SOCIAL INNOVATIONS Journal

Volume 5: 04 March 2021
https://socialinnovationsjournal.com/index.php/sij/article/view/704

Evaluating Outside the Box: Evaluation's Transformational Potential

Scott Chaplowe[1] and Adam Hejnowicz[2]
[1]Evaluation and Strategy Specialist, International Evaluation Academy (IEA), Transformational Evaluation Working Group Coordinator
[2]Research Associate, School of Engineering, Newcastle University, Visiting Research Associate, Department of Biology, University of York

Corresponding author: Scott Chaplowe, Scott@ScottChaplowe.com

Abstract

The call for transformation is a response to the dire global emergency; it is a call for radical innovation at multiple levels if humanity is to survive into the next Century. How can evaluation, a profession in the business of assessment and advising, inform and hasten transformation? As a field that straddles both theory and practice, evaluation is uniquely positioned to support transformational learning and change, but this potential depends on its ability to transform from within. This article identifies four interrelated "boxes" that confine evaluation's transformational potential: a project fixation, a short-term temporal fixation, a quantitative fixation, and an accountability fixation. It also examines the uptake and influence of complex systems analysis in the field of evaluation as a means to "breakout" of these boxes and nudge evaluation towards the inner transformation required for it to contribute to transformational change.

Keywords: evaluation, transformational change, project fixation, quantitative fixation, accountability fixation

Introduction

> "We commit to evaluations that help us learn, understand and support the transformational and systemic changes needed in our countries and the world, as agreed upon in the 2030 Agenda for Sustainable Development. A sustainable balance between the

> social, economic and environmental domains is crucial in light of the existential threats of the climate crisis, mass extinction of species, growing local and global inequity, and ultimately unsustainable use of the resources of the planet," (Prague Declaration on Evaluation for Transformational Change, IDEAS 2019: 1).

Today, each of us is a globally connected witness to a time of profound and inter-related challenges facing humanity (Steffen et al. 2018). The COVID-19 experience provides a clear lens on many of the frailties and failures of the planetary system (Andersen and Rockström 2020). We have reached a global tipping point, where the deteriorating environment is increasingly unable sustain to life as we know it. Although these changes, like climate change, may appear slow relative to the spread of a pandemic they are rapid in earth time, and we face a "near-term collapse" in society if we continue under a business-as-usual (BAU) scenario (Bendell 2018).

The call for transformation is a response to the dire global emergency of our Anthropocene Age. It is a call for radical innovation and progress at multiple levels in society if it is to survive. This call has not gone unheeded in the field of evaluation, as reflected in the opening quote from the Prague Declaration on Evaluation for Transformational Change adopted at the Third International Conference of the International Development Evaluation Association and echoed elsewhere by prominent evaluators (Ofir 2020; Parsons et al. 2020; Patton 2020; Picciotto 2020).

In this article we explore the potential role evaluation can play in transforming society. From education and health care to policing and international development, evaluation has become so ubiquitous in our lives that we often take it for granted (Dahler-Larsen 2011, 3). Given this prominent role, we ask how can evaluation, a profession in the business of assessment and advising, inform and hasten transformation? As a field that straddles both theory and practice, evaluation is uniquely positioned to support transformational learning and change, but as we shall see, this potential depends on its ability to transform from within.

We will first examine the influence of complex systems analysis on evaluation's transformational potential. We will then explore some of the barriers that "box" in evaluation's transformational potential and their potential fixes, and we conclude with some promising trends to monitor. We acknowledge that *transformation* is an unfolding, open-ended process; therefore, we approach it heuristically rather than dogmatically. We also acknowledge and distinguish our use of transformation in evaluation from *transformative evaluation*, an established theoretical framework developed by Mertens (2009) for conducting research and evaluation to support social justice. While there is certainly overlap in principles and theory, we use transformation more broadly to refer to deep, rapid, and radical global systems change, often contrasted with incrementalism, reform, or transition, to convey the magnitude of required change:

> "Unlike a 'transition,' which implies moving from one place or state to another, 'transformation' is more about completely

> reinventing shape or form – like the metamorphosis of a caterpillar to a butterfly," (Waddock et al. 2020, 4).

Cracks in the Box

> "There is a crack in everything. That's how the light gets in."
> (Leonard Cohen, 2019)

To a large degree, evaluation as a discipline and practice coevolved with its evaluand (object of evaluation), reflecting it priorities, shortcomings, and strengths. This has meant an unmistakable focus on program and project evaluation, reflecting evaluation's formalization as a profession in response to the growth in public spending and programing during the 20th Century (Shadish and Luellen 2005). This, however, has engendered a self-limiting project mentality that is not fit-for-purpose for evaluating the complex, global challenges we confront today. As Patton (2020, 188) aptly observes:

> "Critically examining how we got into this situation, evaluation emerges as part of the problem, too often focused only on projects and programs that function within larger systems – but examining those larger system connection and implications would be outside the 'scope' of the evaluator's terms of reference."

Before examining more specifically the confines of the project focus and related walls that box-in evaluation, we first consider promising fracture lines in this mindset with the increasing influence of complexity and systems thinking in evaluation. This will help better frame the limitations of project/program-fixated evaluation for transformational change, and potential remedies.

