Open Research Online

The Open University's repository of research publications
and other research outputs

Towards Systemic Evaluation

Journal Item

oro.open.ac.uk

Towards systemic evaluation

Martin Reynolds[*1], Emily Gates[2], Richard Hummelbrunner[3], Mita Marra[4], and Bob Williams[5]

1. The Open University, Milton Keynes, UK
2. University of Illinois, Urbana-Champaign, Illinois, USA
3. ÖAR Regionalberatung Consultancy, Graz, Austria
4. Department of Political, Social, and Media Sciences, University of Salerno, Italy
5. Independent Evaluator, New Zealand

Abstract

Problems of conventional evaluation models can be understood as an impoverished 'conversation' between realities (of non-linearity, indeterminate attributes, and ever-changing context), and models of evaluating such realities. Meanwhile, ideas of systems thinking and complexity science – grouped here under the acronym STCS – struggle to gain currency in the big 'E' world of institutionalised evaluation. Four evaluation practitioners familiar with evaluation tools associated with STCS offer perspectives on issues regarding mainstream uptake of STCS in the big 'E' world. The perspectives collectively suggest three features of practicing systemic evaluation: (i) developing value in conversing between bounded values (evaluations) and unbounded reality (evaluand), with humility; (ii) developing response-ability with evaluand stakeholders based on reflexivity, with empathy; and (iii) developing adaptive rather than mere contingent use(fulness) of STCS 'tools' as part of evaluation praxis, with inevitable fallibility and an orientation towards bricolage (adaptive use). The features hint towards systemic evaluation as core to a reconfigured notion of developmental evaluation.

Key words: bricolage, complexity science, developmental evaluation, systemic evaluation, systems thinking.

Introduction

(Martin Reynolds)

Following the United Nations launch in September 2015 of the 17 sustainable development goals (SDGs) and 169 targets, attention is now anticipated towards evaluating the implementation of SDGs in the ensuing 15 years. Such evaluation might in large part be regarded as big 'E' evaluation – institutionalised demands from, and services to, policy makers, funders and commissioners, for formalised evaluations of projects, programs, policies and/or other interventions associated broadly with governance of the Anthropocene. Small 'e' evaluation, in contrast, comprise the multitude of human endeavours (including professional practices) engaged with in pursuit of making and developing value judgements. In the formal field of evaluation, evaluation involves making judgements of merit, worth and/or significance (Scriven, 1995). There have developed a range of models/ methods/ techniques/ approaches for bridging the gap between the small 'e' world of making value judgements and the big 'E' world of needing formalised evaluations. Amongst others, these include the

---

* Correspondence to: Martin Reynolds, The Open University, Milton Keynes, MK7 6AA. E-mail: martin.reynolds@open.ac.uk. Tel: +44(0)1908 654894

logical framework approach (LFA) or 'logframes' (Sartorius, 1991), experimental design (cf. Bonell et al.,2006), programme theory evaluation (Rogers *et al*., 2000), theories of change (Weiss, 1995), realist evaluation (Pawson and Tilly, 1997), systems based approaches (Williams and Iraj, 2007), and complexity approaches (Forss *et al*., 2011).

Leading up to the 2015 'International Year of Evaluation'[1] the notion of 'systemic failure' has gained prominence in the media and political discourse in the big 'E' world of governance in the Anthropocene; failure in food security - along with water, energy, environmental, personal/ national/ global – security. Failures are increasingly regarded as systemic though often with little appreciation of what this means outside of it being some other-worldly event. One intuitive notion is that if systems create failure, the fix is to have better monitoring and evaluation of such systems. Typically though, interventions, including evaluation of interventions, are regarded in the big 'E' world as systems based on a simple input-output model (Fig.1).
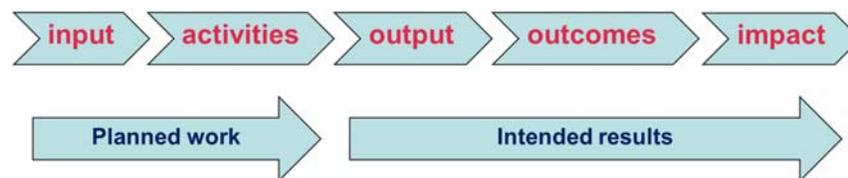

Figure 1 Simple Model Depicting Conventional Logic of Evaluation (Adapted from: http://impactinvesting.marsdd.com/social-impact-measurement/how-social-impact-measurement-tools-and-methods-fit-into-your-logic-model/)

Some little 'e' methods and techniques have accordingly found more favour than others in the big 'E' world. The logical framework approach, for example, conventionally assumes linearity in evaluating outcomes, with little attention to potential feedback effects. Similarly, randomised control trials (RCTs) associated with experimental design, assumes relatively simple causal attribution as the dominant means of making value judgements.

