Dealing with
COMPLEXITY in
DEVELOPMENT
EVALUATION

A Practical Approach

Michael Bamberger | Jos Vaessen | Estelle Raimondo
Editors

Complexity in Development Evaluation

The Framework of the Book

Michael Bamberger,
Jos Vaessen, and Estelle Raimondo

There is a growing awareness that complexity matters in development programs and development evaluation. In this chapter we present our framework for dealing with complexity in development evaluation. We address the question: What do we mean by complexity? An important message is that dealing with complexity in development evaluation does not necessarily require that evaluators learn to apply unfamiliar and highly technical designs. While some evaluations will draw on concepts from complexity science that may be unfamiliar to many evaluators, most of the complexity-responsive approaches we describe are based on the application of established quantitative, qualitative, and mixed methods tools and techniques, but applied within a complexity-responsive framework. What is required is for evaluators and evaluation managers to have a good understanding of the various complexity dimensions that may impact an intervention. We present a framework to make sense of these dimensions. In Section 1 we take the example of a program to combat gender-based violence in El Salvador to illustrate the five dimensions we use to define complexity. In Section 2 we discuss a number of recent trends that underline the need for a complexity focus in international development. In Section 3 we present a number of conceptual frameworks to make sense of complexity as it relates to the practice of evaluation. In Section 4 we pull everything together and present a practical framework for addressing complexity in evaluation. The chapter ends with a discussion of practical applications for policymakers, managers, and evaluators.

1. Experiencing Complexity: Evaluating a Program to Combat Gender-Based Violence in El Salvador

We begin with a real-world example to illustrate what complexity means in practice. Chapter 15 describes a 10- to 15-year program to combat the serious problem of gender-based violence (GBV) in El Salvador (see Box 1.1). This case illustrates the five dimensions of complexity that constitute the framework of this book. GBV is caused or condoned by social attitudes and practices concerning the appropriate behavior of men and women in the home, the community, the workplace, and the wider society and the stresses caused by changes in the economic, social, or political situation of women and men. In the El Salvador case, GBV is caused and perpetuated by historical, political, legal, economic, and sociocultural factors that operate at national, community, and household levels. All of these factors are interconnected, so an intervention addressing one area is affected by multiple other factors. Complexity researchers use terms such as *embeddedness, interconnectedness*, and *path dependence*[1] to describe these relations.

BOX 1.1 A COMPLEX PROGRAM TO COMBAT GENDER-BASED VIOLENCE IN EL SALVADOR

The case describes a 10- to 15-year ongoing program funded by Oxfam USA to combat gender-based violence (GBV) in El Salvador. GBV has very serious consequences at the national, municipal, community, and household levels and has multiple causes relating to the high levels of violence following the civil war, the growth of gangs, and other historical, economic, legal, and sociocultural factors. The political, legal, and judiciary systems have not proven effective in addressing GBV, and there has been only limited action at the municipal level. Moreover, the level of public awareness has been rather low, particularly due to the much greater attention the mass media has given to high levels of general violence resulting from the uneasy peace settlements after the civil war. The Oxfam program that began in 2005 sought to build a broad-based alliance involving political, national, and municipal women's and community organizations. The program combines alliance building, awareness-raising through training and the mass media, and demonstration action programs in seven municipalities. The combined effect of these actions was intended to increase women's self-confidence in their ability to confront and reduce GBV; create effective local alliances between communities, schools, health services, the police, and local government; promote new proactive policies by local government and the judiciary; promote new legislation; and promote broad-based national and regional coverage of GBV in the mass media.

A second source of complexity resides in the nature of the intervention and stems from the fact that it is implemented by many national and municipal government agencies, the police, nongovernmental organizations (NGOs), women's organizations, and community groups; it has many different components and continues to evolve and change over time. Its implementation in seven largely autonomous municipalities adds to the diversity in delivery mechanisms and processes. There are also different government and civil society funding sources. Finally, all of these various components interact and affect each other in multiple ways that are difficult to capture. A third source of complexity relates to the fact that all of the participating agencies have their own governance structures, data collection systems, understandings of what the program entails, and what they hope it will achieve. Some agencies have traditionally worked well together, some have very little experience of coordination with other partners, and there are some recurrent areas of tension (e.g., between police and communities, between some government agencies and some local organizations). In addition, there are other groups interested in or affected by the program. The interaction among all of these stakeholder groups, and more generally the institutional environment, can have an important influence on program priorities, areas of resistance, and how the program operates.

A fourth dimension concerns the causal relations between program inputs and intended outcomes. In El Salvador multiple factors might contribute to a reduction in GBV, including new legislation, improved police training, improved relations

Image 1.1 "No more violence against women." The campaign in action in El Salvador.

SOURCE: Oxfam (2007). Program Strategy Paper for the Program to Prevent Gender Based Violence in El Salvador.

between police and local communities, more effective community awareness and prevention programs, and mass media campaigns encouraging women to be more proactive in preventing GBV. All of these interventions and policy changes interact with local economic, political, and sociocultural factors. Furthermore, different combinations of factors operate in different municipalities, and the same factors have different outcomes in different municipalities. Also, the causal relations may best be described as nonlinear, as an intervention in one area may produce feedback loops that change the very conditions in which the intervention operates. For example, when women gain more control over their lives, by advancing in the labor market or by taking direct action to confront GBV, initial successes may lead to negative reactions from men (or even some women) who feel that women have gone too far. Consequently, initial progress might be followed by increased domestic or public opposition so that some progress is lost (blowback). Results vary over time and across thematic and geographical areas, with slow but steady advances in some areas but negative results in others (e.g., legislation or public policies that limit women's freedom). Evaluating progress over the lifespan of a 10- to 15-year program in which processes of change are diffuse and difficult to track can be very challenging.

In addition to these four sources of complexity stemming from the intervention, its environment, and its processes of change, other factors can influence the degree of complexity of the evaluation process. In this particular case, critical data on, for example, how the police services respond to GBV or how municipal government policies are implemented on the ground is very limited, and it is also difficult to combine and compare data across different agencies. Moreover, the complex and dynamic nature of the political system has a bearing on the choices evaluators make about design and implementation. There are different constraints posed by the evaluation process itself (e.g., the purpose of the evaluation, expectations of stakeholders, time and budget constraints). Box 1.2 articulates the four dimensions of complexity laid out above as well as how these dimensions affect the level of complexity of the evaluation.

BOX 1.2 DEFINING COMPLEXITY OF THE INTERVENTION AND THE EVALUATION

Complexity of an intervention (see Box 1.1 and Table 1.2) is determined by the intersections among four dimensions: (1) contextual factors and the nature of the system within which the program is embedded, (2) the characteristics of the intervention, (3) the institutions and stakeholders and how they interact, and (4) the nature of causality and the processes of change.

In addition *the complexity of the evaluation* is also determined by (a) the purpose of the evaluation, (b) the choice of evaluation design, (c) budget and time constraints, and (d) the value orientation of both stakeholders and evaluators and the methodological preferences of the client(s) and other key stakeholders.

In short, the level of complexity of an evaluation is the result of complexity issues relating to the intervention and its wider context and a number of aspects specific to the evaluation process itself.

When discussing evaluation issues, throughout the book we refer to the distinction between *established evaluation approaches* that are used to evaluate interventions with a relatively low level of complexity (i.e., simple projects) and *complexity-responsive evaluation* approaches that are appropriate for the evaluation of complex interventions. The distinction is explained in Box 1.3.

BOX 1.3 ESTABLISHED EVALUATION APPROACHES AND COMPLEXITY-RESPONSIVE EVALUATION

Throughout the book we refer to *established evaluation approaches* as the quantitative, qualitative, and mixed methods approaches and methods that are widely used in development evaluation (summarized in Chapter 4 and covered in most evaluation textbooks). *Complexity-responsive evaluation* addresses complexity issues by building on established evaluation approaches (and approaches that are commonly used in complexity science; see Chapter 2) and relying on a number of steps that are discussed in subsequent chapters (see Chapters 3 and 7).

---

2. The Growing Need for a Complexity Perspective in Evaluation

2.1 The Move Toward "Complex" Development Interventions

Over the past decade there has been a move toward more complex development interventions on the part of international development agencies, governments, and many NGOs. In 2000 most international development agencies adopted the Millennium Development Declaration,[2] which defined development objectives in terms of eight broad goals (Millennium Development Goals; MDGs) encompassing the main areas of development and which proposed a broad framework for assessing the overall contribution to development of the large numbers of projects and programs being supported by different development agencies.

The MDG initiative was also driven by increasing demands from parliaments, public opinion, civil society, and academia to address the question: Does aid work? These concerns were reflected in the 2002 Monterrey Consensus on Financing for Development, which sought to distribute more money to the world's poorest people while at the same time increasing the efficiency and effectiveness with which aid is managed. The Monterrey Consensus also had important implications for both the focus of development evaluation (assessing the broader impacts of aid rather than the impacts of individual projects) and how evaluations should be managed (greater participation of the recipient country in the management of evaluations and greater cooperation among donor agencies in the conduct of evaluations)

(Morra Imas & Rist, 2009, p. 76). The focus on broader, national development objectives was further formalized in the 2005 Paris Declaration on Aid Effectiveness. As a result of these and subsequent agreements (such as the 2008 Accra Accords[3]), development has evolved toward a more comprehensive agenda, increasingly addressing country policy reforms, capacity building, and global concerns (Morra Imas & Rist, 2009, pp. 77–78).

