

**This is the version accepted by the editor. For the published version see:**

**Ann Larson, Evaluation amidst complexity: eight evaluation questions to explain how complex adaptive systems affect program impact, *Evaluation*, 24(3):353-62. Copyright © 2018. Reprinted by permission of SAGE Publications.**

<http://journals.sagepub.com/doi/10.1177/1356389018781357>

## **Evaluation amidst complexity: eight evaluation questions to explain how complex adaptive systems affect program impact**

### **Abstract**

International development assistance is increasingly being seen as operating within complex adaptive systems (CAS). Evaluators have been developing new methods and approaches that are compatible with the dynamic and unpredictable realities of CAS, which are composed of many separate but interacting agents and groups. However, these new methods are not easily implemented in a conventional commissioned end-of-program evaluation. This paper builds on what is known about how CAS properties affect program performance to propose eight evaluation questions. The findings from the questions will reveal if implementers were aware of and responded effectively to CAS properties. The eight questions can be incorporated within a conventional evaluation to create a plausible narrative for program impact and inform the design and implementation of other programs.

**Key words:** Complex adaptive systems; evaluation; complexity; evaluation questions

**Running title:** Evaluation amidst complexity

# Evaluation amidst complexity: eight evaluation questions to explain how complex adaptive systems affect program impact

## Introduction to complex adaptive systems

Planners, funding bodies and implementers working in international development assistance are increasingly acknowledging that the context in which they work is complex and that positive impact depends on adapting to that complexity (Paina and Peters, 2012; Ramalingam, 2013). Complex adaptive systems (CAS) is the term used to give greater specificity and practical relevance to the concept of complex social and political contexts (Carey et al., 2015). CAS are loosely bounded systems made up of diverse agents operating in different units or groups but interacting with each other and reacting to external forces. The units follow their own rules but also capable of independent action (Axelrod and Cohen, 1999; Page, 2011).

Properties of CAS relevant to the implementation and impact of social programs are described below and summarized in Table 1. These properties draw on the work of authors from diverse social science and evaluation perspectives who use CAS as a metaphor or meso-theory to understand why some intentional efforts to change systems (interventions) are successful and many others are not (Axelrod and Cohen, 1999; Paina and Peters, 2012; Patton, 2011; Plsek and Greenhalgh, 2001; Ramalingam et al., 2008). The list is not meant to be exhaustive but to capture those properties most salient in the literature concerned with how to create change in social systems, particularly in the health and international development sectors.

CAS have histories and norms which create a **path dependency** (Axelrod and Cohen, 1999). Decisions or experiences in the past (historic and recent) constrain the range of possible actions by agents and units. This has a couple of practical implications. First, the same intervention introduced in different systems will be implemented in distinct ways, usually resulting in very different outcomes (Resnicow and Page, 2008). Second, some implementation strategies will be impossible to implement because they conflict with existing formal and informal rules (Bloom, 2014).

A source of complexity in systems is the **diversity** of actors and units. Page (2011) reviewed the theoretical and empirical literature on the role of diversity in the successful adaptation of complex systems. Although too much diversity can be a risk, systems with diverse elements are, in general, more resilient and adaptable. Yet, when introducing a system-wide intervention, some elements will perceive greater benefits and have greater capacity to adopt changes than other elements. In development assistance programs, sources of diversity come from cultural and economic differences between regions and within communities. Another source of diversity is the roles and functions of different groups within a system such as primary and secondary schools or community health workers and doctors.

Another property of CAS is that the different units are regularly interacting with each other, usually in a fairly typical and predictable pattern of **relationships**. The nature of these relationships is shaped by norms and institutional history and by the distribution of power. An intervention such as a new policy or externally introduced program may change those relationships. This change may be embraced or resisted, based not only on how the units respond but also on the effect the change has on the relationships with other units. Such **feedback** will reinforce or dampen the effect of the initial change.

At any time the relationships may foster a relatively stable set of behaviours. When changes occur, such as a new government, donor funding priority or prolonged conflict, multiple relationships will adjust to a new **attractor state**. The system re-orientes to the new reality, resulting in the **emergence** of new behaviours, many of them will be unanticipated and unintended. The emergent behaviours may be the result of small changes to relationships to adjust to the new state without significantly altering the status quo. However, agents may take the opportunity to innovate, **self-organising** to change their behaviours and relationships in ways that may or may not be able to be sustained (Lanham et al., 2013; Sarriot and Kouletio, 2014).