From a small 'e' world perspective of systems thinking and complexity science (STCS), three issues can be summarised in using conventional input-output models for evaluating complex situations. Firstly, as complexity science increasingly testifies, the situation being evaluated is always non-linear – comprising multiple feedback loops (De Haan, 2006; Marra, 2011). Secondly, the causal attribution assumed in the model masks the presence of an attributor; an individual or team of agents (evaluators, policy makers, project managers, sponsors) responsible for *selecting* causal factors and then *attributing* causality. Given the infinite inter-connectedness of reality being evaluated, an 'evaluator' has an inevitably partial position as an attributor in attributing causality. Attributing an outcome to some causal factor is a cornerstone of traditional impact measurement. In complex situations with many actors as well as many factors such attribution can be very problematic (Duignan and Casswell, 1989; Forss *et al*, 2011). Other attributes and attributions may be valid from other perspectives. Thirdly, an intervention – whether

---

[1] Declared in 2013 by EvalPartners, the global movement to strengthen national evaluation capacities.

an intervention being evaluated or an evaluation itself – is a real-world activity subject to ongoing change with uncertain effects. Reality does not stand still. In accordance with laws of requisite variety (cf. Ashby, 1958) good (bounded) models of evaluation need to somehow reflect, accommodate, and adapt to such ongoing change. Using rigid input-output models for evaluation tend by definition not to be adaptive.

This impoverished disjunction between the real ('factual') world of complicatedness, complexity and conflict, and the human (anthropogenic) world of making value judgements, presents a number of challenges particularly amongst practitioners equipped in the small 'e' world of STCS. Such challenges might be regarded in terms of an impoverished 'conversation' amongst small 'e' practitioners to more appropriately serve the big 'E' world of governance in the Anthropocene.

This paper provides an invitation to STCS practitioners for an enhanced conversation on evaluation. Four perspectives on the three challenges regarding issues of conventional evaluation practice noted above are offered from practising STCS evaluators. First, regarding complexity of situations being evaluated, Mita Marra focuses on the role and implications of complexity theory in surfacing non-linearity. Second, Bob Williams looks at the attribution of value and issues of demarcating an evaluand into boundaries of simple, complicated and complex. Addressing the third challenge, Richard Hummelbrunner and Emily Gates consider different systems methods for dealing with changing situations. Richard invites the use of three core systems concepts - inter-relationships, perspectives, and boundaries – as a generic means of grappling with complex evaluands. Emily explores the use of systems ideas for dealing with change resulting from interventions, and particularly change effecting marginalisation of interests. The four 'voices' are kept separate in the narrative below since they do not assume a 'shared' or agreed positioning, but rather provide hints towards some shared principles of systemic evaluation.

Impact evaluation and complexity theory
 (Mita Marra)

Departing from the positivist paradigm, whereby causally attributing impact relies upon randomized control trials or counterfactual analyses, the contribution of complexity to impact evaluation is manifest in three areas: (i) reconstructing program theories — including program activities, outcomes, and contextual factors affecting the extent to which program activities generate desired results (Funnell and Rogers, 2011); (ii) analyzing the evolving interactions between programs and contexts (Pawson and Tilly, 1997), and (iii) exploring change at different units of analysis as a unique or an unexpected phenomenon, which cannot be predicted from the sum of its parts (De Haan, 2006).

Through a complexity framework, evaluators can understand how, why and to what extent policies need to adapt to the environment and stakeholders' perceptions. Diverse agents learn, self-organize, and co-evolve with their environment. Order and progress can emerge naturally from social interactions and they need not be imposed centrally or from outside. Reconstructing a program theory of change needs then to rely on local cultural attitudes to understand how the intervention can work around them and modify those attitudes over time. Several evaluation approaches drawing on

ideas of complexity have been found helpful, including: 'utilization-focused evaluation' (Patton, 2008);'fourth generation evaluation' (Guba and Lincoln, 1994); 'empowerment evaluation' (Fetterman *et al.*, 1996); 'goal free evaluation' (Scriven, 1995); participatory, transformative or culturally inclusive assessment (Mertens, 2008); social experiments (Carpenter, 2005); and ethnographic accounts (Knapp, 1979); developmental evaluation (Patton, 2011).

A complex adaptive approach adds to evaluation the understanding that any intervention will always be locked into different social, economic, political and institutional systems and evaluators need to delve into those systems to examine how they facilitate or undercut different chains of causation (cf. Forss and Marra, 2014; Reynolds, 2015). Adaptation helps evaluators reconstruct a program theory of change from understanding that resources are not the final constraint nor is optimal allocation known in advance. Rather, resources are often hidden or not yet developed and thus need nurturing (Hirschman, 1967).

From a complex adaptive system perspective, causality is generative and dynamic with multiple and nonlinear causation linking interventions with their environment. Multiple program theories can then be developed with the help of evaluators responding to emergent outcomes. Such circumstances require less well-defined methods for evaluators to detect self-organizing patterns involving actors coordinating with each other. Self-organization arises from local circumstances (Stacey, 2005) since the ways people interrelate are influenced by how they find it most effective to complete tasks, given locally available resources and contexts. Patterns of self-organization are powerful because they are rooted in what is required to accomplish tasks on the ground (Lanham *et al.*, 2013). Evaluators need to acknowledge the limits of imposed structures such as implementation designs, project plans, and formal organizational hierarchies. Even in the face of formal rules, procedures, and structures designed to control, self-organization will continue based on needs that exist but that might not be recognized at higher levels of a system.

Relatedly, the notion of embeddedness stresses the role of personal relations and 'networks' of such relations in generating trust and discouraging opportunistic behavior. Embeddedness highlights how frequently actors interact, what previous experiences of interaction have led to, and whether there might be rivalries among actors that could reduce trust. In this perspective, 'participatory' evaluation methods of data collection and action-reflection processes can have a transformative power. By making beneficiaries and other stakeholders actively express their voice, evaluators facilitate the sharing of collective interpretations, which may generate potentially strong drivers of change (Marra, 2015).