Similar trends can be observed in industrialized countries where complex policy initiatives are not restricted to addressing "big challenges" but also focus on more routine initiatives such as school achievement, urban planning, and public health (Forss, Marra, & Schwartz, 2011). National governments are also increasingly involved in regional and global cooperation, working within transboundary legal, fiscal, economic, immigration, and health protocols. These trends significantly increase the complexity of policy and program planning and evaluation. There are also increasing numbers of international organizations involved in the formulation, regulation, and management of these transboundary initiatives.

Many large NGOs also recognize the need for broader and more complex interventions to achieve development goals. For example, the outcome of programs to promote gender equality, social justice, or conflict reduction are affected by a wide range of sociocultural, economic, political, legal, administrative, and often ecological factors, all of which interact in complex and unpredictable ways. There is increasing recognition that established evaluation designs are not equipped to assess the interactions among all of these dimensions, and many of the forward-looking NGOs are beginning to consider the need for complexity-responsive evaluations.

Finally, the post-2015 development agenda emphasizes the need for a global shared responsibility in addressing the world's problems based on a common agenda for sustainable development (encapsulated by the Sustainable Development Goals). This will only increase the need for the evaluation community to address highly complex and international policy initiatives and to take its place as an instrument of learning and accountability.

2.2 New Developments in Data Collection and Analysis

The rapid digitalization of society and the increasing use of new technologies change how people inform themselves about the world around them and how they act. These new sources of information provide new opportunities for more extensive and rigorous analysis of complexity, but also give rise to new challenges as organizations struggle to establish systems to manage the new sources of data (see Chapters 10, 11, and 12).

In addition, new streams of data have become available on a variety of phenomena related to human behavior, including social media and market data, which can help evaluators and policymakers better understand human behavior and how individuals and groups are affected by policy interventions. The rapid evolution of systems for the management and warehousing of multiple sources of data (see Chapter 10), the rapidly increasing use of smartphones and handheld

data collection and analysis technologies (see Chapter 11), and the real-time availability of big data are starting to affect how evaluations are designed and conducted. Some specialists argue that the move toward digitalization and big data will significantly change how both monitoring and evaluation are conducted (see Chapter 12).

There is also a marked increase in the number of evaluations, policy analyses, and policy-oriented academic studies. The number of academic journals continues to grow, and studies are now widely available on the Internet. With the increasing numbers of studies, analyses, and evaluations, the practice of synthesizing evaluative knowledge has gained in importance (see Chapter 9 and Popay, 2006). Review and synthesis approaches now include such methods as narrative review, meta-ethnography, realist synthesis, and systematic review (see, e.g., Popay, 2006). At the same time, the growing number of meta-studies such as systematic reviews has provided new impetus to so-called repositories of knowledge or second-order evidence-producing organizations (Hansen & Rieper, 2010). Examples of such organizations are the Cochrane Collaboration, the Campbell Collaboration, and, particularly relevant for development, the International Initiative for Impact Evaluation (also known as 3ie).

2.3 The Demand for Results-Based Evidence

The increased demand for measurable and credible results, combined with recognition that the evidence base on the effects of development interventions has been rather weak (Evaluation Gap Working Group, 2006), has led many agencies to adopt results-based management (RBM) systems (Kusek & Rist, 2004; Morra Imas & Rist, 2009). These systems are based on intervention logics that provide the basis for the definition and measurement of a set of numerical indicators of program outputs and outcomes with defined milestones by which numerical targets should be achieved. Most international development agencies now use some variant of RBM. Yet the challenge of addressing attribution in RBM systems remains quite daunting, especially when taking into account the many types and levels of interventions for which conventional quantitative counterfactual methods do not work.

The increased recognition that development processes are complex brings new challenges for the institutional and organizational processes underlying RBM. As Patton (2011), among others, makes explicit, tensions can emerge between a somewhat rigid and linear planning and reporting model and a need for managerial and institutional flexibility that complex interventions and contexts demand. Reynolds (2015) argues that RBM systems are designed to provide evidence of the achievement of narrowly defined results that capture only the intended objectives of the agency commissioning the evaluation. Furthermore, he argues that this narrow and inflexible approach, which he calls the "iron triangle of evaluation," is unable to adapt to the broad context within which complex programs operate and address the needs of different stakeholders. A number of reforms to make RBM more complexity-responsive have been introduced in various contexts. In Chapter 13, we present some of these options.

3. What Does Complexity Mean in Development Evaluation?

In this section we discuss a number of conceptual perspectives on complexity that have inspired the development of the conceptual framework of this book, which is presented next.

3.1 The Complexity Science and Systems Perspectives

Systems theory and complexity science represent a variety of approaches to thinking about complexity. There is no common intellectual foundation of thinking about complexity, and many definitions of complexity have been proposed by scholars. Until very recently, evaluators have made only limited use of complexity science, but this is likely to change as there are many potentially important applications. As usefully categorized by Morin (2008), two main strands of complexity theory can be identified in the literature: restricted complexity and general complexity. Box 1.4 summarizes the key elements of the two approaches and explains how in this book we incorporate elements of both approaches.

BOX 1.4 RESTRICTED AND GENERAL COMPLEXITY

Two main types of complexity have been conceptualized in the social sciences: restricted and general. These quite fundamentally different definitions and corresponding worldviews have implications for the type of evaluation approach that can and should be applied:

- *Restricted complexity* refers to a perspective whereby complexity only stems from human interactions. It does not grant any causal power to context and structures, and it is based on a conception of human agency as following a set of patterns and rules. According to this perspective complexity is amenable to mathematical treatment, through, for example, agent-based modeling and nonlinear equations. It can also accommodate modes of inquiry such as quantitative counterfactual evaluation.
- By contrast, the *general complexity* perspective is imbued with the spirit of holism and relies heavily on the concept of multilevel emergence. Complex systems are not just the product of interactions between people but have properties of their own that have to be addressed in and of themselves. One of the implications of this perspective is that variable-based methods are inherently constrained to detect empirical regularities and, instead, careful mapping and comparison of complex and context-dependent causation are required to make sense of complexity in a given geographic and temporal setting. What is needed is an account of how transformation happens by capturing the interactions between various elements (individuals, organizations, institutions, localities) and their respective causal effects on the whole system, making sense of people's agency beyond narrow behavioral rules, and exposing underlying mechanisms.

(Continued)

> (Continued)
>
> In this book we do not make an explicit choice between the two. Both views and their methodological and practical implications have merit in different evaluation contexts. Our conceptual framework—which defines complexity as the interaction of multiple dimensions, including embeddedness, and the nature of the system and institutions—is very much in line with a general complexity perspective. Some of the methodological approaches that we put forth (e.g., realist evaluations, qualitative comparative analysis, case-based methods of unpacking and re-assembling) are also compatible with this view of the world. On the other hand, we also suggest that some tools from complexity sciences (e.g., systems mapping, modeling) or variable-based methodologies (e.g., regressions, [quasi-]experimental designs) can also be valuable in dealing with complexity in development evaluation. The latter are implicitly aligned to a restricted complexity perspective.
>
> ---
>
> SOURCES: Adapted from Byrne (2013), Byrne & Callaghan (2014), Morin (2008).

A number of authors have started translating some of the complexity concepts into an international development language. Ramalingam, Jones, Toussaint, and Young (2008) usefully distinguish between the following three dimensions.[4]

Complexity and systems:

- Systems are interconnected and interdependent.
- Systems are full of feedback loops.
- Emergence: The dynamics of the system are unpredictable.

Complexity and change:

- Change is nonlinear.
- Change is path-dependent but there are also possible system shocks; systems are in continuous flux with quasi-chaotic features.

Complexity and agency:

- Adaptive agents continuously react to the system and each other, which can sometimes lead to patterns of regularity.
- Self-organization can be an emergent property, which may last for a time.
- A result of interconnectedness is co-evolution of actors and parts of the system.
- The move toward participatory planning and evaluation increases the number of agents (NGOs, community organizations, advocacy groups, etc.), which increases the complexity of the evaluand.

While readers may find many of these concepts intuitively clear, Ramalingam et al. (2008) provide a detailed discussion of these aspects. In this book we take a somewhat broader complexity perspective than the now common systems theory or complexity science approaches (Patton, 2011; Ramalingam et al., 2008; Williams & Hummelbrunner, 2011; Wolf-Branigin, 2013). We recognize that complexity is

first and foremost the property of a system, but there are a number of dimensions underlying the day-to-day practice of evaluation in the development context that we consider eminently complex and yet do not fit well in a systems theory approach.