The rise of complex systems analysis in evaluation, and other disciplines, has in no small part been in response to the global scale, urgency, and complexity of today's intractable challenges (Dodds and Bartram 2016; Steffen et al. 2018; UN 2013). In evaluation, these complex challenges are often referred to as ***wicked problem***s, characterized by emergent, nonlinear, and uncertain consequences that defy traditional analysis and solutions (Williams and van t'Hoff 2014, Hopson and Cram 2018). At the global scale, ***super wicked problems*** (e.g., chronic poverty, food insecurity, infectious disease, and climate change) refer to hyper-complex challenges characterized by multiple interacting systems, levels of change (e.g., local, regional, and global), and intersecting interventions and actors (Levin et al. 2012).

As **Figure 1** reflects, these global challenges are interconnected, and layered into this complexity is an assortment of actors ranging from bilateral and multilateral aid organizations, philanthropic foundations, and private donors to civil society organizations, the national public sector, and local populations. Evaluation in such wickedly complex contexts needs to extend beyond the theories of change, timeframes and funding cycles of discrete programs and projects to include multiple perspectives with varying priorities, agendas and resource flows that fluctuate in a

global economy where recession, political change and natural forces result in a high degree of uncertainty. Complex systems analysis provides both a heuristic and suite of tools for navigating such complexity, stressing that interventions occur in a broader context that requires understanding interrelationships, engaging with multiple perspectives, and reflecting on where boundaries are drawn in terms of those interrelationships and perspectives (Williams and Hummelbrunner 2010; Williams 2015; Bicket et al. 2020).

Figure 1: Complex Interconnected Global Challenges (WEF 2020)

The adoption of international agreements premised on the recognition of global interdependence, such as the Paris Climate Agreement and the UN's 2030 Agenda for Sustainable Development and its seventeen Sustainable Development Goals (SDGs), have further propelled the appropriation of systems and complexity-adaptive approaches in the evaluation of international development. Today, a range of systems-appropriate and complexity-adaptive methods are being used in evaluation, including Realist Evaluation, Developmental Evaluation, Outcome Harvesting, Contribution Analysis, Network Analysis, Most-Significant-Change, Process

Tracing and Bayesian Updating, Qualitative Impact Assessment Protocol, and more (BetterEvaluation 2020; Estelle et al. 2016, USAID 2016; Vaessen et al. 2020, Williams 2015). This is accompanied by an expanding literature on this and related topics too exhaustive to cite here, but one notable example is the "Principles for Effective Use of Systems Thinking in Evaluation" published in 2018 by the Systems in Evaluation Topical Interest Group of the American Evaluation Association (SETIG 2018).

In summary, evaluation largely reflects and follows the demands of the industries it serves, largely for the purpose of accountability. This has resulted in evaluation approaches that are predominantly project/program focused. But in response to today's wickedly complex challenges, interventions are increasingly appropriating complex systems analysis into their design and implementation. In turn, evaluation is awakening to the resultant demand to include more than discrete projects and programs in its evaluand, which is an essential prerequisite for its transformational potential. However, as we shall next discuss, bad habits are hard to break.

Four Variants of the Box

The steady rise of complexity and systems thinking presents a potential Kuhnian paradigm shift in how we package, deliver, and evaluate services, whether in international development, health care, education, or a range of other "program" areas. However, habits are hard to change, especially when institutionalized in the industries and related bureaucracies that shape and steer how and why evaluations are commissioned, designed, and implemented (Cox 2019). Consequently, evaluation has been slow to fully embrace complex systems analysis (Williams et al. 2015). In this section we identify four key barriers that box-in evaluation's transformational potential. Far from exhaustive, these four traps are interrelated and point towards what needs to be transformed if evaluation is to be transformational.

The Project Box

Introduced above, the history of evaluation has largely reinforced a project-mentality, shaped by a landscape dominated by single, clearly defined projects and programs provided by single agencies, and funded by single donors. Typically, these interventions are treated as closed systems – *boxes* – designed with a narrow scope or theory of change that excludes the broader context in which they are delivered. Transformation, however, happens in open systems that transcends time, place, political borders, and specialized interests.

A project fixation often fails to *connect the dots* in the larger picture, overlooking important complex interdependencies, spill-over and side-effects, whether they are synergistic or limiting (Patton 2020, 63). Systems transformation requires that we look beyond interventions as the main agent of change and instead consider them as one of and relative to many interrelated factors (Garcia and Zazueta 2015, 32). Hence, the value of evaluation methodologies that stress contribution analysis versus attribution analysis, and assessment that encompasses multi-sector and multi-actors *glocally* (at all levels of engagement, locally and globally).

To a large degree, the project fixation reflects a preoccupation on conceptual (logic) models or frameworks to identify casual linkages for pre-determined results that are then used to access performance and achievement of desired results. In other words, models steer the evaluative questions that evaluators examine, which, in turn, steer evaluative learning and its potential to be transformational. The problem is that the models used to design and guide project and programs – e.g. logical frameworks (and logframes), results frameworks, and Theories of Change (TOC) – typically employ linear casual analysis. This makes measurement more doable, which is good for accountability (another *box* discussed below). However, predetermined, linear models lead to reductionist planning and analysis that reinforce siloed rather than systems thinking. Rather than systemic change, evaluation fixates on an intervention's intended results, overlooking or downplaying other critical considerations.