Programs generally not only work to change behavior but they may also change the conditions that make programs work in the first place. The phenomenon of emergence is rooted in the way evaluators interpret situations and can help them identify how existing components in a system will combine to produce new components, continually changing the composition of a system. Considering emergence, impact evaluation needs then to explore change with the understanding that there will always be imperfect control over policy expected or desired outcomes. Evaluators need to assess those structures and processes through which people interact, exchange information and interpret observations through those

interdependencies, which link agents within micro contexts to social groups and organizations acting within broader environments. Interdependencies highlight those very macro level factors, which influence individual learning unfolding at the local level.

Crucial then is the positionality of the evaluator, who is not objectively separate from the emergent change and its constituent parts. The evaluator is part of the system and has a relationship with both the context and the emergent change (Levers, 2013). The complexity-driven impact evaluation stems from a subjectively shared experience transforming the epistemological frame of all participants and stakeholders, who recognize that social change is un-predictable, unique, and situated (De Haan, 2006; Levers, 2013), and cannot be automatically scaled-up through replicated interventions.

Contingency thinking in evaluation practice

 (Bob Williams)

The defining characteristic of evaluation is the judgement of value; often in terms of merit (an intrinsic context-free value), worth (a contextually determined, place-bound value) and significance (a value that is related to some norm or state of affairs)[2] Many evaluators shorten this to 'what works'. Indeed, there is even a 'what works' clearing house run by the US Department of Education (http://www.w-w-c.org/) that screens and identifies studies of the effectiveness of educational interventions. However in recent years the simplicity of the phrase has been criticised for being over-simplistic, ignoring the fact that what works in some circumstances does not in others. Although widely acknowledged, the criticism has been strongest from a branch of the evaluation field known as "Realist Evaluation" which has modified the phrase to variations of "what works for whom in which circumstances"[3].

In effect this is evaluation's version of the organisational approach known as contingency theory. Gareth Morgan (2006) describes this approach to managing organisations in the following way:

- Organizations are open systems that need careful management to satisfy and balance internal needs and to adapt to environmental circumstances
- There is no one best way of organizing. The appropriate form depends on the kind of task or environment one is dealing with.
- Management must be concerned, above all else, with achieving alignments and good fits
- Different types or species of organizations are needed in different types of environments

In evaluation, the 'contingency' point of view, essentially assumes that once we know and are able to categorise a situation then we can begin to predict or at least form intelligent guesses as to how worthwhile an intervention will be to intended beneficiaries. In situation X it will be a worthwhile intervention and in situation Y it will not. It also helps to identify what are often called 'unintended' or 'unanticipated' effects (cf. Morell, 2010)

---

[2] http://chiyanlam.com/2013/11/11/meritworthsignificance-explained-in-plain-language/ Accessed 14/3/16

[3] http://betterevaluation.org/approach/realist_evaluation. Accessed 14/3/16

But there are still many circumstances when evaluators are faced with the apparently unknown and unknowable, and existing evaluation approaches fail to accommodate them well. Consequently in recent years, many in the evaluation field have turned to the work of complexity theorists and practitioners Stacey, Glouberman, Snowden and especially Zimmerman, to address this aspect of their work. Glouberman and Zimmerman described situations and problems according to three categories : simple, complicated and complex (Glouberman and Zimmerman, 2004).
These descriptors have been modified to suit evaluation's particular needs and circumstances by a range of evaluators interested in the application of systems and complexity ideas in evaluation (e.g. Funnell and Rogers, 2011; Patton, 2010). Typically the categorisations look something like Table 1 prepared by Richard Hummelbrunner and Bob Williams (unpublished):

Table 1 Features of Contingency used for Evaluation
(adapted from original sources: Glouberman and Zimmerman (2004) and Patton
(2011) by Hummelbrunner and Williams, unpublished)

Domain	Characteristics	Cause - effect	M&E approaches
Simple 'known'	• high certainty and agreement • known right answer • best practice 'recipes'	• clear, predictable and controllable	• validate best practice • summative evaluation with linear logic models
Complicated 'know-able'	• some uncertainty and some disagreement • good practices • requires analysis, coordination and expert knowledge	• neither obvious nor predictable • depends on context • alternative routes	• validate effective practice (m & formative evaluation) • summative evaluation based on testable programme theory
Complex 'unknow-able'	• high uncertainty and high disagreement • every situation is unique • requires observing relations and (behaviour) patterns	• only evident in retrospect • depends on initial conditions	• document initial conditions • capture emerging patterns • on-going, rapid feed-back, 'developmental evaluation'

Although this version of a contingency approach is usually related to the *subject* of the evaluation, it has influenced discussions on the appropriate application of evaluation methods in those situations. For simple aspects choose method x, for more complicated choose method Y. Evaluation methods tend to draw on and claim legitimacy from applied social science methodological traditions. Yet those methods are perceived to be strong for simple and complicated aspects of situations, but weak on complex aspects of situations. Its methodological traditions and client base tend to emphasise methods that do not handle complexity successfully. Adopting a contingency approach allows evaluators to maintain their traditional methods (and the traditions of validity that go with them), and apply less traditional, riskier methods only in those aspects of situations that display overtly complex behaviours.