3.2 Distinguishing Between Simple and Complex Interventions

A well-known distinction is the one between simple, complicated, and complex problems (Glouberman & Zimmermann, 2002):

- Simple problems (like following a recipe) might require a certain level of expertise, but they can be mastered and carry a high probability of success in future attempts to replicate the recipe.
- Complicated problems (like sending a rocket to the moon) require coordination and specialized expertise, but outcomes can be predicted with a fairly large certainty.
- Complex problems (like bringing up a child) can have some features of complicated problems, but they are also characterized by interdependency and nonlinearity. Complex systems carry elements of ambiguity and uncertainty as they have the capacity to adapt to changing conditions.

This simple, complicated, complex heuristic framework has been recurrently applied in the evaluation literature (e.g., Bamberger, Rugh, & Mabry, 2012; Patton, 2011; Rogers, 2008) and has been helpful in assessing the degree of complexity of different types of interventions. However, the limited practical applicability due to a significant degree of overlap (and potential confusion) between complicated and complex interventions has also been highlighted in the literature (Patton 2015; Rogers, 2011). Moreover, the use of the terms *simple projects, complicated programs,* and *complex interventions* may also suggest a strong association between the nature (and complexity) of the intervention itself and the nature (and complexity) of processes of change induced by the intervention. This is not necessarily the case. Simple projects can result in causal processes of change that can be considered complex (for a discussion of complexity in processes of change, see Chapter 4). Similarly, large scale, multi-actor, multilevel (e.g., international, national), multi-site (e.g., covering different countries) interventions may affect the behavior of institutions and individuals in ways that are relatively easy to capture.

In this book we use a somewhat different heuristic framework for assessing the degree of complexity, which enables decision makers, managers, and evaluators to assess the level of complexity of a program and its context on the dimensions of complexity discussed in Section 4 of this chapter.

3.3 Pawson's Complexity Framework

Pawson (2013, p. 29) argues that from a realist evaluation perspective, complexity in the social world is ubiquitous. One of the key assumptions of realist evaluation is that programs are complex interventions introduced into complex social

systems (Pawson, 2013; Pawson & Tilley, 1997). For instance, apparently simple interventions can be very intricate, the same intervention can provoke change in a myriad of ways, and each intervention is located "within a dynamic policy and social system surrounding it and within a dynamic cognitive and behavioral system that underpins it" (Pawson, 2013, p. 30). However, Pawson cautions against the risk of aspiring for the holy grail of holistic evaluation and recommends prioritizing the inquiry of seven sources of complexity in relation to policy interventions and their effects (see Table 5.1 in Chapter 5 for the VICTORE checklist).[5] Mapping the various dimensions of complexity according to a checklist such as this one allows the evaluator to delineate the contours of the system and make better informed decisions about empirical data collection and analysis.

We will refer to realist evaluation throughout the book both because it is one of the relatively few evaluation frameworks specifically designed to address complexity and because of its emphasis on the complex behavioral mechanisms (many of which were not intended or anticipated) through which change occurs. The approach also emphasizes the unpredictable nature of these behavioral changes and the fact that these mechanisms can operate quite differently in different program contexts.

4. Toward a Practical Framework for Addressing Complexity in Evaluation

4.1 The Complexity Framework of This Book: Five Dimensions of Complexity

Pawson and Tilley (1997) describe policy interventions as open systems within open systems. It is often not (immediately) apparent where an intervention begins or ends. Programs also first and foremost exist in the mind of the stakeholders involved in their design and implementation. Is the intervention, such as a training program in agricultural land use techniques, simply the aggregate of the inputs, activities, outputs, and different stakeholders involved? Or are there important linkages with, for example, the previous knowledge of the trainers accumulated during past training programs, the social structures in farmer communities, the implementing organization's microcredit component that in the past provided loans to part of the participating farmers, and so on? To what extent are contextual factors like religious beliefs, the availability of non-farm employment opportunities, or cultural differences between program staff and farmers important factors influencing the implementation and potential effectiveness of the program? How far does the influence of the training program in land use techniques go? Does it affect the knowledge levels and behavioral attitudes of farmers? Does it influence land use behavior, yields of crops, and subsequently market volumes and prices of crops? Do farmers and trainers learn from the experience and share this knowledge elsewhere?

The example of the intervention to combat gender-based violence in El Salvador at the beginning of this chapter highlights two key challenges that each evaluation should address: defining *boundaries* for the evaluation and *sense-making of diverse perspectives*. Boundaries refer to two related aspects: the scope of the intended program coverage and impacts as well as the purpose and scope of the

evaluation. For example, is the evaluation intended to assess the overall program effects or only to focus on a particular component or a particular geographic region? Does the evaluation focus only on direct project effects, or does it also consider indirect (secondary) effects?

Sense-making is the process by which people give meaning and value to experience. Complexity-responsive evaluation stresses the importance of understanding the perspectives that different groups use to make sense of the nature and purpose of a program and its results. Evaluators and evaluation stakeholders (e.g., donors, decision makers) need to ask themselves: What is a useful way to delimit the boundaries of an intervention and the boundaries of the evaluation? How is it embedded in and connected to different (historically determined and contemporary) societal processes? How can one make sense of the intervention and the processes of change that may be influenced by it?

While these questions are a good starting point for thinking about complexity, they represent only a first step toward a more comprehensive set of questions around complexity and evaluation. In this book we distinguish between five dimensions of complexity, all interrelated: the nature of the intervention, institutions and stakeholders, causality and change, embeddedness and the nature of the system, and finally the evaluation process itself. We discuss each of these dimensions below (see Figure 1.1).

Dimension 1: The Nature of the Intervention

In the past, evaluations tended to be primarily project-oriented. While this is still an important level at which a lot of evaluative inquiry is taking place, at the same time evaluation is increasingly focusing on higher levels of intervention: sector-wide programs, country programs, multi-country thematic strategies, etc. The defining feature of such interventions is that they involve multiple activities at different levels (e.g., local, national) with different stakeholder groups. In turn, processes of change are likely to be difficult to map as the confluence of different combinations of activities at different levels in divergent social and institutional contexts can lead to a very broad and varied spectrum of change.

BOX 1.5 QUESTIONS TO BE ADDRESSED CONCERNING DIMENSION 1

With regard to this dimension one can ask the following questions:

- What is the purpose of the intervention? To what extent have the multiple activities at different levels been clearly defined and logically linked to assumptions about processes of change contributing to the achievement of the intervention's purpose?
- What is the scope and size of the intervention? What can be said about the number and diversity in activities, actors, locations, etc.?
- To what extent does the intervention include a comprehensive monitoring framework used to periodically collect data? What sources of data are available to capture the intervention processes and activities?

Figure 1.1   The Five Dimensions of Complexity
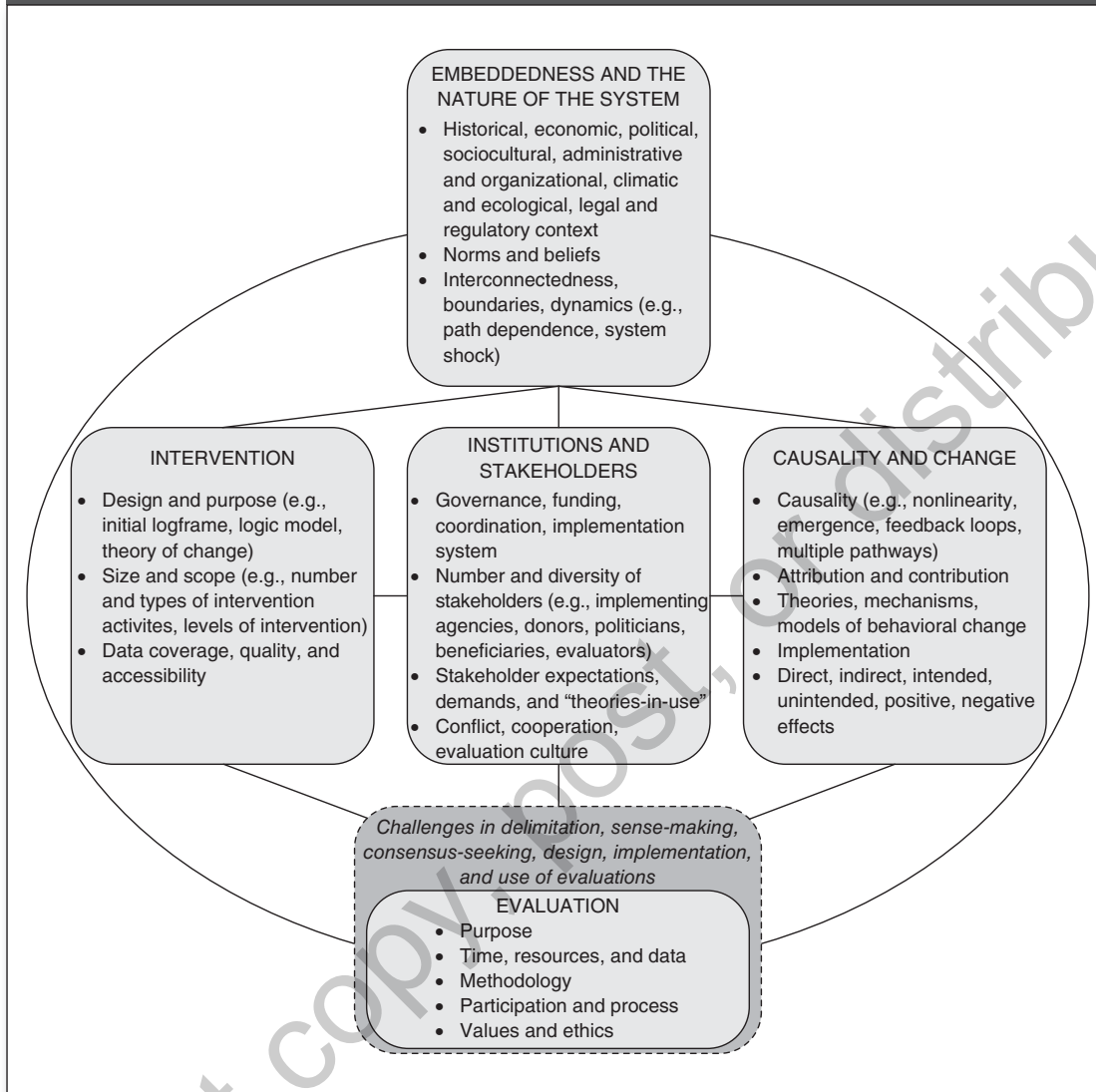
A number of variables can be helpful in characterizing the complexity of interventions:

- *Scale and geographic spread:* In general, programs with larger target populations or populations that are geographically dispersed tend to be more complex.
- *Diversity of the target population:* The more diverse the intervention's population, the more complex it is. Programs that serve a range of different ethnic or cultural groups or clients with different social, psychological, economic, and substance abuse problems tend to be more complex. Programs that seek to address the many different types of homeless populations is an example.
- *Number and diversity of program components:* The more different the components and the greater the diversity, the more complex the program. In general,

a program that provides a wide range of similar services (e.g., a primary health care program) is less complex than a program whose services are more diverse (e.g., a program that provides job placement, education, substance abuse counseling, child care, and housing services to the homeless).

- *Organizational complexity:* Programs that involve a number of different funding agencies, a implementing agencies, and oversight agencies tend to be more complex than a large, multicomponent program managed by a single agency.
- *Range, clarity and logical coherence of program objectives and theory of change (design):* The broader the range of program objectives, the greater the complexity. Complexity is also affected by any lack of clarity in the definition of objectives, the overall theory of change, or the program design and implementation strategy.

Dimension 2: Institutions and Stakeholders

The social and institutional structure of an intervention poses one of the most challenging features of an evaluation. Interventions are human constructions and each aspect of planning, funding, design, coordination, implementation, reporting, and evaluation is subject to negotiations and interactions among different stakeholders. Complexity is thus affected by how the program is governed and coordinated; the number and diversity of stakeholders; the number and diversity of funding sources; stakeholder expectations, their theories of how the program will operate (theories of change), and the level of consensus among them; and the level of conflict or cooperation among agencies. The extent to which stakeholders consider a program complex will also be important for the evaluation.

Within the context of an evaluation, it is essential to carefully map the institutional system of an intervention (i.e., the stakeholders involved in processes of governance, coordination, management, implementation, etc.). Moreover, apart from the formal institutional context, there are many informal channels through which stakeholders interact and negotiate on particular aspects of an intervention.

BOX 1.6 QUESTIONS TO BE ADDRESSED CONCERNING DIMENSION 2

With regard to this dimension, one can ask the following questions:

- What are the types and number of stakeholders associated with a particular intervention (e.g., donors, decision makers, managers, beneficiaries)?
- What are the characteristics of the systems of governance, coordination, implementation, and so on that bring these stakeholders together?
- What are the main assumptions and expectations of different stakeholder groups regarding how an intervention works or should work? To what extent do stakeholders differ in terms of these expectations and assumptions?
- To what extent can different stakeholder arenas surrounding an intervention be characterized as cultures of conflict, cooperation, learning, and so on?

Dimension 3: Causality and Change

Causal change processes are often nonlinear and characterized by emergence[6] and feedback loops.[7] Moreover, intervention outputs can affect multiple causal pathways. For example, conditional cash transfers to poor households can lead to higher school enrolment rates and have motivational effects on students. The more complex the intervention, the more likely that a complex array of causal pathways is (in)directly influenced by the intervention. Different sense-making and measurement techniques should be considered by evaluators when trying to capture and analyze complex processes of change.

BOX 1.7 QUESTIONS TO BE ADDRESSED CONCERNING DIMENSION 3

With regard to this dimension, one can ask the following questions:

– In a particular intervention context, what are the defining characteristics of causal change processes?
– How do these characteristics affect the prospects for meaningfully analyzing the attribution of change to particular intervention outputs? To what extent should evaluations employ a contribution perspective (i.e., to study the confluence of causal factors influencing a particular process of change)?
– What theories, mechanisms, or models of behavioral change can be meaningfully employed to better understand and analyze processes of change?
– To what extent are implementation processes across an intervention homogeneous? Are there any implementation failures? How does this affect the analysis of causal change?
– What is the full range of changes influenced by an intervention? What are the direct, indirect, positive, negative, intended, and unintended effects of an intervention?

It is useful to distinguish between the following parameters for delineating the effects of an intervention (Leeuw & Vaessen, 2009):

- *The level of effects:* Institutional level versus beneficiary level. Some development interventions focus on the institutional level (e.g., the quality of governance, national policies, or legislation) and through these changes affect the well-being of communities and individuals (beneficiary-level change). Other interventions (e.g., vaccination programs, conditional cash transfer programs, rural roads) can directly affect the beneficiary level.
- *The timing of effects:* Short-term versus long-term. Some effects (e.g., changes in attitudes and capacities of individuals) can occur in the short term, while others (e.g., changes in the species diversity in a particular ecosystem) can take many years to become manifest.

- *The nature of causality and direct and indirect effects:* Causal processes between interventions and (expected) change can be relatively straightforward and direct or can be long and diffuse, with many other factors influencing potential changes further down the causal chain. The following are characteristics commonly associated with processes of change from a complexity perspective (see, e.g., Ramalingam et al., 2008): nonlinearity, irreversibility, feedback loops, emergence, path dependence, multiple causes explaining a single outcome, the same outcome being achieved through different combinations of causes, the same set of causes producing different outcomes in different settings, and nonproportionality—the notion that small interventions can result in large effects (the "tipping point") or large interventions can produce very small effects (inertia).

From an evaluation perspective, a key task is making sense of the complexity in processes of change that may be influenced by a development intervention. Sometimes this involves striking a difficult balance between empirical reality (the potential direct and indirect effect to which an intervention may contribute) and stakeholder priorities (see Leeuw & Vaessen, 2009). Again, the reconstruction of the theory of change is a good way to start. By reconstructing the causal assumptions connecting different components of an intervention to processes of change, one can start to build a picture of the nature and level of effects of an intervention. From the point of view of the policymaker and program manager, it is important to distinguish between processes they can control or influence and those they cannot. Finally, the inherent diversity in the nature of change processes at different levels (individual, institution, community, ecosystem, etc.), evaluation designs should encompass multiple methods of data collection and analysis.

Dimension 4: Embeddedness and the Nature of the System

Development interventions, the communities and regions in which they are implemented, the groups in society that are (in)directly affected by them, and the stakeholders involved in funding and implementing the interventions do not exist in a vacuum. They exist in contexts with particular economic, political, and social characteristics; they are affected by the divergent norms, beliefs, and opinions of different stakeholder groups. To illustrate, two projects on sexual and reproductive health, which on paper look quite the same in terms of budget, activities, content, and so on and are implemented in different regions, can turn out to be completely different interventions in terms of the way they are perceived and acted on by target groups and health workers. For example, the local gender norms regarding a woman's right to make decisions about her health-seeking behavior, the involvement of husbands in health consultation processes, the prevalence of extramarital relationships, and the use of birth control measures are but a few factors that define the context of such a project.

BOX 1.8 QUESTIONS TO BE ADDRESSED CONCERNING DIMENSION 4

With regard to this dimension, one can ask the following questions:

– What are the key historical, economic, political, sociocultural, administrative and organizational, climatic and ecological, legal and regulatory contextual factors that have a bearing on a particular intervention?
– What are the prevailing norms and beliefs of key stakeholder groups or specific actors that may have a bearing on the intervention?
– What are the boundaries of the intervention system? More specifically, what is the sphere of influence of an intervention? At what levels (local, regional, national, international) can one expect the process of change to occur?

Dimension 5: The Evaluation Process

The interplay between the complexity of the intervention, the institutional context, the (potential) processes of change influenced by an intervention, and the way these three dimensions are embedded in broader economic, social, cultural, and political systems affects the opportunity space for evaluators (and other evaluation stakeholders[8]). What creative solutions can evaluators (and other evaluation stakeholders) employ to improve their understanding and delimitation of intervention contexts and the design, implementation, and use of evaluations?