As a result of the properties of CAS, change is unpredictable and **nonlinear**. Outcomes are disproportionate to the initial changes: small inputs may invoke a large response, and large inputs may result in a small or delayed response.

**Table 1: Properties and behaviours of complex adaptive systems**

- Path dependency
- Diversity of actors and units (groups of actors) which are similar at different levels
- Multiple dynamic, interdependent relationships with feedback loops
- Self-organising emergent behaviours within and between units
- External shocks and shifts in attractor states
- Non-linear outcomes

Paina and Peters (2012) give a description of health systems as a CAS which is a good illustration of these properties. Health systems involve diverse providers and serve patients and populations with diverse needs. The system is made up of sole practitioners (paid and unpaid), primary care centres, clinics, and hospitals. National and sub-national health departments and facilities are influenced by politics; for example, investing in new clinics and hospitals are popular with the electorate and a source of income for those who award building contracts. The form of the system and how it operates is shaped by a legacy of funding decisions, past policies and regulations, national ideology, and discrete local identity. In low income countries, health systems are also heavily influenced by donor priorities, which regularly change. How the healthcare providers and community members respond to the changes brought about by an international development assistance program will depend on their norms, opportunities and constraints, and their perception of the consequences of adopting or resisting the change. Groups within the health system can create change by organising, such as advocacy groups or unions.

There are a growing number of examples of international development assistance programs being designed and implemented with a greater sensitivity to the context and complexity. Donor organisations are trailing several mechanisms. Some examples are tools to analyse system level influences on the desired outcome (Ramalingam et

al., 2014); processes for awarding long term, flexible funding (Hartmann et al., 2013); and improved monitoring practices to facilitate rapid decision making (Britt, 2013; Ladner, 2015).

But, in practice, explicit strategies to anticipate and work within CAS are not applied in most international development programs funded by large bilateral and multilateral donors. There are two main reasons for this lack of attention.

First, efforts to initiate and implement programs typically focus on the intervention and not the context, ignoring the relationships between the intended beneficiaries and implementers (Harris et al., 2014). Evidence-based solutions are expected to work across diverse settings, regardless of context (Subramanian et al., 2011). Failure is attributed to lack of political will or fidelity to the original intervention protocols. Tools purporting to give steps for successful implementation stresses that ‘simple’ interventions will be implemented more successfully than interventions that involve many stakeholders and actors and complex processes, ignoring the inherent complexity of the systems to be changed (Cooley et al., 2012; ExpandNet, 2009). Untested, local solutions are considered too risky (Sarriot et al., 2016).

Second, the features of CAS are most apparent when change is introduced and participants respond to it. At any point in time, such as when a program is being designed, a complex system may appear static. The degree that, for example, a health intervention might depend on the fluid relationships between, for example, nurses and managers, budgeting and procurement processes and stock-outs, religious leaders and community members are not considered up front. When those relationships result in unanticipated obstacles, implementers are not able to be sufficiently adaptable to respond.

The implication for evaluation is that many programs have not been designed to be sensitive to a dynamic context. It is the implementers who have to adapt top-down designed programs to local realities which change during the life of the program.

## **Role of evaluation**

The job of evaluators is to develop a plausible causal chain from intervention to results. Increasingly evaluators, facing unexpected, nonlinear results and the interdependence of various stakeholders and program components, are considering how to address the complexity of the context in which programs operate (Walton, 2016). Even the widely-used evaluation criteria formulated by the OECD's Development Assistance Committee (DAC) for evaluating international assistance programs (Development Assistance Committee, 1991) has recently been criticised for being too linear in its model of causality and too insensitive to the diversity of values and world views (Hieider, 2017).

To measure complex dynamics and to understand how they affect program impact, evaluators have developed new approaches that take into account CAS properties (Carey et al., 2015). Preskill (2014) and colleagues have proposed a set of principles to guide complexity-sensitive evaluation. Proponents of realist evaluation argue that it is possible to uncover how context influences the pathways to change by closely studying which mechanisms were effective for which groups under what circumstances (Westhorp, 2014). Patton (2011) has championed developmental evaluation, in which evaluators work closely with implementers providing real-time information to enable programs in complex or rapidly changing situations to adapt.