This is represented in **Figure 2**, where the liner results chain from inputs to intended impact typically becomes the focus of evaluation to the exclusion of other influencing factors and unintended positive and negative consequences. In essence, "As the intervention's TOC is likely to direct the attention of the evaluator to the expected causal links, it can also act as a blinder to chains of causality that had not been contemplated in the intervention design, and are rarely apparent by project completion," (Garcia and Zazueta 2015, 41). Such boilerplate approaches risk narrowly confining evaluation to accountability to the exclusion of transformational learning and change.

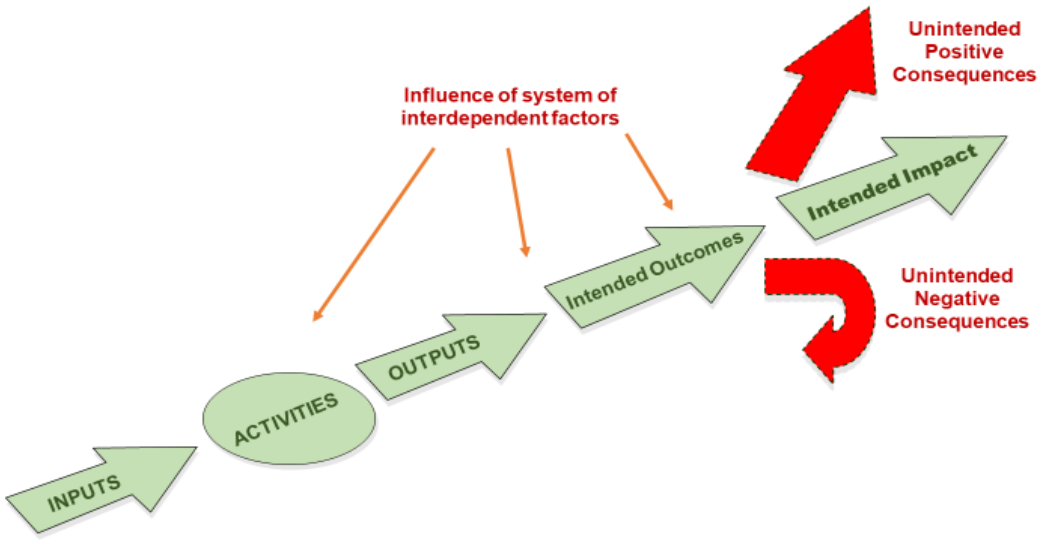

Figure 2: Linear Intervention Design
Figure courtesy of Chaplowe (2017)

The above summary is not absolute, and with the influences of complex systems analysis there are more examples of dynamic modelling of causal relationships, such as interlinked logic models or TOC with multiple branches and feedback loops. However, we need to remember that conceptual models are just that: theories of what people think will happen rather than actual maps of reality. Instead, open theories of change are better suited to design and evaluate

transformation, with attention to unintended consequences, trade-offs, and future forecasting. At the global systems level, if we are to achieve global systems transformation, we need to swap out Theories of Change for Theories of Transformation (TOT) that are adaptive and transcend programs and projects (Patton 2020, 154).

The Temporal Box

A project is designed with a given timeline, often dictated by the funding cycles of the donor. This preconceived timeframe and payment schedule is based on how the design model predicts change over time. It reflects a mechanistic casual model that evokes static predictability, order, and timing. However, complex systems do not behave according to project budget cycles. Instead, there is a myriad of emergent, intervening variables that can affect the intended intervention logic. For instance, consider how the following unpredictable events have significantly impacted the timing (and budgets) of programs and projects: the 2008 Great Recession, the presidency of Donald Trump, and COVID-19.

Rather than homing in on the predetermined timeframes of purposive theories of change, evaluation is more capable to support transformation if it is nimble, responsive to immediate and ongoing evaluative feedback that allow interventions to be more receptive to change. Such monitoring ***as*** evaluation is more iterative and ongoing than eventful evaluations like midterm or final evaluations. It is more responsive to complex operating environments, supporting emergent transformative learning, strategy testing, and course correction during implementation. It also couples well with a "developmental" approach to evaluation, where evaluators not only assess but also engage to support innovation and adaptation (Patton 2020).

Another temporal concern extends beyond the project/program timeframe to consider the degree to which evaluations assess longer-term impacts beyond funded implementation. Conventional summative (final) evaluations are typically commissioned with an endpoint or exit strategy in mind, which narrows assessment to the short- and mid-term outcomes. Meanwhile, follow-up with ex-post evaluation, (one to several years later), are rare, meaning that assessment of higher-level results and longer-term consequences are frequently left unevaluated (Zivetz et. Al. 2017).

For the evaluation of change this is significant, as it neglects the evaluation of future unintended consequences within the human and natural ecosystem. For instance, final evaluation of an agricultural export program that includes chemical fertilizers and pesticides may focus too narrowly on planned results and relevant key performance indicators (KPIs) such as productivity (e.g., crop yield) and profitability (e.g., farm income), excluding downstream costs on the local ecosystem due to resultant topsoil erosion, groundwater pollution, and biodiversity loss, and vulnerability to overexposure to the cash crop market, which loops back and harms the local farming economy over time.

A notable example of this shortsightedness in evaluation is the OECD DAC Criteria for Evaluating Development Assistance, the most widely adopted evaluation criteria in the world.

