second interpretation is that in fact the theory of how systems change happen within STCS requires further development. For example, the three articles do not appear to draw upon ideas from complexity science to theorize change in systems over time. Even within systems thinking traditions, different approaches will give different emphasis to system structures compared to the exercise of power to make decisions and allocate resources (Jackson, 2019).

## Evaluation theory

Several articles make suggestions for what evaluation should be about and how it should be conducted. A theme related to discussion of systems change is the idea that evaluation can support systems change by supporting ongoing learning and adaptation cycles. This is evaluation as part of the program theory for achieving systems change. Of course, the idea of evaluation as an ongoing activity to support adaptation has been well discussed in recent times through, for example, adaptive action (Eoyang & Holladay, 2013) and developmental evaluation (Patton, 2011). Interestingly, within the articles by Vang, More and Nicklin (Chapter 7) and Wilkins, Kossover-Smith, Hogan, Espinosa and Wilson (Chapter 5), we see attempts to support use of STCS within evaluation. Of course, guidance on what STCS evaluation practice should be is dependent on the boundaries that define STCS (see theme one).

The article by Hepi and colleagues (Chapter 4), provides a different example with implications for evaluation theory. Drawing upon ideas in critical systems and kaupapa Māori methodology, also a critical theory, the article suggests that an important role for STCS informed evaluation is to identify power and conflict within program theory and defining

an evaluand. Beyond this, Hepi and colleagues describe processes for giving more power to indigenous partners and participants of the program. Linked to ideas of systems change, the power to challenge and propose a different mental model is often described as having most influence on the system as a whole (Meadows & Wright, 2008). Reeders and Brown (Chapter 2) also seek to give power to those involved within peer programs in articulating program theory.

### **Theme 3: Using STCS and evaluation to support systems change**

The concept of systems change has appeared within the preceding two themes. Across the articles within this issue, ideas of system change featured so heavily that we decided it was worth pulling together ideas presented so far and reflecting further. In fact, taking the articles in this issue as our sample, we ask is evaluation from an STCS perspective synonymous with systems change?

An explicit link between systems change and evaluation is made within Chapters 9, 10, and 11. Gates and Fils-Aime (Chapter 9) and Richardson and Patton (Chapter 10), both describe organizations with an explicit purpose of supporting systems change in areas of health and food systems respectively. Both draw upon ideas from systems thinking and both highlight evaluation within the program theory of systems change. As already noted, aspects of STCS are drawn upon within the program theory of systems change, while other aspects are not utilized. An explicit example of the selective use of STCS was shown by Bustamante, Vidueira, and Baker (Chapter 6), which highlighted that The Economics of Ecosystems and Biodiversity for Agriculture and Food Evaluation Framework (TEEBAgriFood) evaluation framework included numerous STCS concepts, but excluded several principles derived from complexity sciences that consider dynamics of systems over time. Potentially such principles could support program theory for achieving systems change.

Another type of systems change is suggested within the articles by Wilkins and colleagues (Chapter 5) and Vang, Moore and Nicklin (Chapter 7). While not described as systems change, they both describe attempts at increasing use of STCS within evaluation practice, changing the system of evaluation to an STCS orientation. Both articles describe incremental changes utilizing existing structures and formalizing consideration of STCS within these structures. Hargreaves (Chapter 8) suggests something similar by offering bricolage as a mixed design framework to create space for STCS, whilst side-stepping what she views as entrenched paradigm conflict in applied STCS evaluation within the US context.

As discussed previously, Hepi and colleagues (Chapter 4), Reeders and Brown (Chapter 2) and Reed and colleagues (Chapter 3) all utilize STCS in articulating program theory and defining evaluands. Whilst these authors do not describe their work as systems change, we could view their efforts of challenging and broadening the boundaries of programs as seeking a different mental model underpinning the programs. Something that Gates and Fils-Aime (Chapter 9) identify as an explicit aim of the ReThink Health initiative in their systems change efforts.

Across all articles, what can be seen is a role for evaluation in seeking to change systems through: broader and equity focused mental models underpinning programs; organizational structures that may shift systems over time; supporting networks and partnerships; and, evaluation as an ongoing activity that supports learning and adaptation. What isn't seen, at least within space limits provided by articles, is articulation of systems change that across this full breadth of the roles of evaluation, grounded deeply in STCS theory. We consider this an area for further development that we are likely to see in the near future as strands of practice and theory are brought together.