However, these approaches are time-consuming and require planning from the inception of the program. The convention in the evaluation of international development assistance programs remains the commissioned mid- or end-of-program review by an external evaluator. Even though it is now commonplace to acknowledge that social programs operate within CAS, conventional evaluations still do not incorporate complexity-aware concepts or approaches (Kania et al., 2013). Evaluators report that they do not have the skills, resources or the mandate from their clients to incorporate the new evaluations into their conventional evaluation work (Walton, 2016). Ambitious theory-driven complexity consistent evaluations are resisted by funders and can be the first component of large programs to be cut (Douthwaite et al., 2017).

## **CAS-sensitive evaluation**

This paper proposes how conventional evaluations can be tweaked to be CAS-sensitive. Although the context in which the programs operate may be complex, understanding and applying CAS-principles to evaluation need not be complicated (Atun, 2012). First, the paper describes how CAS properties can affect the processes, outputs and outcomes of programs, offering examples of how these can have a positive or negative effect drawn from the literature and the author's evaluation practice. Second, the paper presents eight CAS-sensitive evaluation questions that align with DAC evaluations and CAS properties, which elucidate how program implementers incorporated and responded to the CAS context.

The evaluator's task is to describe, analyse and critique if and how a program resulted in intended and unintended change. In most cases, the program designers and implementers were confident that change was possible, even likely: the purpose was aligned with local needs and had the support of influential stakeholders; similar interventions had been found to work in other settings; and the program had been carefully planned with attention to financing, supplies, and work force capacity. Despite high standards of design, most international development assistance programs fall short of the high expectations of change.

The signs that a program is not having the changes expected because it is operating within a CAS can take several forms. Table 2 describes those signs and their underlying CAS properties. The signs are ones often documented by evaluators, such as delayed starts, confusion about purpose and roles, and lackluster uptake. A reference to a published study or evaluation is given for each sign or symptom for the interested reader to see an example. Of course, the converse is also true. Programs may succeed precisely because they are sensitive to the complex adaptive nature of the context in which they work.

**Table 2: Influence of complex adaptive system properties on programs**

<b>Signs and symptoms</b>	<b>CAS property and possible program responses</b>
<p>Slow progress in starting implementation, especially delays in necessary policy reforms or new regulations. National governments struggled to adopt their policies to top-down policies for integrated community case management advocated by donors (Bennett et al., 2014).</p>	<p><b>Path dependency.</b> It is difficult to change established practice. Time is often required to allow decision makers reconcile the new directions with current and sometimes longstanding arrangements. New policies and regulations will be most resisted if they seem to contradict lessons from previous experience.</p>
<p>Small or lack of adoption of new practices despite different strategies to change behaviour. Local groups may organise themselves to use the inputs and processes in a different way. Efforts to introduce electronic patient records are usually resisted by primary care staff because they cannot be easily adopted but many small scale innovations flourish (Greenhalgh et al., 2011).</p>	<p><b>Nonlinear outcomes and emergent behaviour.</b> Current behaviour is reinforced through a myriad of implicit and tacit norms and rules. When these are breached, there is likely to be resistance from individuals who benefited from the previous arrangements. While change may occur, it may take longer to manifest than expected. If change does occur, it can often happen quickly, accelerated by <b>emergent behaviours</b> that were not part of the original program design.</p>
<p>Resistance or lack of support</p>	<p><b>Diversity.</b> The units that make up a complex</p>