The definition for the Sustainability criteria was limited to, "measuring whether the benefits of an activity are likely to continue after donor funding has been withdrawn," (DAC 2020). The criterion was revised in 2019, but its primary intent remains the same: a conceptualization that stresses the continuity of achieved results while excluding unintended consequences on the human and natural systems (Patton 2019, 2020).

> "It is a logic of moving from one condition (a problem) to a new condition (a solution) in a way that the problem does not recur and the solution lasts. This is how evaluators have come to think and practice, but this way of conceptualizing and evaluating sustainable change is a fundamental barrier to transformation," (Patton 2019, 106).

The Quantitative Box

"Obsessive Measurement Disorder" (OMD) occurs when the production of evidence-based data undermines the very interventions they are supposed to support (Natsios 2010, 4). It is an over-reliance on quantitative measurement of pre-determined, measurable goals that burdens and distracts from a spectrum of learning opportunities. In project/program evaluation, it reinforces a narrow focus on whether it is done "right" versus whether the "right" thing is being done in the first place (Roche and Madvig 2016, 32).

Whereas it is often asserted that, "What gets measured gets done", another witticism reminds us that, "Not everything that can be counted counts, and not everything that counts can be counted." Reality is not a binary concept that can be counted, nor wrapped up into neat, quantitative boxes with key performance indicators (KPIs) that measure whether funded interventions achieve impact. The same attributes that make systems complex make interventions within hard to measure.

The "tyranny of metrics" (Muller 2018) is a box with two significant traps. First, it engenders excessive bureaucratization and "proceduraliztion" (Muller 2018; Anderson et. al. 2012, 67) that burdens and distracts organizations and project teams, promoting a top-down agenda that handicaps flexibility and innovation which is crucial for transformation.

Second, a metric fixation reflects an over reliance on quantitative methods that seek measures with objectivity and certainty. It is characterized by evaluation methods that employ experimental and quasi experimental designs and statistical analysis to determine causal impact of an intervention on target population. Probably the most prominent example of this category of evaluations is Randomized Controlled Trials (RCTs): experiments that compare the effects of an intervention to a randomly selected portion of the target population against a group that did not receive it. Sometimes referred to as the *gold standard* for impact assessment (Webber and Prouse 2018), they have had enormous value and contribution to scientific advancement, illustrated by the 2019 Nobel Prize in Economics to three economists using RCTs to address poverty by breaking it down into smaller and more precise areas to analyze.

However, RCTs are not a gold standard, and instead can be a *fool's gold* when inappropriately used (Patton 2015, 93-95). RCTs are limited to assessing effects of single treatments on individual outcomes where they control for all other observable and non-observable characteristics that can influence the casual change process (Bamberger 2016, 65-67). This means, 1) they do not to assess effects of multiple inputs on multiple outcomes that characterize comparisons in highly complex open systems, and 2) they are unable to detect unintended outcomes (changes outside the intervention logic). Both shortcomings do not accord well with the complexity and systems thinking that underpin transformational evaluation. Just because it may not be possible to establish direct causation and net results from an intervention, does not mean the intervention does not contribute to transformation (Garcia & Zazueta, 2015). As Natsios (2010, 4) reminds us:

> "(T)hose development programs that are most precisely and easily measured are the least transformational, and those programs that are the most transformational are the least measurable."

The qualitative-quantitative debate is not new territory in the field of evaluation (Green and Henry 2005), and there is general consensus that the good practice is to adopt mixed method approaches, especially as interventions become larger in scope and complexity, and beyond any one method to assess the many different dimensions (Bamberger et al. 2016, 114). However, in practice, there remains a heavy preference towards quantitative, experimental methods, which can hamper evaluation's transformational potential.

The Accountability Box

Evaluation is embedded in the political economy, and therefore subject to the same market and power forces that shape its evaluand (Nielsen et. al. 2018). In other words, the political economy shapes the industry of evaluation, including prospects for complexity adaptive, transformative evaluation. As Williams (2015) warns, despite the overture towards more systems-savvy methods in evaluation, for the most part evaluations are primarily commissioned and used for accountability purposes rather than adaptive learning, innovation, and improvement.

An *obsessive accountability disorder* largely overlaps with and fuels the other boxes. An emphasis on accountability favors commissioning methodologies and technologies for simple, measurable designs or framings of project-based interventions, where measurement is more narrowly focused on a linear chain of desired results confined to a particular timeline dictated by funding cycles. As such, evaluators are accustomed and inclined to use experimental and quasi-experimental approaches that focus on the causal relationship between intended results by controlling other factors that could otherwise explain change.

In its study on the impact of the commissioning process on complexity-appropriate evaluation, the Centre for the Evaluation of Complexity Across the Nexus (CECAN) identifies several barriers that inhibit complexity-appropriate methods (Cox 2019). One important source of barriers relates to risk-adverse attitudes related to: experimentation (in terms of innovation)

amidst tight budgets and timelines; insufficient knowledge about and capacity in new methods and technologies to commission and manage complexity approaches; concerns as to whether the approach will deliver and needing to justify methods to key stakeholders; and the overall risk of assessing delivery of unknown evaluation methods.