However, contingency approaches are controversial in the systems and complexity fields. Some reject it outright; everything is complex (cf Mowles, 2014; Reynolds, 2015). The systems field has had its own battles with contingency. Jackson and Keys (1984) proposed a system of systems methodologies that allocated particular systems methods to specific characteristics of situations that also created much controversy (cf. Midgley, 2000).

While the anti-contingency argument may be ontologically correct, the success of so many evaluations based on ontological assumptions of complicatedness or simplicity still need explaining. I believe the contingency approach has its uses from an epistemological position (Kurtz and Snowden, 2003). Epistemologically we can view aspects of situations as if they were simple or complicated, even if they are

ontologically complex. Yes things are complex but we can gain valuable insights by considering aspects of situations *as if* they were simple or merely complicated.

There are of course risks with this particular approach to contingency– the judgment call to frame the situation as complicated or simple may be inappropriate. It may mask potentially important contrasting viewpoints; perspectives that may turn out to be important sources of conflict. There are two further benefits of being able to use methods suitable for 'complicated' and 'simple' aspects of ontologically complex situations. One is essentially political; clients, donors, colleagues are more familiar with those methods and thus consider them more valid, thus the solutions although perhaps not 'perfect' (as much as any resolution to a complex situation can be perfect) are more acceptable. This is not trivial, I have had colleagues publicly criticise and refuse to be associated with a particular methodology designed to research complex aspects of situations. Ultimately we had to abandon that methodology in the face of criticisms over its validity. The other potential benefit of not regarding all situations as complex is that methodologies and methods designed to work in complex aspects of situations can be much more resource-intensive than those designed from a complicated and simple world view. If we can use such methods in at least part of our work in complex situations, then it leaves more left over for methodologies and methods associated with evaluating within complexity.

Systems thinking for evaluation practice

 (Richard Hummelbrunner)

Connecting two broad fields – systems thinking and evaluation – can be challenging. During recent years, as interest in systems ideas grew within the evaluation field, attempts were made to provide points of entry for evaluators who neither have the time nor the interest to thoroughly explore the systems field before choosing an approach or method.

In compiling an anthology on "Systems Concepts in Evaluation" (Williams and Imam, 2007) practitioners who applied systems thinking to evaluation gathered to discuss the two fields with a remit towards articulating 'systemic evaluation'. They proposed that the application of systems thinking in evaluation should be characterized by the use of three concepts, which Midgley (2000) had identified in his analysis of three successive waves of historical development, each focussing on particular concepts of systems thinking:

- Inter-relationships (1st Wave): This is probably the most familiar systems concept, partly because it is also the oldest: how things are connected, by what, to what and with what consequence, stems from the earliest thinking about systems (e.g. cybernetics, systems dynamics).
- Perspectives (2nd Wave): Inter-relationships are not neutral but can be interpreted differently and methods were developed that helped explore the implications of different perspectives that could be taken on a situation (e.g. soft systems methodology).
- Boundaries (3rd Wave): Perspectives are not neutral either and the dominant perspective defines a system's boundary, determining what is relevant and what is not, what is included and what lays outside. Approaches were

developed that explicitly focus on boundary issues like power or marginalization (e.g. critical systems thinking).

Taken together these three concepts constitute the essence of systems thinking and distinguish a systems approach from other ways for dealing with complexity. These concepts are not new to evaluation yet the systems field can offer specific contributions, for which a closer look at the specific methods or techniques is needed, that each of these concepts can be associated with. The concepts can also be used to identify questions that could be asked when thinking about a situation systemically and to frame evaluations in systemic terms by maintaining a threefold focus: *Understanding* inter-relationships, *engaging* with multiple perspectives and *reflecting* on boundary decisions.

When understood in this way, systems thinking is valid for many evaluation approaches and not confined to a particular niche. It can be braided with evaluative thinking in line with tendencies in the evaluation field towards 'multi-methodology"; methods from the systems field can be used alongside other 'traditional' evaluation methods.

Most evaluators gain access to the wide field of systems thinking through a single gate; typically being familiar with only one systems approach and its associated methods. This bears the risk of only using a single approach or of taking a one-size fits all attitude by attempting to fit the situation at hand to a particular systems approach or method, rather than exploring the systems universe for something that might fit better. Bearing in mind the three basic system concepts helps evaluators to maintain a broader focus to safeguard against these risks, yet they may require additional guidance for choosing appropriate methods from the systems field (some guidance is for instance offered by Williams and Hummelbrunner, 2011).

Using systems as tools for evaluation can generate problems. Within the systems field, historically there are two ways of conceiving systems (Checkland and Poulter, 2006): either as entities (ontologies) or as heuristic devices (epistemologies). All systems methods can be located along a continuum between ontological and epistemological ideas. Ison (2010) has emphasized the potential for confusion between system as 'thing' and system as 'process', and cautions against using systems and situations interchangeably; in effect constraining the effectiveness of systems practice. He advises practitioners (and this would include evaluators engaging with systems thinking) to not start with systems but with situations - which might be understood as complex, confusing or wicked. Making a conscious choice of engaging in a systemic inquiry of this situation, requires avoiding the traps of confusing situations with systems.

Engaging in a systemic evaluation of a situation requires three orientations (Williams and Hummelbrunner, 2011):

1. Reflectiveness: being aware (and checking the validity) of assumptions, mental models and values - and how they affect what we see or hear. Evaluations should be conceived as reflective practice processes consisting of successive reflective loops.