<p>among some stakeholders or intended beneficiaries; this may not be widespread, but rather concentrated in particular types of communities or workplaces. Barriers to accessing mainstream health care, and strategies to reduce them have been the focus of many programs to improve Indigenous Australian's health (Hayman, 2010).</p>	<p>adaptive system are diverse and will not react the same way to change. Different strategies are needed for different groups varying in, for example, management capacity, poverty levels, religion or culture.</p>
<p>Obstacles, sometimes relatively simple ones, cannot be overcome. Attempts to address absenteeism among teachers and health workers is affected by the different forms of incentives used and the importance workers, managers and the public put on attendance (Banerjee and Duflo, 2006).</p>	<p><b>Feedback loops.</b> Individuals and units do not know how to interact with others or work together to resist the change. Programs need to either incorporate all parts of the system or empower some to respond.</p>
<p>What gains that have been made are unlikely to continue after the program ends. Specific efforts are required to sustain changes (Sarriot and Kouletio, 2014).</p>	<p><b>Interdependence and external shocks.</b> A system can adjust to a temporary influx of resources. When those are withdrawn, the system will revert to previous practices unless there has been a permanent change in how the units and actors relate to one another either through <b>emergent behaviour</b> or</p>

	changing the incentives and norms that govern behaviour.
--	--

The retrospective nature of conventional evaluations makes them particularly well suited to uncovering how CAS function and respond to change. While the response of a CAS, can, in theory, be predicted probabilistically, in practice ‘the only way to know what a complex system will do is to observe it’ (Plsek and Greenhalgh, 2001: 627). The approach proposed uses the tools of conventional evaluations of interviewing stakeholders and triangulating responses with quantitative data and documentation.

The evaluation questions in Table 3 draw out how program design and implementation anticipated and responded to CAS properties and the barriers and opportunities that arose. The questions are aligned with the DAC evaluation criteria, offering a path to making them more complexity-sensitive.

**Table 3: Eight CAS-sensitive evaluation questions**

<b>Was the program...</b>	<b>DAC criteria</b>	<b>CAS property</b>
grounded in history and current priorities?	Relevant	Path dependency
informed by the dynamic relationships between implementers and beneficiaries and between within and between units who have different functions?	Effective	Interdependence and feedback loops
effective in influencing those dynamics to enable the intended change to occur, such as by introducing extrinsic or intrinsic incentives?	Effective	Interdependence and feedback loops
responsive to external shocks, such as new policies, program funding or process changes, new stakeholders, etc.?	Effective	External shocks and attractor states
accommodating of the diversity of	Efficient	Diversity

intended beneficiaries and other agents and units in its design and implementation by employing different approaches depending on capacity and circumstance?		
monitoring, reviewing and taking action based on regular information to ensure that over time the program was having the intended effect?	Efficient	Nonlinear outcomes
aware of and supportive of self organising and emergent behaviours relevant to the intervention?	Sustainable	Emergent behaviour
engaged in what would happen when the program ended?	Sustainable	Interdependence, feedback loops and path dependency

Most skilled evaluators would be able to incorporate these evaluation questions into their existing evaluation plans. The questions can be answered with both quantitative and qualitative data and supplement, rather than replace, other evaluation questions. Adopting them has implications for evaluation design (Adam et al., 2012; Kania et al., 2013; Preskill et al., 2014). Exploring the different experiences of diverse agents and their relationships with each other implies the participants of evaluations also need to be diverse. The emphasis on temporal change implies the collection of both qualitative and quantitative data.

The CAS-sensitive questions can be used in a variety of ways. The most straightforward would be as an interview guide with stakeholders, adapted to be specific to the program being evaluated and the relevant stakeholders. In the author's experience, stakeholders including intended beneficiaries and those somewhat removed from the program are very receptive to describing their relationships with

others and how, if at all, they have changed over time. When encouraged, they are also usually willing to discuss reasons for adopting or resisting the changed behaviour that was expected.

The questions can also be, and were first developed, as a framework for analysing a program's capacity for working within a CAS. Drawing on a wide range of data sources including but not limited to interviews, the eight evaluation questions can produce plausible causal chain to show where the program may have nudged behaviour and how change was assisted or impeded by pre-existing relationships and history.

The eight questions can also be employed across multiple case studies, using careful comparison of cases, to explicitly take account the context-specific sources of complexity and how they contributed to change under which circumstances, over what period of time and in what order (Byrne, 2013). Case studies of different interventions in the same country or the same interventions in across different CAS are particularly effective in producing examples of good practice (Larson et al., 2015).

## **Discussion**

Use of the eight CAS-sensitive evaluation questions described in this paper has three advantages. First, they can produce meaningful findings within the confines of conventional evaluations. Their application is practical and not resource intensive.