A troubling characteristic of a fixation on accountability in evaluation is that seeks to replace judgement with standardized measurement. The marketplace can also consign evaluation to a descriptive, tick-box, accounting exercise that steers clear of judgment rather than providing judgment that steers decision making (House 2014; Schwandt 2015). Such "valuephobia" is paradoxical in that it undermines evaluation's very core tenets to "determine merit, worth, value, or significance," (Scriven 2016, 29; 1991, 235). Reliance on quantitative metrics can side-track and release decision-makers from the responsibility of making hard judgement calls based on subject matter expertise, experience, and systems analysis, and instead rely on the allure and implication of numbers. Accountability to projects and their measurement supersedes and distracts from wider systems impacts and implications required if evaluation is to support transformational change. Perversely, "the snake of accountability eats its own tail," (Muller 2018, 154).

To a large extent, the current fixation on accountability traces to the neoliberalism of Reaganomics and Thatcherism promoting marketlike conditions and the use of performance metrics to determine reward and punishment to uphold accountability (Vedung 2010). In international development, this influence is epitomized by business models such as results-based management, value-for-money, and payment for results.

Certainly, individual, and organizational performance accountability serves intervention implementation and plays an important role ensuring funders that their investment is being put to good use. But, as previously noted, problems arise with the cumulative creep to a *tyranny of metrics* (Muller 2018; Anderson et al. 2012, 67). In development, an industry fraught with political/economic agendas that often supersede the collective good, donor monitoring, evaluation and reporting requirements can undermine and erode organizational autonomy and accountability to its own mission. This is especially concerning for local civil society organizations that are vulnerable to the capriciousness and impositions of funding, with some becoming more attuned and accountable to donors' needs than the people they are meant to represent and serve (Chaplowe and Engo, 2007). Such "clientism" (House 1995, 29) mistakenly assumes it is best to do whatever is in the clients' interests – essentially losing sight of the forest (mission) for the trees (donor).

Looking Beyond the Box

> "A new wave in evaluation history is about to break. A new mindset, new methods, and new evaluation processes are being summoned to explore and address the challenges of global pandemics, growing inequities, and existential environmental risks.

> This is part of a broader paradigm shift underway in science where interdisciplinarity has become the norm rather than the exception." (Picciotto 2020, 54)

Given our emphasis on complex systems analysis that stresses the importance of change as unfolding and uncertain, we would be amiss to predict or prescribe a specific approach for evaluation to best support transformational change. Instead, we have framed four notable barriers that box-in evaluation's transformational potential and pointed at corrective possibilities with the increasing attention given to systems and complexity thinking. **Table 1** summarizes key elements of the barriers that box-in and potential bridge that support evaluation's transformational potential.

TABLE 1 – Transformational Barriers & Bridges

Barriers	Bridges
1. **Purpose** – Accountability for intervention goals; performance reporting to funders	**1. Purpose –** Accountability for the planet; learning for innovation and transformational change
2. Perspective – reductionist, mechanistic	**2. Perspective** – holistic, comprehensive
3. Overarching principles – market driven, human-centric, separation from living systems, scarcity, tragedy of the commons	**3. Overarching principles** –cooperation, sustainability, regenerative, abundance, and prosperity of the commons
4. **Primary User** – donor/funder	4. **Primary User** – implementing team/s + key partners
5. **Engagement** – external, independent evaluation team	5. **Engagement** – multiple actors/levels/perspectives
6. Scale/evaluand – micro focus on intervention	**6. Scale/evaluand –** macro focus beyond intervention
7. **Design/modelling** - Predetermined, linear intervention logic (e.g., logic models and theories of change)	7. **Design/modelling** - Complexity-adaptive, systems-savvy models and theories of transformation
8. **Data collection methods** – preference for quantitative metrics to support counterfactual analysis	8. **Data collection methods** – mixed methods
9. Measurement – quantitative focus on predetermined goals, KPIs and targets	**9. Measurement** – mixed methods alert for emergent, unintended consequences/outcomes
10. Analytical framework – positivist objectivity; experimental designs and failure adverse	**10. Analytical framework** – systems and complexity thinking; experimentation and failure tolerant
11. **Evaluation Timeframe** – eventful, limited to intervention implementation, culminating in summative evaluation	11. **Evaluation Timeframe** – real-time, iterative, and ex-post evaluation beyond intervention

Source: Adapted from Chaplowe, Hejnowicz and Marlene, Laeubli-Loud 202

As we have noted, a key factor affecting evaluation's transformative potential is an evaluation marketplace where business-as-usual does not favor complexity-appropriate evaluation approaches. Fortunately, as also noted, there are cracks in the box that confines evaluation for transf
rmative, reflected by the increase in scholarly research, publications, and application of complexity-appropriate evaluation. A particularly exciting development to monitor are transdisciplinary initiatives partnering evaluators with academics and other practitioners across a global network to co-create and share new learning in complex systems analysis and promoting transformational change. Two promising examples include *Blue Marble Evaluation* (BME, bluemarbleevaluation.org) and *The Centre for the Evaluation of Complexity Across the Nexus* (CECAN, cecan.ac.uk).

Named after Michael Quinn Patton's seminal book (2020), BME is a global initiative based on four overarching and twelve operating principles to better understand and further develop human response to global ecosystem change through evaluation. It seeks to prepare evaluators for the local to global challenges of today and tomorrow, connecting them with others from various disciplines and practices related to transformational change. CECAN, hosted by the University of Surrey (UK), seeks to transform the practice of policy evaluation to better serve today's complex world by pioneering, testing, and promoting innovative policy evaluation approaches and methods across a range of domains, including food, energy, water, and the environment. It works through a series of 'real-life' case studies with interdisciplinary project teams comprised of social scientists, policy makers, policy analysts and experts from other fields.