2. Respect and trust self-organization: paying due attention to emergence and value differences from plans, as these can provide useful clues for improvements. Evaluations should look at the entire range of processes triggered and beyond original intentions.
3. Think–act-think circularity: enacting non-linear praxis rather than linearity: Since many interventions are still predominantly conceived in a 'linear' fashion, evaluations should transform them into 'circular' ones, i.e. structured as a system and linked in a recursive logic to the relevant operational context.

However, for many evaluators this notion of being systemic is perhaps the most difficult aspect of applying systems thinking in practice. One can learn about concepts and methods, but 'being systemic' confronts our ideas about expertise, values and certainty in ways that are often more emotional than cognitive. For instance, the three orientations mentioned above require coping with recursive evaluation designs or reviewing intentional designs - and make us uncomfortably aware of our (inevitable) leaning towards linear and mechanistic thinking. Some evaluators (and evaluation commissioners alike) are comfortable about these challenges and ambiguities whilst others are less comfortable.

Attending to the marginalised in evaluation practice
 (Emily Gates)

A significant consideration when evaluating social policies and programs is determining whose or what interests an evaluation should serve (Greene, 1997, 2011; Schwandt, 2003). Critical systems heuristics (CSH)[4] - a strand of critical systems thinking (Midgley, 2007; Ulrich, 2002a, 2002b; Ulrich, 2012),- can help those conducting evaluations to practically and morally address this challenge by facilitating processes through which an evaluation can attend to whose and what interests are and should be included, excluded, and marginalized in an evaluation. This facilitates reflection and discussion about potential unintended consequences (Williams, 2015) and contributes to high quality evaluations.

Determining whose or what interests (i.e. perspectives, values, worldviews) are and should be relevant (and not relevant) to an evaluation is challenging. Evaluators typically take one of two different orientations to addressing this challenge (Datta, 2011). One orientation is characterized by a "belief in the public interest or common good that transcends the diversity, a highest common denominator, and a role for evaluators as sources of unbiased, fair information relating to this interest" (*ibid* p. 279). This approach aspires for evaluators and evaluations to be politically unbiased and impartial. Another orientation contends that impartiality is not feasible, and, instead, argues for advancing particular values and interests in evaluation (*ibid* p. 273). This approach typically aspires for stakeholder participation and inclusion of diverse viewpoints and values in an evaluation, particularly, the voices and experiences of historically and contemporarily underrepresented and disenfranchised

---

[4] CSH, as discussed here, can broadly be characterized as a critical strand of applied systems thinking that focuses on the normative core of professional practice (Ulrich, 2012) represented in the works of C. West Churchman (1968, 1971, 1979), Werner Ulrich (1983), Gerald Midgley (2000), and Martin Reynolds (Ulrich and Reynolds, 2010).

groups in society (Datta, 2011). A key difference between these two orientations is that the first espouses impartiality and the latter advances partiality towards particular views, values, and interests.

Critical systems heuristics takes a rather different orientation to the challenge of determining whose or what interests should be served by a social inquiry or intervention. CSH contends that any social inquiry or intervention (including an evaluation) is inevitably selective – considering *some* facts and values relevant – and partial – benefiting *some* groups and interests better than others (Ulrich and Reynolds, 2010). Addressing whose and what interests to privilege should be determined through a process that critically examines alternative options for inclusion while considering implications for who or what may be excluded and/or marginalized by these options. The aim is not for evaluators and evaluations to be impartial or partial towards particular interests; but, instead, to be *critically partial* – making transparent, questioning, and justifying the selectivity and partiality of an evaluation, and/ or whose and what interests are included, excluded, and marginalized.

There are two processes evaluators can use to support critical partiality - boundary critique (BC) and critical systems heuristics (CSH)[5]. BC is a process of questioning, debating, and justifying boundaries – the assumptions, decisions, or judgments made about what empirical considerations (e.g., 'factual' judgments of observations, facts, information) and what normative considerations (e.g., value judgments of norms, perspectives, interests, worldviews) are relevant and not relevant. Critiquing boundaries involves surfacing implicit boundaries; examining potential practical, political, and ethical consequences; considering alternative boundaries; and ultimately making transparent and justifying the boundaries used while remaining open to contestation and revision (Ulrich, 2002a). CSH is a process for making boundaries explicit through constructing reference systems as ways of framing an intervention or evaluation. This involves answering a set of twelve questions[6] in two modes: a normative mode regarding what "ought to be" and a descriptive mode regarding what "is" (Ulrich, 2002b; Ulrich and Reynolds, 2010).

By using BC and CSH in an evaluation, evaluators can engage in an explicit and systematic process of attending to inclusion, exclusion, and marginalization in an evaluation. Typically, this would involve critiquing boundaries and constructing a reference system of the policy, program, or initiative being evaluated (i.e. first-order intervention) and of the evaluation (i.e. second-order intervention). First and second-order use of BC and CSH both call attention to potential practical, political, and ethical consequences of the intervention and of the evaluation. While examining potential intended and unintended consequences of interventions might be typically part of an evaluation, CSH urges evaluators to also consider the evaluation as an intervention and to just as carefully examine potential consequences of an evaluation. Carrying out BC and CSH in an evaluation strengthens the credibility of an evaluation – by ensuring that multiple views and interests are considered; the defensibility of an

---

[5] For evaluations using boundary critique see Midgley (1996) and Midgley *et al.,* (1998). For evaluations using critical systems heuristics see Reynolds (2007; 2014).