Second, asking the questions leads the evaluators to focus on how the implementers responded to the unexpected challenges and opportunities from CAS properties. From this information, the evaluator can draw a convincing narrative of how program outcomes and impacts were (or were not) achieved.

Third, an important objective for conventional evaluations is to generate lessons to improve the design and implementation of future programs. This is particularly important in light of the resistance of international development organisations to adopting more complexity-sensitive approaches.

Fourth, and most ambitiously, the findings can fill a current knowledge gaps by building the evidence base for social change within CAS to inform replication, scale-up,

sustainability, and generalisation to other settings. It is common to design programs based on a theory of change that takes into account complexity, but there is a lack of evidence on the accuracy of these models (Carey et al., 2015).

The eight evaluation questions are limited to addressing the influence of programs attempting to change a CAS in some way. Programs become, at least temporarily, a part of the CAS they are attempting to change. The programs themselves may be simple, complicated or complex but the extent of their complexity is not relevant to the evaluation approach proposed here. There is no reason to expect that complex programs are more effective in responding to CAS properties than simple programs. All programs need to be responsive.

Incorporating the eight questions is not a substitute for much deeper complexity-consistent evaluation. The approach proposed here draws the boundaries of inquiry closely around the scope of program being evaluated. It does not use methodological or theoretical approaches which would enable the evaluation to capture the complexity of the wider systems in which the program operated or the finely grained insights that come from a realist evaluation of the mechanisms of change that work under what circumstance and for whom.

## **Conclusions**

Evaluators typically become involved at the end and have a narrow mandate to assess retrospectively a program's relevance, efficiency, effectiveness and sustainability. A big part of the job is to give implementers and funders a narrative that explains obstacles, achievements and recommends next steps and lessons learned.

The complexity of the context in which a program is operating is often invoked as a reason for its failures. How CAS properties can affect program impact is well-understood (Barder, 2012). The key message of this paper is that evaluators can use a complexity lens to understand and facilitate program success – if we ask the right questions.

## Acknowledgements

An early version of this work was presented at the Australasian Evaluation Conference in Melbourne in 2015. It has its origins in work done on scaling up national MNCH interventions for a USAID flagship program. I acknowledge the advice and collegiality of Anne LaFond, Jim Ricca, Jessica Posner and Eric Sarriot in formulating the ideas in this paper and John Maxwell Owen for his insightful comments on an earlier draft.

## Funding Acknowledgement

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

- Adam T, Hsu J, de Savigny D, et al. (2012) Evaluating health systems strengthening interventions in low-income and middle-income countries: are we asking the right questions? *Health Policy and Planning* 27: iv9-iv19.
- Atun R. (2012) Health systems, systems thinking and innovation. *Health Policy and Planning* 27: iv4-iv8.
- Axelrod R and Cohen MD. (1999) *Harnessing Complexity: Organizational Implications of a Scientific Frontier*. New York: The Free Press.
- Banerjee A and Duflo E. (2006) Addressing Absence. *Journal of Economic Perspectives* 20: 117-132.
- Barder O. (2012) Development and Complexity. *PowerPoint of Kapuściński Lecture*.
- Bennett S, George A, Rodriguez D, et al. (2014) Policy challenges facing integrated community case management in Sub-Saharan Africa. *Tropical Medicine & International Health*: n/a-n/a.
- Bloom G. (2014) History, complexity and health systems research. *Social Science & Medicine* 117: 160-161.
- Britt H. (2013) Complexity-aware monitoring. Ver 2.0. *Discussion Note. Monitoring and Evaluation Series*. Washington DC: USAID.
- Byrne D. (2013) Evaluating complex social interventions in a complex world. *Evaluation* 19: 217-228.
- Carey G, Malbon E, Carey N, et al. (2015) Systems science and systems thinking for public health: a systemic review of the field. *BMJ Open* 5: e009002.
- Cooley L, Ved R and Fehlenberg K. (2012) *Scaling Up—From Vision to Large-Scale Change: Tools and Techniques for Practitioners*, Washington DC: MSI.