Another overall promising trend to monitor is the renewed interest in *traditional Indigenous and Aboriginal worldviews***,** which are challenging the norm with alternative paradigms that support transformational learning and change in and through evaluation (Chouinard and Hopson 2016; Rowe 2019; Smith 1999). Globally, indigenous people have accumulated valuable knowledge systems that embody key aspects of transformational learning and change, especially regarding the interrelationships between natural and human systems, premised on the innate connectedness and equality between the human and non-human world (land, plants, human beings, stars, water, air, etc.). Only now are these non-Western worldviews and cosmologies beginning to gain traction, permeate and influence the field of evaluation. Initiatives such as the multi-stakeholder partnership EvalIndigenous (2020) are promoting the recognition of the different world views by advancing the contribution of Indigenous evaluation to global evaluation practice.

While not exhaustive, we will end our discussion identifying a final trend to monitor that would be neglectful omit. We are currently living through a Data Revolution. The rapid development of new tools, techniques, and data types is growing exponentially, with significant repercussions for all fields of practice that interface with information generation, management, and use. At the leading edge of this data revolution are the core trio of the data science toolbox: Big Data, Machine Learning, and Artificial Intelligence (York and Bamberger 2020; Giest 2017; USAID 2019; Vinuesa et al. 2020).

Take Big Data for example. This encompasses a huge, diverse sets of data, often generated continuously and over long periods of time, from multiple sources, including social media streams, internet searches, GPS location data, digital financial transactions, and satellite and remote sensor images. The data amassed and the data analytics it offers were unimaginable at the start of this decade, but now present us with new possibilities for more sophisticated types of analysis for evaluation and transformational change like the integration of multiple data sets for complex systems analysis and predictive modelling.

Yet, despite the potential benefits, the use of big data and other advances in data science is not a panacea. As discussed earlier with conceptual models, mathematical modelling, even when enhanced by advances in data science, is not a substitute for reality. Last year, in response to the misuse of mathematical modelling during the COVID-19 pandemic, 22 authors published a manifesto in *Nature* underscoring the inherent uncertainty in mathematical modelling, and cautioning about the politicization and misuse of models:

> "Modellers must not be permitted to project more certainty than their models deserve; and politicians must not be allowed to offload accountability to models of their choosing," (Saltelli et al. 2020, 483).

Conclusion

In conclusion, we circle back to our opening quote taken from the Prague Declaration on Evaluation for Transformational Change. The declaration acknowledges the formidable challenges that humanity faces, and affirms a commitment to evaluation that supports the transformational and systemic changes required if we are to prevail. In this article, we submit that evaluation's potential role in affecting transformational change will largely depend on its ability to transform itself. We identify four interrelated "boxes" or *bad habits* that constrain evaluation's transformational potential: a project fixation, a short-term temporal fixation, a quantitative fixation, and an accountability fixation. Alongside, we also examine the uptake and influence of complex systems analysis in the field of evaluation as a means to "breakout" of these boxes and nudge evaluation towards the inner transformation required for it to contribute to longer-term transformational change. Ultimately, acknowledging that the transformation of evaluation and the world are interconnected and will unfold unpredictably, we nevertheless contend that it would be much better if evaluation, a profession intended to improve things through assessment, can take a more proactive role in nurturing the transformations that are vital for our shared tomorrow.

Bios

Scott Chaplowe is an evaluation and strategy specialist with over 22 years' experience working with civic, public, and private organizations promoting socially and environmentally responsible

solutions for sustainable development. His background in Geography, with a focus in cultural and political ecology, informs his specialization in complex systems analysis and adaptive design and evaluation.

Adam Hejnowicz, PhD, is a postdoctoral research associate in the School of Engineering at Newcastle University and a visiting research associate in the Biology Department at the University of York. He works with the York Environmental Sustainability Institute (YESI) and is a member of the Centre for the Evaluation of Complexity Across the Nexus (CECAN).

Citations

For further reading and additional resources on the topic, we recommend the webizes of: Blue Marble Evaluation, https://bluemarbleevaluation.org; The Centre for the Evaluation of Complexity Across the Nexus, https://www.cecan.ac.uk/eppns; Zenda Ofir's Evaluation for Development, https://zendaofir.com/tag/transformation; and Scott Chaplowe's M&E Resources, https://scottchaplowe.com.

Andersen, Inger, and Johann Rockström, J. 2020. "COVID-19 Is a Symptom of a Bigger Problem: Our Planet's Ailing Health". TIME Magazine, June 5th, 2020.

Anderson, Mary B., Dayna Brown, and Isabella Jean. 2012. Time to Listen: Hearing People on the Receiving End of International Aid. CDA Collaborative Learning Projects Publication.

Bamberger, Michael, Marco Segone and Florencia Tateossian. 2016. Evaluating the Sustainable Development Goals With a "No one left behind" lens through equity-focused and gender-responsive evaluations. EvalPartners.

Bamberger, Michael, Jos Vaessen, and Estelle Raimondo. 2016. Dealing with Complexity in Development Evaluation – A Practical Approach. SAGE publications.