## DISCUSSION

An aim of this *NDE* issue was to highlight diverse contemporary uses of STCS within evaluation and consider what future developments in the field may look like. Across the articles, we see a range of applications and have identified three broad themes: (1) setting boundaries and positioning within STCS; (2) theory scaffolding, using STCS to enhance program and evaluation theory; (3) using STCS and evaluation to support systems change. Across the themes we see the diversity within the STCS fields, and diversity in application within evaluation. In practice, boundaries need to be placed around what is being referred to as STCS. Such boundaries should of course be relevant for the job at hand. One kind of boundary is to utilize a specific STCS methodology (for example see Hepi et al., Chapter 4). Another kind of boundary is to use principles derived widely from across the STCS field (for example see Richardson and Patton, Chapter 10). Such principles offer wide scope for practice, but little specificity to judge quality or guide choice in methods. Bustamante, Vidueira and Baker (Chapter 6) provide a framework from which boundaries of STCS can be tested, and we see this as a useful innovation for evaluation practitioners and commissioners alike. Another limitation of wide and general principles focused STCS boundaries is lack of engagement with the theory underpinning STCS approaches. Such theory could usefully contribute to both program theory and evaluation theory. For example, by drawing upon ideas of complexity science to consider how systems remain stable or change over time, we could identify ways that programs may operate, as well as criteria against which we could judge the ongoing impact of programs (Signal et al., 2013). Currently such ideas from complexity science sit outside of systems thinking guidance such as AEA Systems TIG principles (Systems in Evaluation Topical Interest Group of the American Evaluation Association, 2018).

Drawing upon both our personal experiences, and the set of examples within this issue we propose a modestly radical agenda of moving away from systems thinking and evaluation and their presumed relationship of importing systems thinking *into* evaluation. Let us unpack what we mean by this.

First, we believe that in order to move towards greater use and innovation of systems and complexity within evaluation, it is necessary to expand focus well beyond systems thinking and systems concepts. Systems thinking is a summary term, suggesting a systemic rather than systematic orientation and a term that is inclusive of the many and varied approaches within hard, soft and critical systems approaches (Ison, 2010; Jackson, 2019). We contend that the term systems thinking, and principles of systems thinking, is a useful doorway into a systemic orientation. Such a doorway is needed in our opinion. Systemic approaches to evaluation are not the norm and systems thinking and systems principles provide a useful entry point. However, once through the door, the hard work requires deeper engagement with more detailed methods and associated theory underpinning different systems and complexity traditions and approaches. Such deeper engagement will support theory scaffolding, building upon understanding of systems with topic specific theory. Deeper engagement will also support the refinement of evaluation theory associated with STCS use within evaluation. For example, whilst understanding multiple perspectives is important, how do we navigate between perspectives to inform core evaluation business such as valuing? Hepi et al. (Chapter 4) provide us an example where evaluation activity explicitly grounded in critical systems theory and indigenous values led the research team to articulate and justify the value and power bases on which their evaluative processes and claims were grounded. In our view, this signifies a depth of engagement with a systems theory and methodology and an integration of this into the core evaluative process. The result is a refinement of evaluation theory of valuing more broadly to incorporate considerations of power.

Second, drawing on our observations of trends in the field and case examples featured in this issue, we also call for moving beyond evaluation as a bounded inquiry defined by judging the value of a particular intervention or process. Evaluation as a field and professional practice has long extended beyond the systematic assessment of policies, programs and other bounded interventions, yet this way of framing evaluation persists. In the context of systems and complexity informed evaluation, this frame too narrowly bounds what STCS is applied to. Given the origins and development of systems and complexity traditions and approaches as discussed in Chapter 1 (Gates et al., this issue), the power and value of STCS cover a range of tasks, including identifying, framing and structuring problems or opportunities for change, crafting mutual understanding of problems and needs for change; constructing models to inform strategy and change efforts; learning and leveraging learning to inform action and adaptation; and, throughout all these tasks, examining and deliberating about the value of what was done and should be done next. In other words, STCS informed evaluation should move beyond the implicit linear assumptions about evaluation at the end of an intervention design, implementation and evaluation cycle. Such a shift in the role of evaluation is called for within developmental and blue marble evaluation (Patton, 2011, 2019), and other approaches to systems change (Abercrombie et al., 2018; Lowe & Plimmer, 2019). To help us consider the roles for both STCS and evaluative thinking in supporting both program theory and evaluation theory, we suggest systems thinking, STCS principles, and in depth theory and method of specific STCS approaches can all be useful. Each provides a gaze illuminating areas where STCS can support service and systems design, implementation and evaluation.