- Development Assistance Committee. (1991) Principles for evaluation of development assistance. Paris: OECD.
- Douthwaite B, Mayne J, Dougall C, et al. (2017) Evaluating complex interventions: a theory-driven realist informed approach. *Evaluation* 23: 294-311.
- ExpandNet. (2009) Practical guidance for scaling up health service innovations. World Health Organization.
- Greenhalgh T, Russell J, Ashcroft RE, et al. (2011) Why National eHealth Programs Need Dead Philosophers: Wittgensteinian Reflections on Policymakers' Reluctance to Learn from History. *Milbank Quarterly* 89: 533-563.
- Harris D, Batley R and Wales J. (2014) The Technical Is Political: What Does this Mean in the Health Sector? : ODI and University of Birmingham.
- Hartmann A, Kharas H, Kohl R, et al. (2013) Scaling Up Programs for the Rural Poor: IFAD's Experience, Lessons and Prospects (Phase 2) Washington DC: The Brookings Institution.
- Hayman N. (2010) Strategies to improve Indigenous access for urban and regional populations to health services. *Heart, Lung and Circulation* 19: 367-371.
- Hieider C. (2017) *Rethinking evaluation - have we had enough of R/E/E/I/S?* [Blog post]. Available at: <https://ieg.worldbankgroup.org/blog/rethinking-evaluation>.
- Kania A, Yelland GS, Nguyen DTK, et al. (2013) Capturing the complexity of evaluations of health promotion interventions: a scoping review. *Canadian Journal of Program Evaluation* 27.
- Ladner D. (2015) Strategy testing: an innovative approach to monitoring highly flexible aid programs. *Working Politically in Practice Series. Case Study No. 3*. San Francisco: The Asia Foundation
- 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.
- Larson A, McPherson R, Posner J, et al. (2015) *Scaling Up High-Impact Health Interventions in Complex Adaptive Systems: Lessons from MCHIP*. Available at: <http://www.mcsprogram.org/resources/scaling-high-impact-health-interventions-complex-adaptive-systems-lessons-mchip/>.
- Page SE. (2011) Diversity and Complexity. Princeton NJ: Princeton University Press.
- Paina L and Peters DH. (2012) Understanding pathways for scaling up health services through the lens of complex adaptive systems. *Health Policy and Planning* 27: 365-373.
- Patton MQ. (2011) *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*, New York: The Guildford Press.
- Plsek PE and Greenhalgh T. (2001) The challenge of complexity in health care. *BMJ* 323: 625-628.
- Preskill H, Gopal S, Mack K, et al. (2014) Evaluating Complexity: Propositions for

- Improving Practice. Washington, DC: FSG.
- Ramalingam B. (2013) *Aid on the Edge of Chaos: Rethinking International Cooperation in a Complex World*, London: Oxford University Press.
- Ramalingam B, Jones H, Reba T, et al. (2008) Exploring the science of complexity: ideas and implications for development and humanitarian efforts, 2nd edn. . *Working Paper 285*. London: Overseas Development Institute.
- Ramalingam B, Laric M and Primrose J. (2014) From Best Practice to Best Fit: Understanding and Navigating Wicked Problems in International Development. *Shaping Policy for Development Working Paper*. London: ODI.
- Resnicow K and Page SE. (2008) Embracing chaos and complexity: a quantum change for public health. *American Journal of Public Health* 98: 1382-1389.
- Sarriot E and Kouletio M. (2014) Community Health Systems as Complex Adaptive Systems: Ontology and Praxis Lessons from an Urban Health Experience with Demonstrated Sustainability. *Systemic Practice and Action Research*: 1-18.
- Sarriot EG, Sacks ER and Larson A. (2016) Self-organization and emergence in global health: Insights from practice, blind spots and possible contributions from complexity science. *Design & Nature and Ecodynamics* 11: 644-653.
- Subramanian S, Naimoli J, Matsubayashi T, et al. (2011) Do we have the right models for scaling up health services to achieve the Millennium Development Goals. *BMC Health Services Research* 11: 336.
- Walton M. (2016) Expert views on applying complexity theory in evaluation: opportunities and barriers. *Evaluation* 22: 410-423.
- Westhorp G. (2014) *Realist Impact Evaluation: An Introduction* London: Overseas Development Institute.