BOX 1.9 QUESTIONS TO BE ADDRESSED CONCERNING DIMENSION 5

With regard to this dimension, one can ask the following questions:

– What is the purpose of the evaluation? To what extent should complexity be addressed to fulfill the expectations of evaluation stakeholders?
– What are the time, data, and resource constraints that affect the design and implementation of the evaluation? To what extent can complexity challenges be addressed given these constraints?
– What combination of methods can be used to adequately address pertinent complexity challenges? What (innovative) methods can be used to adequately address the challenges of attribution and contribution in evaluation?
– What mechanisms are used to involve stakeholders in the process? What are the costs and benefits of participatory evaluation processes in a specific evaluation context?
– What are the explicit and implicit values of evaluators determining the characteristics of the evaluation? To what extent do they coincide with or diverge from the values of other stakeholders? What are the ethical obligations of evaluators and other stakeholders in a particular evaluation context?

Table 1.1 presents examples of six programs (four included in this book) to illustrate the first four dimensions of complexity. The fifth dimension—how the specific constraints and opportunities of the evaluation context together with an assessment of the first four dimensions may result in a complexity-responsive evaluation design and implementation—is a recurrent aspect that is discussed in most chapters of the book.

Table 1.1 Examples of Evaluations That Address Different Dimensions of Complexity		
Evaluation	**Source**	**Examples of different dimensions of complexity**
1. A 10-year program to combat gender-based violence in El Salvador	Discussed above and in Chapter 15	*Embeddedness and the nature of the system:* Gender-based violence is caused and perpetuated by the high levels of violence following civil war, the growth of gangs, and historical, cultural, and legal factors.
2. Coordination of international programs to combat human trafficking	Chapter 17	*Institutions and stakeholders:* There were multiple problems relating to difficulties of coordination among many different agencies, with different agendas and operating procedures and often working with limited resources.
3. General budget support to seven developing countries	Chapter 19	*Causality and change:* There is no generally accepted theory of change to track the complex processes of change with multiple and different sets of factors in each country. Additional challenges for causal analysis include contagion of effects from other programs, definition of boundaries, and diverging country contexts.
4. A systematic review of microcredit programs and women's empowerment	Chapter 16	*Causality and change:* Change and resistance to change are affected by multiple cultural, political, historical, institutional, and economic factors, making it very difficult to isolate the effect of a particular intervention (microcredit). Also, change takes place over a long period of time and often cannot be captured within the evaluation time horizon.
5. A multi-year, multi-agency program to make Ontario tobacco free	Schwartz & Garcia (2011)	*Complexity in the intervention:* The program included 12 prevention areas, each managed by a number of different agencies. Many of these were new and technically challenging, and often available documentation was limited.
6. Evaluating complex strategic development interventions: the challenge of child labor	Perrin & Wichmand (2011)	*Complexity in the intervention:* Fourteen activities to combat child labor are identified, and many programs include several of these activities, each with a different implementation process of affecting child labor over different time horizons and at different levels, many of which are interlinked.

SOURCE: All of these examples come from either this book (chapter number is given) or Forss et al. (2011).

4.2 A Checklist for Assessing the Level of Complexity of a Program or Intervention

In the previous section we identified four dimensions of complexity relating to an intervention and its context. Each of the four dimensions includes a number of categories. *Complexity* is a relative term, and a program and its context may be characterized as having a high degree of complexity on some categories but a low degree on others. The checklist presented in Table 1.2 describes for each category what would be considered a low level and what would be considered a high level of complexity. These low and high levels are derived from the framework discussed earlier. A program and its context can be rated on an ordinal scale from 1 (*low*) to 5 (*high*) for each category. It is also possible to use a summary rating for each dimension.

Table 1.2 Checklist for Assessing Levels of Complexity							
	Level of Complexity						
	Low	←		→			**High**
		Complexity rating					
		1	2	3	4	5	
Dimension 1: The nature of the intervention							
1. Objectives	Few and relatively clearly defined						Multiple, broad, and often not clearly defined
2. Size	Affecting small population						Affecting large population
3. Stability of program design	Relatively stable						Emergent design
4. Implementation procedures	Clearly defined in project design						Often not clearly defined and changing
5. Services or components	Relatively few						Large number
6. Technical complexity	Low						High
7. Social complexity	Low						High
8. Duration	Clear start and end dates						No clear end date and sometimes no clear start date
9. Testing of program design	Well tested and used many times						Relatively new and untested

		Level of Complexity						
	Low	\u2190				\u2192	**High**	
		Complexity rating						
		1	**2**	**3**	**4**	**5**		
Dimension 2: Institutions and stakeholders								
10. Budget	Use of funds clearly defined						General budget support with no clear definition of services to be funded	
11. Funding and implementing agencies	Relatively few						Large number	
12. Stakeholders	Relatively few and with similar interests						Many and diverse	
Dimension 3: Causality and change								
13. Causal pathways	Single causal pathway						Multiple causal pathways (nonlinear, interconnected, recursive feedback loops)	
14. Certainty on outcomes	Relatively high degree of certainty						Low degree of certainty	
15. Agreement and clarity on appropriate actions to address problems	High level of agreement and clarity						Low level of agreement and clarity	
Dimension 4: Embeddedness and the nature of the system								
16. Clarity on contextual factors that influence implementation and change	Contextual factors known and knowledge on their potential influence available						Contextual factors and their potential influence on implementation and change largely unknown	
17. Processes of behavioral change	Simple processes that are well understood						Multiple mechanisms to promote complex behavioral change	

The checklist requires field testing and refinement and can of course be adapted to the requirements of a particular program that is being evaluated. However, the practical value of this framework is that it provides guidance to an agency in deciding whether a program and its context should be considered as having a high, medium, or low level of complexity for the purpose of deciding which evaluation design to use. Further discussion of these issues is introduced in subsequent chapters, especially Chapters 3 and 7, which provide more detailed guidance on how to deal with complexity issues. In general, the lower the level of complexity, the more feasible it is to use established evaluation designs; the higher the level of complexity, the greater the need to develop a complexity-responsive evaluation design.

A point of clarification is in order when assessing the level of complexity. Some scholars may argue that all programs and the way they are embedded in multilayered societal processes, structures, belief systems, and so on are complex. We do not fundamentally contest this point of view. For us, and for the purposes of evaluation, the key question is: How complex are a program and its context given the purposes and (potential) uses of the evaluation? This is a rather different question; it is about how much of the underlying complexity surrounding programs should be taken on board to produce credible evaluation findings and recommendations that serve their purpose and use. In sum, assessing complexity is about taking into account the operationally relevant level of complexity of different characteristics of interventions and their contexts (see Figure 1.1.) while at the same time keeping in mind the question of how complexity-informed the evaluation should be, given its purpose and use.

5. Practical Applications

The following are some practical guidelines for agencies commissioning and managing evaluations of programs and interventions that may have complex elements.

- It is important to be familiar with the basic concepts of complexity. Not all programs are equally complex, but there is sometimes a tendency for agencies or evaluators to believe that their programs are very complex. So before commissioning an evaluation, it is important to understand to what extent and in what ways a program and its context are complex and how important the dimensions of complexity are in practical terms. While a theorist can always show that a program has many dimensions of complexity, these may not be important from an operational or policy perspective.
- Our message is not that complexity can be ignored, but rather that it can be introduced in a phased approach. For example, in the first phase of an evaluation, careful judgment is required to decide to what extent different dimensions of complexity should be addressed given stakeholder priorities and resources. On the basis of discussions between evaluators and stakeholders about the needs for further information for program improvement

or roll-out (e.g., scaling up, replication), and gaps in knowledge about the effectiveness of the program in particular settings, a more elaborate complexity-responsive evaluation can be designed in phase 2.

- Table 1.2 presents a checklist for assessing the level of complexity of a program and its context on the different dimensions of complexity. The checklist can be used to determine the extent to which and in what ways the program and its context should be considered complex from an evaluation perspective. Box 1.10 presents an overview of the main frameworks of guidance on complexity-responsive evaluation that are discussed in this book.

BOX 1.10 GUIDANCE ON THE DESIGN OF COMPLEXITY-RESPONSIVE EVALUATIONS

In several chapters we provide guidance on complexity-responsive evaluation. The reader may wish to consult the following three main frames of reference:

- Table 1.2 in this chapter provides an initial complexity checklist that helps decision makers and evaluators assess the extent and nature of complexity of a program and its context.
- Chapter 3 presents a step-by-step approach to managing a complexity-responsive evaluation process (see Figure 3.1 for a summary).
- Chapter 7 discusses a five-level approach to unpacking complex programs into evaluable parts and reassembling them into an overall evaluative perspective (see Figure 7.1 for a summary).

References

Bamberger, M., Rugh, J., & Mabry, L. (2012). *RealWorld Evaluation: Working under budget, time, data, and political constraints* (2nd ed.). Thousand Oaks, CA: Sage.

Byrne, D. (2013). Evaluating complex social interventions in a complex world. *Evaluation, 19,* 217–228.

Byrne, D., & Callaghan, G. (2014). *Complexity theory and the social sciences: The state of the art.* Milton Park, UK: Routledge.