The third and final implication of identified themes is to emphasize the importance of working in teams. If use of STCS and evaluation within programs and systems change initiatives is going to increase in frequency and continue to develop as a field, engagement is required across a breadth of STCS approaches, with evaluation theory and method and with the substantive topic area of focus. This requires a team. It also requires processes that create spaces for sharing power between perspectives and for creativity (Innes & Booher, 2010). A team can provide a link between those engaged in theory and those engaged with on the ground doing (whether service delivery, advocacy, philanthropic funding or other roles). Sanderson (2006) suggested complexity informed evaluation practice should draw upon pragmatist philosophy to mix theory with practice, bringing together knowledge of context with knowledge of what has worked elsewhere and theories of how things might work. Bringing together case examples, with explicit linking to underpinning theory and method, was a purposeful approach within this issue to support working evaluators to engage with some depth of STCS informed evaluation.

## CONCLUSION

In closing, this issue sought to provide case-based evidence to reflect on contemporary use of systems thinking and complexity science for evaluation theory and practice. Bringing together ten cases as articles, provides a modest contribution within a growing area of evaluation activity. Even so, the examples in this issue illustrate a range of applications of conceptual and methodological contributions from the STCS fields to various evaluative tasks, including program theory, developing indicators for monitoring, framing the evaluation scope, building organizational capacity, supporting systems change, conducting meta evaluations among others. Across the cases, co-authors drew on aspects of research within the STCS and evaluation fields as well as substantive, methodological, and context-specific knowledge relevant to the case and circumstances at hand. This tremendous variation

suggests that the answer to the question of what difference STCS makes for evaluation is—it depends. It depends on the following:

- What blend of concepts, methods, theories, and approaches are being drawn on;
- From where (or which traditions) within the STCS fields;
- For what evaluative purposes and tasks;
- How STCS and evaluative processes are brought together;
- By whom and with what team expertise and capacities;
- In what institutional and contextual circumstances;
- And why or to what end?

This variation leaves those of us interested in STCS with the challenge to keep learning from both fields (STCS and evaluation), innovating through their interaction, and adapting them to the substantive areas and contexts in which we work. For researchers studying and practitioners sharing examples, it may be important to break from the presumption of applying STCS generally to evaluation, to more deeply engaging with specific theories and methods within STCS and leveraging them for specific evaluation tasks to realize the depth of potential systems and complexity informed evaluation.