[6] Questions address four sources of influence on an intervention or evaluation: 1) motivation (e.g. where does a sense of purposefulness and value come from?; 2) power (e.g. who is in control of what is going on and is needed for success?); 3) knowledge (e.g. what experience and expertise support the claim?); and 4) legitimacy (e.g. where does legitimacy lie?) (Ulrich and Reynolds, 2010).

evaluation – by making transparent and justifying the evidence, values, and boundaries on which an evaluation is based; and the legitimacy of an evaluation – by fostering awareness of and responsibility for potential consequences contributes to the legitimacy of an evaluation.

However, using BC and CSH in an evaluation poses several challenges for evaluators. One challenge comes from the way many evaluations are commissioned. Evaluators often work for specific evaluation commissioners under the requirements of a contract that limits the scope of work, and in short time frames with scarce resources. For political and logistical reasons, commissioners may initially resist (or even not hire) evaluators who consider multiple ways to frame a policy or program or ensure that the perspectives, interests, and worldviews of groups potentially negatively affected are considered. Another challenge has to do with the inherent contestability of boundaries, as groups with different worldviews, values, and interests may make and push for different boundaries in an evaluation. Evaluators may want to consider building space into an evaluation for those involved and affected to reflect on and deliberate about boundaries. A final broader challenge for the evaluation field is re-considering the role and responsibilities of professional evaluators to incorporate a critically partial orientation. In CSH, dealing critically with boundaries and taking responsibility for decisions about who or what is excluded and marginalized is a core responsibility of all professionals (Ulrich, 2002a; 2002b).

Summary: features of systemic evaluation

(Martin Reynolds)

Three features of *systemic* evaluation might be discerned from the challenges noted by the four practicing evaluators above. The features are a development of more general features of systems thinking signalled by Hummelbrunner. They provide a counter-play to the three shortcomings of *systematic* linear input-output evaluation approaches referred to in the introduction. Each feature serves the challenge of encouraging the use of (small 'e') evaluation tools of systems thinking and complexity science for (big 'E') mainstream evaluation practice. The three features refer to: (i) the context of using STCS tools – developing 'conversation' between systems and reality (between bounded 'values' and unbounded 'facts'); (ii) the users of STCS tools – developing responsibility of practitioners as reflexive agents of systemic change; and (iii) the usefulness of STCS tools – developing adaptability as part of innovativeness in evaluation practices.

Firstly, the context of possibly using STCS tools currently comprises real world interdependent issues of securing the implementation of sustainable development goals. Michael Patton has responded to the challenge of SDGs in rethinking the evaluand in terms of 'the Blue Marble' – the famous image of planet Earth taken in 1972, by the crew of Apollo 17 (Patton, 2016; http://www.utilization-focusedevaluation.org/blue-marble-evaluators). The Blue Marble perspective means thinking globally, holistically, and systemically: in essence, thinking of the world and its peoples as the evaluand. This means thinking beyond national states. Patton notes that the image illustrates no human induced national borders, cultures, or races, and hence requires 'global evaluators' to bound their valuations. Blue Marble illustrates the long-serving systems adage that a system is merely a *map* of a situation or

territory, not to be confused with the actual *territory*. Blue Marble represents the *territory* not the *map*. Systemic evaluation speaks not of just doing a map of the territory to gain value judgements, but rather *conversing* between the systematic 'doing' of bounded maps and continually systemically 'being' aware of the realities. Partitioning realities at the outset in terms of viewing them as simple, complicated or complex through contingency thinking, is itself a bounded value judgement. As Williams and Hummelbrunner note, there may be good practical reasons for making such judgements in an evaluation, but such value judgements ought not to be confused with judgements of fact. Complex and conflictual issues of emergence and unpredictability highlighted by Mara, and unintended consequences and detrimental effects highlighted by Gates, are real world constituents of all evaluands (situations being evaluated); all part of the Blue Marble.

Second, users of STCS tools for systemic evaluation invite a particular sense of responsibility or more precisely response-ability. As all four practitioners have variously hinted, developing meaningful conversation amongst stakeholders in an evaluand requires acknowledgement of the evaluator role as a partial change agent; a political rather than some 'objective' impartial agent. Engaging with perspectives as part of systemic evaluation may therefore require not only empathy in being able to 'hear' concerns, but also reflexivity in making transparent evaluators' own partiality and indeterminate effect on the evaluand as part of the evaluation. With the growth of big 'E' evaluation, professional evaluation societies and bodies have correspondingly grown in the past twenty years – particularly under the auspices of the American Evaluation Association and the European Evaluation Society, alongside similar international bodies covering Africa and South America. An increasing sense of professional identity amongst evaluators provides a nexus of conversation between big 'E' norms and practices and various small 'e' insights and initiatives arising from different fields of practice. From a STCS perspective this would invoke an inevitable fallibility of evaluators. As part of the ongoing conversation over whether evaluation might be modelled as a profession rather than a trade, Thomas Schwandt (2015) suggests that evaluators are engaged in a political as well as an ethical capacity. Schwandt therefore suggest moving towards a 'democratic' as against 'technical' professional model of evaluation practice (Pers. Comm. January, 2016).