Evaluation Gap Working Group. (2006). *When will we ever learn? Improving lives through impact evaluation.* Washington, DC: Center for Global Development.

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

Glouberman, S., & Zimmerman, B. (2002). Complicated and complex systems: What would successful reform of Medicare look like? (Discussion Paper No. 8). Retrieved from http://c.ymcdn.com/sites/www.plexusinstitute.org/resource/collection/6528ED29-9907-4BC7-8D00-8DC907679FED/ComplicatedAndComplexSystems-ZimmermanReport_Medicare_reform.pdf

Hansen, H. F., & Rieper, O. (2010). Institutionalization of second-order evidence-producing organizations. In O. Rieper, F. L. Leeuw, & T. Ling (Eds.), *The evidence book: Concepts, generation, and use of evidence.* New Brunswick, NJ: Transaction.

Kusek, J., & Rist, R. (2004). *Ten steps to a results-based monitoring and evaluation system.* Washington, DC: World Bank.

Leeuw, F. L., & Vaessen, J. (2009). *Impact evaluations and development: NONIE guidance on impact evaluation.* Washington, DC: Network of Networks on Impact Evaluation.

Morin, E. (2008). *On complexity.* Cresskill, NJ: Hampton Press.

Morra Imas, L. G., & Rist, R. C. (2009). *The road to results: Designing and conducting effective development evaluations.* Washington, DC: World Bank.

North, D. C. (1990). *Institutions, institutional change and economic performance.* Cambridge, UK: Cambridge University Press.

Patton, M. Q. (2011). *Developmental evaluation: Applying complexity concepts to enhance innovation and use.* New York, NY: Guilford Press.

Patton, M. Q. (2015). Book review: Evaluating the complex: Attribution, contribution and beyond. *American Journal of Evaluation.* Advance online publication. doi:10.1177/1098214015569758

Pawson, R. (2013). *The science of evaluation: A realist manifesto.* Thousand Oaks, CA: Sage.

Pawson, R., & Tilley, N. (1997). *Realistic evaluation.* Thousand Oaks, CA: Sage.

Perrin, B., & Wichmand, P. (2011). Evaluating complex strategic development interventions: The challenge of child labor. In K. Forss, M. Marra, & R. Schwartz (Eds.), *Evaluating the complex: Attribution, contribution and beyond* (pp. 243–282). New Brunswick, NJ: Transaction.

Popay, J. (2006). *Moving beyond effectiveness: Methodological issues in the synthesis of diverse sources of evidence.* London, UK: National Institute for Health and Clinical Excellence.

Ramalingam, B., Jones, H., Toussaint, R., & Young, J. (2008). *Exploring the science of complexity: Ideas and implications for development of humanitarian efforts* (Working Paper 285). London, UK: Overseas Development Institute.

Reynolds, M. (2015). (Breaking) the iron triangle of evaluation. *IDS Bulletin, 46,* 71–86.

Rogers, P. J. (2008). Using Programme theory to evaluate complicated and complex aspects of interventions. *Evaluation, 14*(1), 29–48.

Rogers, P. J. (2011). Implications of Complicated and complex characteristics for key tasks in evaluation. In K. Forss, M. Marra, & R. Schwartz (Eds.), *Evaluating the complex: Attribution, contribution and beyond* (pp. 33–53). New Brunswick, NJ: Transaction.

Schwartz, R., & Garcia, J. (2011). Intervention path contribution analysis (IPCA) for complex strategy evaluation: Evaluating the smoke-free Ontario strategy. In K. Forss, M. Marra, & R. Schwartz (Eds.), *Evaluating the complex: Attribution, contribution and beyond* (pp. 187–208). New Brunswick, NJ: Transaction.

Williams, B., & Hummelbrunner, R. (2011). *Systems concepts in action: A practitioner's toolkit.* Stanford, CA: Stanford University Press.

Wolf-Branigin, M. (2013). *Using complexity theory for research and program evaluation.* Oxford, UK: Oxford University Press.

Notes

1. Change has a certain starting point, and the nature of that starting point (e.g., the different characteristics of institutions, rules, norms, and beliefs) determines the change process. The term *changes at the margin* is sometimes used to indicate that change is often not abrupt but gradually arises from deviations in a number of issues at the same time (for a discussion of this issue, see North, 1990).

2. http://www.unorg/millenniumgoals.

3. http://www.unctad.org/en/docs//tdxii_accra_accord_en.pdf.

4. For reasons of space and simplicity, some of the terms are explained slightly differently than in the original text, also taking into account the literature on institutional change in the field of institutional economics (e.g., North, 1990).

5. The seven dimensions of the VICTORE framework are volitions, implementation, context, time, outcomes, rivalry, and emergence (Pawson, 2013, ch. 5).

6. Emergence can be defined as a situation in which collective phenomena (e.g., mass migration, climate change) are the outcomes of interactions between individual elements of a system that are seemingly chaotic or random.

7. Something causes something else to happen, which in turn reinforces the initial causal factor. Feedback loops can be circular or more complex.

8. For example, donors, decision makers, program managers, beneficiaries.

CHAPTER 2

Toward More Complexity-Responsive Evaluations: Overview and Challenges

Estelle Raimondo, Jos Vaessen, Michael Bamberger

Dealing with complexity in development evaluation requires more than a good understanding of the various complexity dimensions that characterize an intervention and its context. It requires a set of methodological approaches that can address particular aspects of complexity in the evaluation design and implementation. A key question raised in this chapter is thus: How can established evaluation approaches be adapted to take account of complexity issues while operating within the real-world conditions of data, budget, and time constraints? We start by presenting a few scenarios under which development evaluations are typically conducted and discuss some of the challenges encountered in established evaluation approaches from the perspective of complexity. In the remainder of the chapter we explore several strategies for developing complexity-responsive evaluations. One of these strategies relies on approaches stemming from complexity science; another consists of strengthening established evaluation approaches to take better account of complexity.

In Chapter 1 we pointed out that complexity is a multifaceted construct with no single agreed-upon definition. This diversity of views on the nature of complexity translates into a range of possibilities for developing methodological approaches that are well suited to a complexity perspective. The idea that there is still no widely accepted methodological road map to deal with complexity poses a challenge and can be difficult to communicate to decision makers and managers. At the same time, the absence of such a road map opens the door to renewed creativity. For instance, in this chapter we will see that some methods stemming from complexity science can usefully contribute to making sense of certain complexity dimensions, such as nonlinear causal processes or systems relationships. Yet these techniques have their own limitations and might be less

helpful for dealing with other complexity dimensions, such as complex institutional processes or conflicting stakeholder relations. Moreover, the choice of evaluation design must often balance the need for methodological soundness and real-world challenges of time, resources, and data constraints while ensuring the use of evaluation findings.

To arrive at a complexity-responsive evaluation approach, a key question is thus: How can established evaluation methods be adapted to take account of complexity issues? We do not propose a major overhaul of the field of development evaluation to tackle complexity. Instead our approach is premised on the twin ideas that complex interventions do not always warrant complex evaluation designs and it is possible to strengthen established evaluation designs and techniques to make them complexity-responsive. In this chapter and the remainder of the book, we argue that all development evaluations should be complexity-responsive by meeting three minimum requirements: (1) demonstrate awareness of the larger system in which the intervention and its evaluation are embedded; (2) understand how the various dimensions of complexity will affect the design, implementation, and use of the evaluation; and (3) explicitly address and deal with some of the complexity dimensions in the evaluation design, process, and ultimately findings. While all evaluations should strive to be complexity-responsive, this need not necessarily be technically, institutionally, or financially demanding. Consequently, we refrain from using the term *complex evaluations*.

The next chapter goes one step further by discussing in more detail how evaluations can be designed to include complexity issues in a systematic and stepwise manner, taking due account of the demands and restrictions of a particular evaluation setting. Subsequent chapters discuss in detail a number of methodological approaches—some already known and used in development evaluation, others less known—that are well-equipped to address one or more of the complexity dimensions discussed in Chapter 1.

1. Common Scenarios in Development Evaluation

Development evaluation encompasses a broad field of practices in very divergent circumstances (e.g., in terms of stakeholders, nature of the evaluand, target group, etc.). At the risk of oversimplification we distinguish between five common evaluation scenarios. For lack of a better term, and because a unifying heuristic is sometimes necessary, in what follows we use the term *established evaluation approaches*, which share the following common features:

- They are well known within existing evaluation systems and build on existing evaluation capacities in development organizations.
- They are composed of well-known data collection and analysis methods.
- They offer the flexibility to be adjusted to real-world constraints regarding data availability, budget, and time restrictions and political challenges.
- They intend to provide information that is readily usable by program managers and decision makers.

The majority of development evaluations conducted to date fit into one or more of the following evaluation scenarios. These five scenarios do not provide a comprehensive picture of all development evaluations, and they sometimes overlap. But they paint in broad brush strokes some typical situations for development evaluation managers and practitioners.