Bendell, Jem. 2018. "Deep Adaptation: A Map for Navigating Climate Tragedy" IFLAS Occasional Paper No.2.

BetterEvaluation.org. 2021. "Approaches" https://www.betterevaluation.org/en/approaches. (retrieved 2021-02-22).

Bicket, Martha, Ian Christie, Nigel Gilbert, Alex Penn, Dione Hills, and Helen Wilkinson. 2020. "Complexity and what it means for policy design, implementation and evaluation". CECAN Evaluation and Policy Practice Note (EPPN) Note No. 16. Centre for the Evaluation of Complexity Across the Nexus.

Chaplowe, Scott G, and Ruth Bamela Engo-Tjéga. 2007. "Civil society organizations and evaluation: Lessons from Africa". Evaluation 13, no. 2: 257-274.

Chaplowe, Scott G. 2017. "Monitoring & Evaluation Training Slides". Geneva.

Chaplowe, Scott. G, Adam P. Hejnowicz, and Marlene, Laeubli-Loud. (Forthcoming, May 2021) "Evaluation as a pathway to transformation: Lessons from Sustainable Development". In Palgrave Handbook on learning for Transformation. Edited by Aliki Nicolaides, Saskia Eschenbacher, Petra Buergelt, Yabome Gilpin-Jackson, Marguerite Welch, and Mitsunori Misawa. Palgrave MacMillan.

Chouinard, Jill Anne, and Rodney Hopson. 2016. "A Critical Exploration of Culture in International.

Cohen, Leonard. 1992. Antehm. The Future. Columbia.

Cox, Jane. 2019. "How Does the Commissioning Process Inhibit the Uptake of Complexity-Appropriate Evaluation?" Centre for the Evaluation of Complexity Across the Nexus (CECAN): Surrey, UK.

Dahler-Larsen, Peter. 2011. The Evaluation Society. California: Stanford University Press
Development Evaluation. Canadian Journal of Program Evaluation 30, no. 3: 248-276
Dodds, Felix, and Jamie Bartram. (Eds.). 2016. The Water, food, Energy and Climate Nexus: Challenges and an Agenda for Action. Earthscan, Routledge. London.

EvalIndigenous. 2021. "EvalIndigenous." https://www.evalpartners.org/evalindigenous/about. (retrieved 2021-02-22).

Garcia, Jeneen R, and Aaron Zazueta. 2015. Going beyond mixed methods to mixed approaches: A systems perspective to asking the right questions. Institute of Development Studies Bulletin 46, no. 1: 30-43.

Giest, Sarah. 2017. "Big data for policymaking: Fad or fastrack?" Policy Science 50:367-382
Green, Jennifer and Gary Henry. 2005. "Qualitative-Quantitative Debate in Evaluation." In Encyclopedia of Evaluation Edited by Sandra Mathison, 345-350.Thousand Oaks, CA: Sage Publications.

Hopson, Rodney K, and Fiona Cram. (Eds.). 2018. Tackling Wicked Problems in Complex Ecologies: The Role of Evaluation. California: Stanford University Press.

House, Ernest R. 2014. Evaluating: Values, Biases, and Practical Wisdom. (A volume by Evaluation and Society). Charlotte, North Carolina: Information Age Publishing.

House, Ernest R. 1995. Principled evaluation: A critique of the AEA Guiding Principles., In W. R. Shadish, D. L. Newman, M. A. Scheirer, and C. Wye (Eds.). Guiding principles for evaluators. New Directions for Program Evaluation, no. 66. p. 29.

IDEAS. 2019. "Prague Declaration on Evaluation for Transformational Change". Adopted by the International Development Evaluation Association (IDEAS), October 4, 2019.

Levin, Kelly., Benjami Cashore, Steven Bernstein, and Graeme Auld. 2012. "Overcoming the tragedy of super wicked problems: constraining our future selves to ameliorate global climate change". Policy Sciences 45: 123–152.

Mertens, Donna. 2009. Transformative Research and Evaluation. New York: Guilford Press.

Muller, Jerry Z. 2018. The Tyranny of Metrics. Princeton: Princeton University Press.

Natsios, Andrew. 2010. The clash of the counter-bureaucracy and development. Center for Global Development.

Nielsen, Steffen Bohin, Sebastian Lemire and Christina A. Christie (Eds.). 2018. The evaluation marketplace: Exploring the evaluation industry. New Directions for Evaluation, 160: 13–28. Wiley Publications.

Ofir, Zenda. 2020. Transformation Blog, "Evaluation for Development: A focus on Transformation and the Global South" https://zendaofir.com"

Parsons, Beverly, Lovely Dhillon, and Matt Keene (Eds.). (2020). Visionary Evaluation for a Sustainable, Equitable Future. Information Age Publishing.

Patton, M. Q. 2020. "Evaluation Criteria for Evaluating Transformation: Implications for the Coronavirus Pandemic and the Global Climate Emergency." American Journal of Evaluation 42:1.

Patton, Michael Quinn. 2020. Blue Marble Evaluation: Premises and Principles. New York: Guilford Press.

Patton, M. Q. 2019. "Transformation to global sustainability: Implications for evaluation and evaluators". Special Issue: Evaluating Sustainability: Evaluative Support for Managing Processes in the Public Interest. New Directions for Evaluation 162: 103–117.