Finally, systemic evaluation might regard the usefulness of STCS tools beyond reified notions of constituting a toolbox, 'delivering' on a contingency basis. STCS tools might be regarded instead as a continually developing craft-skill set; a means for adapting and customising existing practices –conventional evaluation practices as well as practices associated with STCS - as part of innovative evaluation. Such craft skills are captured by a metaphor used by Patton (2011 pp. 264–304) from the French anthropologist Levi-Stauss in the 1960s; the art of a bricoleur. As Patton explains, this is a travelling nineteenth-century rural craftsman in France skilled in using and adapting tools at hand, but with, as well, a sense of inventiveness and creativity for purposeful intervention – the art of bricolage (Reynolds, 2015 pp.80-81). Amongst the four practitioners, Marra signals a feature of complexity thinking tools as being positively less-well defined, Williams and Hummelbrunner comment on how systems might be used as either ontological or epistemological devices, and Gates signals the potential adaptability of one set of tools – critical systems heuristics and boundary critique – for different professional practices. More generically the systems idea can be used to depict 'natural' systems, human-made artefactual (purposive/ mechanistic)

systems, and human-enacted (purposeful/ social) systems (see 'A Delicate Balance' this issue). Whilst evaluation is a human purposeful activity system, systemic evaluation might adopt and adapt variations of any system conceptualisations in order to make and develop appropriate value judgements.

Together the three features outlined here hint towards systemic evaluation as core to a reconfigured notion of what Patton refers to as developmental evaluation (Patton, 2011).

References

Ashby WR. 1958. Requisite variety and its implications for the control of complex systems. *Cybernetica*, **1**: 83-99.

Bonell C, Hargreaves J, Strange V, Pronyk P, Porter J. 2006. Should structural interventions be evaluated using RCTs? The case of HIV prevention. *Social Science & Medicine* **3**(5): 1135-1142.

Carpenter J, Harrison G, List JA (eds). 2005. *Field Experiments in Economics*. Stamford. CT: JAI Press.

Checkland PB, Poulter J. 2006. *Learning for Action. A short definite account of soft systems methodology and its use for practitioners, teachers and students.* Wiley: Chichester. Churchman CW. 1968. *The Systems Approach*. Basic Books: New York,

Churchman CW. 1979. *The Systems Approach and Its Enemies.* Basic Books: New York

Churchman CW. 1971. *The Design of Inquiring Systems: Basic Concepts of Systems and Organizations.* Basic Books: New York

Datta LE. 2011. Politics and evaluation: more than methodology. *American Journal of Evaluation* **32**(2): 273-294.
De Haan J. 2006. How emergence arises. *Ecological Complexity* **3**: 293-301.

Duignan P, Casswell S. 1989. Evaluating community development programs for health promotion: problems illustrated by a New Zealand example. *Community Health Studies* **13**(1):74-81.

Fetterman D, Kaftarian SJ, Wandersman A. 1996. *Empowerment Evaluation: Knowledge and Tools for Self-Assessment and Accountability*. SAGE: London

Funnell SC, Rogers PJ. 2011. Purposeful Program Theory Purposeful Program Theory: Effective Use of Theories of Change and Logic Models. In Patton MQ. (ed). *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*. Guilford Press: New York

Forss K, Marra M, Schwartz R. (eds). 2011. *Evaluating the complex: Attribution, contribution, and beyond* (Vol. 1). Transaction Publishers: New Brunswick, NJ

Forss K, Marra M (eds) 2014. *Speaking Justice to Power. Ethical and Methodological Challenges for Evaluators*. Transaction Publishers: New Brunswick, NJ

Glouberman S, Zimmerman B. 2004. *Complicated and complex systems: What would successful reform of Medicare look like*? In: Forest GP, McIntosh T, Marchildon G (eds) Health Care Services and the Process of Change.21-53. Originally published as Discussion Paper No 8 (2002) Ottawa: Commission on the Future of Health Care in Canada *Romanow Papers* **2,** 21-53
Greene JC. 1997. Evaluation as advocacy. *American Journal of Evaluation* **18**(1).

Greene J. 2011. Values-engaged Evaluations. In Segone M (ed), UNICEF:192-206.
Guba EG, Lincoln YS. 1994. *Fourth Generation Evaluation*. Sage: London

Hirschman AO. 1967. *Development Projects Observed*. The Brookings Institution: Washington, DC.

Ison R. 2010. *Systems Practice: How to Act in a Climate-Change World*. Springer: London.
    Jackson MC 2003. *Systems Thinking : Creative Holism for Managers*. Wiley: London

Jackson MC, Keys P. 1984. Towards a system of systems methodologies. *Journal of the operational research society* **35** (6) 473-486.

Knapp MS. 1979. Ethnographic contributions to evaluation research. In Cook TD,   Reichardt CS (eds), *Qualitative and quantitative methods in evaluation research* . Sage: Beverly Hills, CA; 118-139

Kurtz CF, Snowden DJ. 2003. The new dynamics of strategy: Sense-making in a complex and complicated world. *IBM Journal of Research and Development* **42**(3): 462–483.

Lanham HJ, Leykum LK, Taylor BS, et al. 2013. How complexity science can inform scale-up and spread in health care: understanding the role of self-organization in variation across local contexts. *Social Science & Medicine* **93**: 194–202.

Levers, MJD. 2013. Philosophical Paradigms, Grounded Theory, and Perspectives on Emergence, *Sage Open* **3**(4): 1-6, DOI: 10.1177/2158244013517243

Marra M. 2011. Some insights from complexity science for the evaluation of complex policies. In: Forss K, Marra M,Schwartz R (eds), *Evaluating the Complex. Beyond Attribution and Contribution*. Transaction Publishers: New Brunswick, NJ

Marra M. 2015. Cooperating for a more egalitarian society: Complexity theory to evaluate gender equity, *Evaluation* **21**(1): 32–46.