Some of these scenarios have more latitude than others for developing and implementing an evaluation design that is amenable to dealing with complexity, but it is possible to address complexity within each scenario. The main issue is often to arbitrate between scope and depth.

- *Rapid evaluations:* These are conducted within a limited number of working days. While some of them may be well funded and deploy many evaluators concurrently in various places (most are conducted with limited resources), the rapidity of the exercise often means limited depth. Most commonly, these are retrospective evaluations typically relying on reviews of secondary data combined with key informant interviews and often short field visits to conduct case studies in a number of program locations through focus groups, in-depth interviews, or rapid surveys. Usually there are no systematic baseline data to draw on.
- *Large-scale, long-term evaluations:* These multi-year programs (e.g., programs to combat smoking, homelessness, or child labor) may include substantial funding and time for an evaluation. Many of these evaluations involve a number of stages to adapt the evaluation design as the program evolves and often to contribute to refining the program design and implementation strategy (formative evaluation). Typically, the evaluation involves ongoing consultations with stakeholders, and a significant amount of time is spent on developing and refining the program theory. These evaluations typically use multiple sources of evidence collected through a wide range of data collection methods.
- *Experimental manipulation and/or reliance on primary data collection:* In recent years there has been growing interest in the use of experimental (randomized controlled trials) and quasi-experimental designs to assess more systematically the effects of program interventions. There is a continuing debate about the feasibility and appropriateness of experimental approaches to program evaluation. However, it is generally agreed that experimental designs work best when complex interventions can be unpacked so that experimentation can be applied to a single component (see Chapter 7), but it is probably not possible to use these approaches in isolation to assess the overall impact of a complex, multi-component intervention (Vaessen, 2011). Other evaluation designs also rely heavily on primary data collection. These include ethnographic studies as well as quasi-experiments and statistical analysis based on survey data.
- *Systematic reviews[1]:* These synthesize the findings of existing evaluations (that meet certain standards of relevance and rigor) so as to provide an overall assessment of the theory of change or the impacts of an intervention (e.g., microcredit) on one or more outcomes (e.g., the creation of small enterprises, women's economic empowerment). While systematic reviews have great potential as a tool for addressing complexity in evaluation, they are best suited to assess the relationships between specific inputs

and specific outcomes, rather than to assess the combined effects of multiple inputs (including the interactions among inputs) on multiple outcomes.

- *Participatory evaluations:* Stakeholder participation is a potentially important aspect of any type of evaluation. Over time, a set of evaluation approaches was developed that puts participation at the center of evaluative inquiry. Participation in evaluation can serve multiple purposes. First, it can improve the quality and depth of data collection. Second, participatory models are often used to enhance the relevance, ownership, and use of evaluations by stakeholders. Finally, participatory evaluations can also support transformative processes such as stakeholder empowerment and control over the program that is the subject of the evaluation (see, e.g., King, Cousins, & Whitmore, 2007). While potentially offering a number of distinct advantages, participatory models of evaluation are sometimes also criticized for their costliness or susceptibility to bias.[2]

Table 2.1 provides more background information on the five evaluation scenarios.

Table 2.1 Five Common Scenarios for Development Evaluation		
Scenario	**Common features**	**Examples**
1. Rapid evaluation	• Short time frame for conducting the evaluation and limited resources • Usually retrospective • Limited access to baseline data • Relies heavily on secondary data and key informant interviews • May include case studies on a few countries or program sites • Makes use of rating scales (e.g., OECD-DAC)	• SIDA evaluation of its HIV/AIDS programs • UNICEF evaluation of its HIV/AIDS programs; UNESCO evaluation of its HIV/AIDS program • IEG Project Performance Assessment Reports.
2. Large-scale, long-term evaluation	• Design evolves and changes as program evolves and is better understood • Often has formative function to help refine the definition of program objectives and implementation • High level of stakeholder involvement • Investment of time in developing program theory model • Multiple streams and sources of data	• Poverty impact of European support to Tanzania (Toulemonde, Carpenter, & Raffier, 2011) • Multi-agency response to homelessness in Australia (Wilkins, 2011) • Swiss smoking prevention program (Spinatsch, 2011) • Smoke-free Ontario strategy (Schwartz & Garcia, 2011) • Child labor prevention (Perrin & Wichmand, 2011) • Evaluating the effectiveness of weather-indexed crop insurance in Ethiopia (Bamberger, 2015)

(Continued)

Table 2.1 (Continued)		
Scenario	**Common features**	**Examples**
3. Experimental manipulation	• Randomized controlled trials when subjects can be randomly assigned to treatment and control groups • Quasi-experimental designs when randomization is not possible but when treatment and comparison groups can be statistically matched • Works better for assessing effects of single treatments on individual outcomes but not to assess influence of multiple inputs on multiple outcomes • Design is often combined with other (qualitative) methods	• Deworming in Kenya (Miguel & Kramer, 2004) • Education in India (Banerjee, Cole, Duflo & Linden, 2007) • Monitoring corruption in Indonesia (Olken, 2007) • Returns to capital and access to finance in Mexico (McKenzie & Woodruff, 2008) • Microcredit in the Philippines (Karlan & Zinman, 2007) • Indian Panchayat Reform Program (Bamberger et al., 2010)
4. Systematic review	• Review of all evaluations that meet acceptable standards of rigor and relevance • Current methodologies focus on assessing effects of specific interventions on specific outcomes • Difficult to apply to assessing effects of multiple inputs on multiple outcomes • Can be combined with theory-based approaches to review and synthesis	• Effects of conditional cash transfers on child nutrition in Latin America (Leroy, Ruel, & Verhofstadt, 2009) • Effects of microcredit, micro-savings, and micro-leasing on promoting access of poor people and especially women to economic opportunities (Stewart el al., 2012)
5. Participatory evaluation	• Often used in evaluations with a formative purpose • Often used in community-based programs • Can be costly to implement in evaluations of multi-site, multi-actor interventions	• Supporting indigenous governance in Colombia (Estrella & Gaventa, 1998) • Various examples (Chambers, 2007) • Use of empowerment evaluation to build the capacity of Aymara women artisans in Puno, Peru (Sastre-Merino, Vidueira, Diasz-Puente, & Fernandez-Mora, 2014)

2. Limitations and Challenges Encountered in Established Evaluation Approaches From the Perspective of Complexity

As argued by Mowles (2014), many evaluations that claim to be sensitive to complexity issues are in fact paying lip service to it. In other words, there is often a disconnect between the theoretical discussion of complexity and the approaches that are actually used to address complexity in evaluation (Forss & Schwartz, 2011; Funnell & Rogers, 2011; Patton, 2011; Rogers, 2011). What follows are some

examples of common pitfalls encountered in established evaluation approaches from the perspective of dealing with complexity. Broadly, they fit into two categories: overlooking and oversimplifying some dimensions of complexity.

2.1 Overlooking Complexity

- Most development evaluations do not address the issues discussed in the literature on the nature of complexity.
- Most program theory models (see Chapters 5 and 6) do not adequately address complexity and are developed and presented in similar ways for simple and complex interventions.
- Methodologies for identifying unanticipated outcomes are not incorporated into most evaluation designs.
- Differences between target populations and institutional contexts across interventions pose particular challenges for the generalizability of evaluation findings; these challenges are often disregarded in established evaluations (Pritchett & Sandefur, 2013).

2.2 Oversimplifying Complexity

- Many established evaluation approaches are intrinsically reductionist; they harvest particular variables from a complex context and end up decontextualizing the intervention while focusing on the effect of a controlled variation in one part (i.e., an intervention) on specific outcomes. Yet as highlighted by Byrne and Callaghan (2014, p. 173), the presence of emergence implies that such reductionist approaches are missing a vital part of the processes of change and effects of an intervention.[3] In Chapter 7 we discuss the circumstances under which such deconstructions are feasible.
- The power dynamics, the conflicting perceptions, and the changing nature of interactions among many different stakeholders and actors are rarely addressed in evaluations and the focus tends to be on the influence of a small number of stakeholders, each examined individually.
- Most designs do not adequately address the problem of assessing changes that take place over long periods of time. Woolcock (2013), among others, argues that the important piece of information that evaluation should bring to the program manager is not an estimate of impact at a given point in time but rather an accurate picture of the likely impact trajectory over time. While most established evaluation approaches implicitly assume that the impact trajectory is linear, in reality most development interventions have an impact trajectory that either follows a J-curve (things get worse before improving) or has a tipping point (i.e., the idea that things do not change until they dramatically do). Moreover, development interventions seeking to be transformational at the level of norms and culture (e.g., interventions seeking to change gender relations) are subject to particularly long causal pathways of change.

The challenges presented above are illustrative of the limitations of established evaluation approaches from the perspective of complexity. Under certain

circumstances, methodologies emerging from complexity sciences can be usefully leveraged to address some of the complexity dimensions of our framework. Below we introduce some of these approaches and discuss their respective strengths and limitations.