Patton, Michael Quinn. 2017. "A Reading Guide." In Evaluation for Agenda 2030: Providing Evidence on Progress and Sustainability, edited by Rob D. van den Berg, Indran Naidoo, and Susan D. Tamondong. Exeter, UK: IDEAS. pp. xi - xvii.

Picciotto, Robert. 2020. “Evaluation and the big data challenge.” American Journal of Evaluation 4, no.2: 166-181.

Raimondo, Estelle, Jos Vaessen, and Michael Bamberger. 2016. “Chapter 2: Towards More Complexity-Responsive Evaluations: Overview and Challenges,.” In Dealing With Complexity in Development Evaluation: A Practical Approach. Edited by Michael Bamberger, Jo Vaessen, and Estelle Raimondo. SAGE Publications.

Roche, Chris, and Annette Madvig. 2016. “Working towards transformational development and the sustainable development goals”. Working Paper, La Trobe University: Australia.

Rowe, Andy. 2019. Sustainability-ready evaluation: A call to action. New Directions for Evaluation 162(Summer), 29–48.

Saltelli, Andrea., Gabriele Bammer, Isabelle Bruno, Erica Charters, Monica Di Fore, Emmanuel Didier, Wendy Espeland et al. 2020. “Five Ways to Ensure that Models Serve Society: A Manifesto”. Nature 582: 482-484.

Schwandt, Thomas A. 2015. Evaluation Foundations Revisited: Cultivating a Life of the Mind for Practice. California: Stanford University Press.

Scriven, Michael. 1991. Evaluation Thesaurus. SAGE Publications.

Scriven, Michael. 2016. Roadblocks to recognition and revolution. American Journal of Evaluation 37, no.1: 27–44.

SETIG. 2018. “Principles for Effective Use of Systems Thinking in Evaluation.” Systems in Evaluation Topical Interest Group, American Evaluation Association.

Shadish, William. R., and J. K. Luellen. 2005. “History of Evaluation”. In Encyclopedia of Evaluation, Edited by Sarah Mathison,183-186. SAGE Publications Inc.

Smith, Linda Tuhiwai. 1999. Decolonizing Methodologies: Research and Indigenous Peoples. London: Zed Books Ltd.

Steffen, Will, Johann Rockström, Katherine Richardson, Tim M. Lenton, Carl Folke, Diana Liverman. et al. 2018. “Trajectories of the Earth System in the Anthropocene”. Proceedings of the National Academy of Sciences 115: 8252-8259.

UN. 2013. World Economic and Social Survey 2013: Sustainable Development Challenges. Department of Economic and Social Affairs. United Nations, New York.

UNRISD. 2016. “Policy Innovations for Transformative Change: Implementing the 2030 Agenda for Sustainable Development.” United Nations Research Institute for Social Development: Geneva.

USAID. 2016. “SPACES MERL – Systems and Complexity White Paper”. United States Agency for International Development: Washington D.C.

Vaessen, Jos, Sebastian Lemire, and Barbara Befani. 2020. Evaluation of International Development Interventions: An Overview of Approaches and Methods. Independent Evaluation Group. Washington, DC: World Bank.

Vedung, Evert. 2010. Four Waves of Evaluation Diffusion. American Journal of Evaluation, 16(3), 263-277.

Vinuesa, Ricardo, Hossein Azizpour, Iolanda Leite, Madeline Balaam, Virginia Dignum, Sami Domisch, Anna Felländer et al. 2020. “The role of artificial intelligence in achieving the Sustainable Development Goals”. Nature Communications 11, Article no:233

Waddock, Sandra, Steve Waddell, Bruce Goldstein, Bjorn Ola-Linner, Niko Schapke, and Coleen Vogel. 2020. Transformation: How to spur radical change. In Scrutton, A. (Ed), Our Future on Earth Report, Future Earth. (pp.82-90).

Webber, Sophie and Carolyn Prouse. 2018. The New Gold Standard: The Rise of Randomized Control Trials and Experimental Development. Economic Geography, 94:2, 166-187

WEF. 2020. The Global Risks Report 2020. World Economic Forum: Geneva.

Williams, Bob, and Richard, Hummelbrunner. 2010. Systems Concepts in Action: A Practitioner's Toolkit. California: Stanford University Press.

Williams, Bob, and Sjon van ’t Hof. 2014. Wicked Solutions. A Systems Approach to Complex Problems (2nd ed.).

Williams, Bob. 2015. “Prosaic or Profound? The Adoption of Systems Ideas by Impact Evaluation”. IDS Bulletin 46, no. 1: 7-16.

Williams, Bob. 2015. “Prosaic or profound? The adoption of systems ideas by impact evaluation”. Institute of Development Studies Bulletin 46, no. 1: 7-16

York, Peter and Michael Bamberger. 2020. “Measuring Results and Impact in the Age of Big Data: The nexus of evaluation, analytics, and digital technology.” The Rockefeller Foundation: New York.

Zivetz, Laurie, Jindra Cekan, and Kate Robbins. 2017. Building the evidence base for post project Evaluation: Case Study Review and Evaluability Checklists. https://valuingvoices.com/wp-content/uploads/2013/11/The-case-for-post-project-evaluation-Valuing-Voices-Final-2017.pdf. (retrieved 2021-02-22).