Mertens D 2007. Transformative paradigm: mixed methods and social justice. *Journal of Mixed Methods Research* **1**(3): 212–25.

Midgley G. 1996. Evaluating Services for People with Disabilities: A Critical Systems Perspective. *Evaluation* **2**(1): 67-84.

Midgley G. 2000. *Systemic Intervention: Philosophy, Methodology and Practice.* Kluwer Academic/Plenum Publishers: New York

Midgley, G. 2007. Systems Thinking for Evaluation. In:Williams B. Imam I. (eds) *Systems Concepts in Evaluation: An Expert Anthology* EdgePress of Inverness: Point Reyes, CA; 11-34

Midgley G, Munlo I,Brown, M. 1998. The theory and practice of boundary critique: developing housing services for older people. *Journal of the Operational Research Society* **49**(5): 467–478.
    Morgan G. 2006. *Images of Organization.* Sage. London

Morell JA. (ed) 2010. *Evaluation in the Face of Uncertainty: Anticipating Surprise and Responding to the Inevitable*. Guilford Press: New York

Mowles C. 2014. Complex, but not quite complex enough: The turn to the complexity sciences in evaluation scholarship. *Evaluation* **20** (2): 160-175

Patton MQ. 2008. *Utilization-focused Evaluation*, 4th edn. Sage: Thousand Oaks, CA

Patton MQ. 2011. *Developmental evaluation: Applying complexity concepts to enhance innovation and use*. Guilford Press: New York.

Patton MQ. 2016. A Transcultural Global Systems Perspective: In Search of Blue Marble Evaluators. *Canadian Journal of Program Evaluation* **30**(3).
    Pawson R,Tilly N. 1997. *Realist Evaluation*. London: SAGE.

Ramalingam B. 2013. *Aid on the edge of chaos: rethinking international cooperation in a complex world*. OUP: Oxford.

Reynolds M. 2007. Evaluation based on critical systems heuristics. In: Williams B, Imam I. (eds). *Using Systems Concepts in Evaluation: An Expert Anthology*. EdgePress: Point Reyes CA, USA; (101–122).

Reynolds M. 2014. Equity-focused developmental evaluation using critical systems thinking. *Evaluation —The International Journal of Theory, Research and Practice* **20**(1): 75-95.

Reynolds M. 2015. (Breaking) The Iron Triangle of Evaluation. *IDS Bulletin* **46**(1): 71–86.

Rogers PJ, Petrosino A, Huebner T A, Hacsi TA. 2000. Program theory evaluation: Practice, promise, and problems. *New directions for evaluation*, *2000* (87): 5-13

Sartorius RH. 1991. The logical framework approach to project design and management. *Evaluation practice* **12** (2)139-147.

Scriven M. 1995. The logic of evaluation and evaluation practice. *New Developments for Evaluation*, 68. 49-70. Jossey Bass: San Francisco, CA

Schwandt T A. 2003. In Search of the Political Morality of Evaluation Practice. *Studies in Educational Policy and Educational Philosophy*, 1-6.

Schwandt TA. 2015. *Evaluation Foundations Revisited: Cultivating a Life of the Mind for Practice*. Stanford University Press: Stanford, CA

Stacey RD. (ed) 2005. *Experiencing Emergence in Organizations: Local Interactions and the Emergence of Global Patterns*. Routledge: London.

Ulrich W. 1983. *Critical Heuristics of Social Planning: a new approach to practical philosophy*. Stuttgart (Chichester), Haupt (John Wiley - paperback version).

Ulrich W. 2002a. Boundary Critique. *The Informed Student Guide to Management Science*. 41-42.Thomson Learning: London.

Ulrich W. 2002b. Critical Systems Heuristics. In Daellenbach HG, Flood RL (eds), *The Informed Student Guide to Management Science* 72.Thomson Learning: London; 72

Ulrich W. 2012. *CST's Two Ways: A Concise Account of Critical Systems Thinking* Retrieved from http://wulrich.com/downloads.html (accessed June, 2016)

Ulrich W, Reynolds M. 2010. Critical Systems Heuristics. In Reynolds M, Holwell S (eds) *Systems Approaches to Managing Change: A Practical Guide* Springer: London; 243-292

Weiss CH. 1995. Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community initiatives for children and families. *New approaches to evaluating community initiatives: Concepts, methods, and contexts*: 65-92.

Williams B. 2015. Prosaic or profound? The adoption of systems ideas by impact evaluation. *IDS Bulletin* **46** (1) 7-16.

Williams B. Iraj I. (eds) 2007. *Systems concepts in evaluation: An expert anthology*. American Evaluation Association/EdgePress: Point Reyes.

Williams B, Hummelbrunner R. 2011. *Systems Concepts in Action: a practitioner's toolkit*. Stanford University Press: Stanford.

---

[1] The Open University, Milton Keynes, UK

[2] University of Illinois, Urbana-Champaign, Illinois, USA.

[3] ÖAR Regionalberatung Consultancy, Graz, Austria

[4] Dept. of Political, Social, and Media Sciences, University of Salerno, Italy

[5] Independent Evaluator, New Zealand