## ORCID

Mat Walton  <https://orcid.org/0000-0002-0680-7214>

Emily F Gates  <https://orcid.org/0000-0003-2352-5064>

Pablo Vidueira  <https://orcid.org/0000-0003-0348-8825>

## REFERENCES

- Abercrombie, R., Boswell, K., & Thomasoo, R. (2018). *Thinking Big: How to use theory of change for systems change*. <https://www.thinknpc.org/resource-hub/thinking-big-how-to-use-theory-of-change-for-systems-change/>
- Abercrombie, R., Harries, E., & Wharton, R. (2015). *Systems Change: A guide to what it is and how to do it*. <https://www.thinknpc.org/resource-hub/systems-change-a-guide-to-what-it-is-and-how-to-do-it/>
- Barbrook-Johnson, P., Proctor, A., Giorgi, S., & Phillipson, J. (2020). How do policy evaluators understand complexity? *Evaluation*, 26(3), 315–332. <https://doi.org/10.1177/1356389020930053>
- Bicket, M., Christie, I., Gilbert, N., Hills, D., Penn, A., & Wilkinson, H. (2020). *Magenta book 2020 supplementary guide: Handling complexity in policy evaluation*. CECAN
- Byrne, D. (1998). *Complexity theory and the social sciences: An introduction*. Routledge.
- Byrne, D. (2013). Evaluating complex social interventions in a complex world. *Evaluation*, 19(3), 217–228. <https://doi.org/10.1177/1356389013495617>
- Byrne, D., & Callaghan, G. (2014). *Complexity theory and the social sciences: The state of the art*. Routledge.
- Byrne, D., & Ragin, C. C. (Eds.). (2009). *The SAGE handbook of case-based methods*. SAGE.
- Cilliers, P. (1998). *Complexity and postmodernism: Understanding complex systems*. Routledge.
- Davidson, S., & Morgan, M. (2018). *Systems change framework*. <https://preventioncentre.org.au/wp-content/uploads/2015/01/Systems-Change-Overview-w-Practices.pdf>
- Donaldson, S., & Lipsey, M. (2006). Roles for theory in contemporary evaluation practice: Developing practical knowledge. In I. F. Shaw, J. C. Greene, & M. M. Mark (Eds.), *The SAGE handbook of evaluation* (pp. 57–75). SAGE.
- Eoyang, G. H., & Holladay, R. J. (2013). *Adaptive action: Leveraging uncertainty in your organization*. Stanford University Press.
- Flaherty, E. (2019). *Complexity and resilience in the social and ecological sciences*. Palgrave MacMillan.
- Foote, J., Midgley, G., Ahuriri-Driscoll, A., Hepi, M., & Earl-Goulet, J. (2021). Systemic evaluation of community environmental management programmes. *European Journal of Operational Research*, 288(1), 207–224. <https://doi.org/10.1016/j.ejor.2020.05.019>
- Gates, E. F. (2016). Making sense of the emerging conversation in evaluation about systems thinking and complexity science. *Evaluation and Program Planning*, 59, 62–73. <https://doi.org/10.1016/j.evalprogplan.2016.08.004>
- Gates, E. F. (2018). Toward valuing with critical systems heuristics. *American Journal of Evaluation*, 39(2), 201–220. <https://doi.org/10.1177/1098214017703703>

- Gerrits, L., & Marks, P. (2017). *Understanding collective decision making a fitness landscape model approach*. Edward Elgar.
- Hawe, P. (2015). Lessons from complex interventions to improve health. *Annual Review of Public Health*, 36(1), 307–323. <https://doi.org/10.1146/annurev-publhealth-031912-114421>
- Hawe, P., Alan, S., & Riley, T. (2009). Theorising interventions as events in systems. *American Journal Of Community Psychology*, 43, 267–276.
- Haynes, P. (2015). *Managing complexity in the public services* (2nd ed.). Taylor & Francis.
- Innes, J. E., & Booher, D. E. (2010). *Planning with complexity: An introduction to collaborative rationality for policymaking*. Routledge.
- Ison, R. (2010). *Systems practice: How to act in a climate change world*. The Open University.
- Ison, R., & Straw, E. (2020). *The hidden power of systems thinking: Governance in a climate emergency*. Routledge.
- Jackson, M. C. (2019). *Critical systems thinking and the management of complexity*. Wiley.
- Lowe, T., & Plimmer, D. (2019). *Exploring the new world: Practical insights for funding, commissioning and managing in complexity*. <https://collaboratecic.com/exploring-the-new-world-practical-insights-for-funding-commissioning-and-managing-in-complexity-20a0c53b89aa>
- McGill, E., Marks, D., Er, V., Penney, T., Petticrew, M., & Egan, M. (2020). Qualitative process evaluation from a complex systems perspective: A systematic review and framework for public health evaluators. *PLoS Medicine*, 17(11), e1003368. <https://doi.org/10.1371/journal.pmed.1003368>
- Meadows, D. H., & Wright, D. E. (2008). *Thinking in systems: A primer*. Chelsea Green.
- Miller, R. L. (2016). On messes, systems thinking, and evaluation: A response to patton. *American Journal of Evaluation*, 37(2), 266–269. <https://doi.org/10.1177/1098214015626294>
- Morell, J. A. (2010). *Evaluation in the face of uncertainty: Anticipating surprise and responding to the inevitable*. Guilford Press.
- Oakden, J., Walton, M., & Foote, J. (2020). Contracting public health and social services: Insights from complexity theory for Aotearoa New Zealand. *Kōtuitui: New Zealand Journal of Social Sciences Online*, 16(1), 180–195. <https://doi.org/10.1080/1177083X.2020.1822422>
- Patton, M. Q. (2011). *Developmental evaluation: Applying complexity concepts to enhance innovation and use*. The Guilford Press.
- Patton, M. Q. (2019). *Blue marble evaluation*. Guilford Press.
- Ragin, C. C., & Amoroso, L. M. (2011). *Constructing social research: The unity and diversity of method*. Sage Pine Forge Press.
- Room, G. (2011). *Complexity, institutions and public policy*. Edward Elgar Publishing.
- Sanderson, I. (2006). Complexity, “practical rationality” and evidence-based policy making. *Policy and Politics*, 34(1), 115–132.
- Shearn, K., Allmark, P., Piercy, H., & Hirst, J. (2017). Building realist program theory for large complex and messy interventions. *International Journal of Qualitative Methods*, 16(1), 1609406917741796. <https://doi.org/10.1177/1609406917741796>
- Signal, L., Walton, M., Ni Mhurchu, C., Maddison, R., Bowers, S., Carter, K., Gorton, D., Heta, C., Lanumata, T. S., Mc Kerchar, C. W., O’Dea, D., & Pearce, J. (2013). Tackling “wicked” health promotion problems: A New Zealand case study. *Health Promotion International*, 28(1), 84–94. <https://doi.org/10.1093/heapro/das006>
- Systems in Evaluation Topical Interest Group of the American Evaluation Association. (2018). *Principles for effective use of systems thinking in evaluation practice*. <https://www.systemsinevaluation.com/resources>
- Torres-Cuello, M. A., Pinzón-Salcedo, L., & Midgley, G. (2018). Developing a systemic program evaluation methodology: A critical systems perspective. *Systems Research and Behavioral Science*, 35(5), 538–547. <https://doi.org/10.1002/sres.2561>
- Walton, M. (2014). Applying complexity theory: A review to inform evaluation design. *Evaluation and Program Planning*, 45(0), 119–126. <https://doi.org/10.1016/j.evalprogplan.2014.04.002>
- Walton, M. (2015). Setting the context for using complexity theory in evaluation: Boundaries, governance and utilisation. *Evidence and Policy*, 12(1), 73–89. <https://doi.org/10.1332/174426415x14298726247211>
- Walton, M. (2016). Expert views on applying complexity theory in evaluation: Opportunities and barriers. *Evaluation*, 22(4), 410–423. <https://doi.org/10.1177/1356389016667890>
- Westhorp, G. (2012). Using complexity-consistent theory for evaluating complex systems. *Evaluation*, 18(4), 405–420. <https://doi.org/10.1177/1356389012460963>
- Williams, A. (2020). *Political hegemony and social complexity*. Palgrave Macmillan.
- Williams, B., & Hummelbrunner, R. (2011). *Systems concepts in action: A practitioner’s toolkit*. Stanford University Press.
- Williams, B., & Iman, I. (Eds.). (2006). *Systems concepts in evaluation: An expert anthology*. American Evaluation Association.
- Williams, B., & van ’t Hof, S. (2016). *Wicked solutions: A systems approach to complex problems*, 2nd Ed. <https://gumroad.com/l/wicked>

## AUTHOR BIOGRAPHIES

**Mat Walton**, PhD, is the Technical Lead in the Social Systems Team at the Institute of Environmental Science and Research (ESR) in New Zealand, which specializes in the application of systems thinking to social science research and evaluation in the areas of environmental health, public health, and social policy.

**Emily F. Gates**, PhD, an assistant professor at Boston College, conducts evaluations of educational and health interventions and researches systems thinking, values and equity in evaluation.

**Pablo Vidueira**, PhD, is an evaluation consultant, researcher, and professor affiliated with the Global Alliance for the Future of Food, the Comillas Pontifical University, and the Technical University of Madrid.

**How to cite this article:** Walton M., Gates, E. F., & Vidueira, P. (2021). Insights and future directions for systems and complexity-informed evaluation. In E. F. Gates, M. Walton, & P. Vidueira (Eds.), *Systems and Complexity-Informed Evaluation: Insights from Practice*. *New Directions for Evaluation*, 2021, 159–171.  
<https://doi.org/10.1002/ev.20459>