3. Methodological Approaches Stemming From a Complexity Science Perspective

Complexity science is an umbrella term encompassing a vast and eclectic field of academic inquiry with roots in various philosophical branches and disciplines as well as many ramifications in the social sciences. In the field of complexity science a number of well-defined methods for studying complexity have been developed, ranging from network analysis to soft systems methodologies. Most of these approaches share the objective of modeling multi-component systems that organize spontaneously. In recent years, many social science disciplines have attempted to adapt some of these methods of inquiry to the study of the social world. The work of the Santa Fe Institute of Science for a Complex World, for example, is particularly praiseworthy in its endeavor to provide insights into the complexity of human behavior, institutions, and social systems.

3.1 Succinct Overview of Complexity Science Approaches

As usefully summarized by Williams (2015), the development of the field of complexity science can be characterized by three main waves. The first wave, dating back to the 1960s, was mostly concerned with dynamic interrelationships and has inspired methods such as systems dynamics and complex adaptive systems modeling. The second wave, starting in the 1970s, paid closer attention to the issue of multiple perspectives, recognizing that there were often many ways of understanding each relationship in a given system. This wave inspired the soft systems methods. The third and most recent wave emphasizes the importance of delineating boundaries to otherwise limitless systems; it gave rise to critical systems thinking (Williams, 2015, p. 9). Given the variety of models and techniques that are associated with complexity science, our goal here is not to be exhaustive but rather to present a sample of approaches that have some degree of compatibility with evaluation and the potential to help address some of the complexity dimensions covered in this book's framework (laid out in Chapter 1). Additionally, we attempt to represent the diversity within complexity science. Table 2.3, later in the chapter, summarizes these approaches. Below we discuss various methodological approaches, organized by complexity dimension (see Chapter 1).[4]

a. Dimension 1: Nature of the Intervention: Dealing With Complex Theories of Change

- **Systems Mapping**

Systems mapping is an umbrella term to describe a range of methods aimed at providing a visual representation of a system. Systems mapping helps identify the various parts of a system as well as the links between these parts that are likely to change. In a subsequent step, the evaluator attempts to find measurement methods to

check whether the anticipated system changes have actually occurred. Systems maps are closely related to theories of change (TOC), but they differ from the majority of TOC and logic models by doing away with the assumption of successionist causal relationships[5] between inputs, outputs, and outcomes. Systems mapping can be done via a participatory process. For instance, CARE International recently engaged in an overall assessment of its internal systems for collecting, brokering, and transferring information and used systems mapping to document and monitor its envisaged systems change. Three maps were produced by a consulting firm: a baseline map, an intervention map, and a conclusion map. Figure 2.1 displays the intervention map. The advantage of such a map is that several dimensions of a theory of change can be represented at the same time, and complex connections at several levels of intervention can be illustrated. For example, the map in Figure 2.1 displays the CARE USA level along with the Regional level and the Country level, and the arrows show how these different levels interact in top-down, bottom-up, and sideways directions. External actors with an influence on the strategy can also be integrated in the systems, in this example, the Media/Public as well as U.S. government policymakers.

b. Dimension 2: Institutions and Stakeholders: Dealing With a Large Number and Diversity of Stakeholders

- **Social Network Analysis**

Social network analysis (SNA) is particularly useful for modeling complex stakeholder relationships and describing how information flows and decisions are made, including through the dynamic use of feedback. SNA studies the connections within systems and acknowledges that social units are interdependent and have multiple layers of relationships by relying on concepts stemming from graph theory and matrix operations (Wolf-Branigin, 2013, pp. 64–65). Drew, Aggleton, Chalmers, and Wood (2011) report on their experience with using SNA to evaluate a particularly complex intervention in the framework of the Global Dialogue on Sexual Health and Well-Being. At the core of this large program lies a network of individuals, the Sexual Policy Watch, based in Rio de Janeiro and composed of key informants who are connected to various organizations. The evaluators thus resorted to SNA to identify (1) the organizations that each key informant was involved in, (2) the geographical and thematic focus of their work, and (3) their links to up to 10 other key informants working in the field of sexual health. The mapping process was produced iteratively with various informants, entered into a database, and analyzed through tables of attributes. The process helped to visualize the various circles of influence and to get a sense of the magnitude of the outreach of the program.

c. Dimension 3: Causality and Change: Dealing With Various (Nested) Levels of Effects

- **Agent-Based Modeling**

Various agent-based modeling (ABM) techniques have been developed to accommodate for multiple levels of reality (e.g., the macro-level of the external environment of the system, the meso-level of an organization, the micro-level of

34

Figure 2.1 Example of a Systems Map
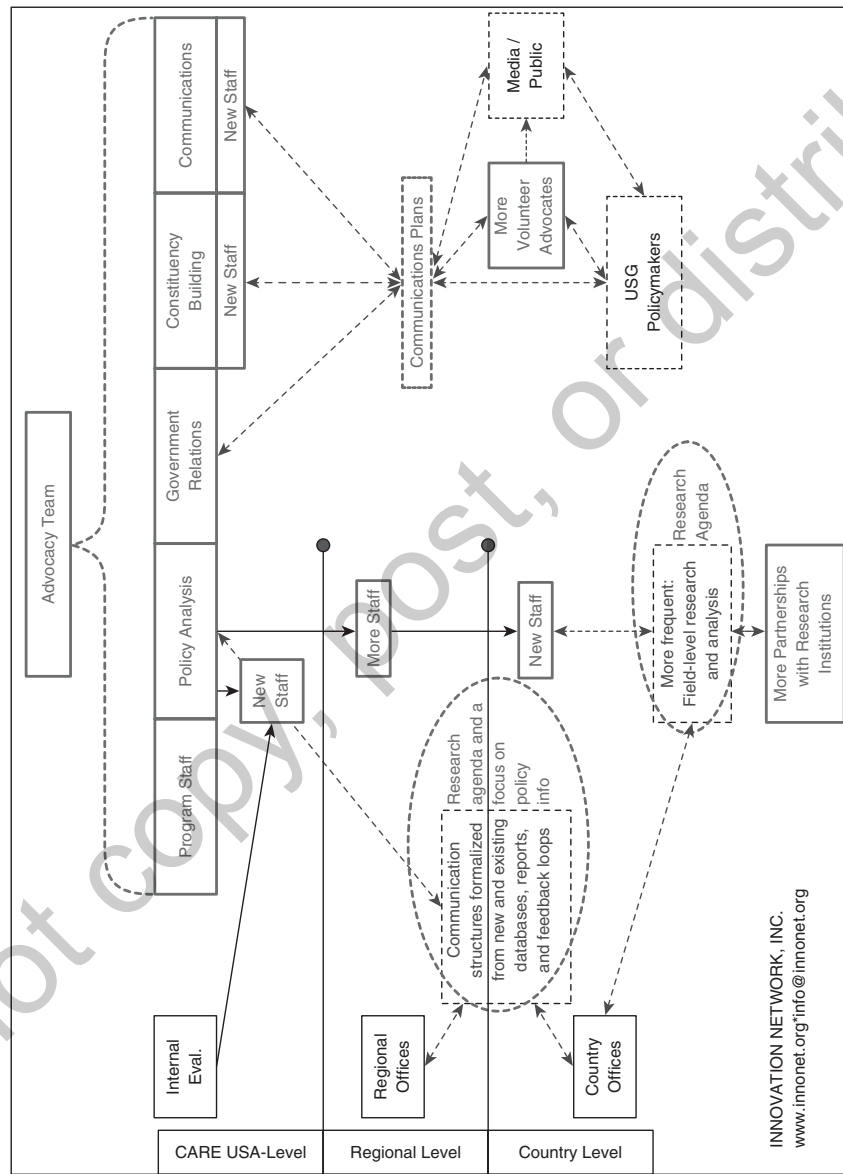
SOURCE: Innovation Network (n.d.).

individual behavioral characteristics; Byrne & Callaghan, 2014). ABM consists of computationally simulating the interactions, preferences, and characteristics of individual agents to assess their effect on entire systems. While ABM has not yet been widely applied as an evaluation tool, it has proven useful to analyze program implementation in which actors go through repeated interactions with the liberty to pick their strategy (Hermans, Cunningham, & Slinger, 2014). ABM offers a formal approach that forces evaluators to simplify the description of social interactions down to their most basic features. For example, ABM has been used to assess the effectiveness of a greenbelt at the border between a rural and urban area as a way to delay the urbanization beyond a particular fringe and to preserve environmental diversity outside the city (Brown, Page, Riolo, & Rand, 2004). A model was developed with simple assumptions about the trade-off between the greenbelt placement, its effect on the development rate beyond the belt, and how these factors interact with agents' heterogeneous characteristics and preferences as well as their varied but imperfect level of information.

d. Dimension 4: Embeddedness and the Nature of the System: Dealing With Interconnectedness and Dynamic Boundaries

- **System Dynamics**

System dynamics (SD) is an approach for capturing the dynamic behavior of (social) systems. It borrows from engineering and management and relies on the analysis of relationships between stock and flow variables that are quantified. A simulation model is then crafted by assembling all the elements to express the various relations as mathematical equations.

Applied to evaluation, SD can be particularly useful when attempting to capture how complex systems are affected by development interventions and affect these interventions in return. SD can help improve our understanding of dynamic interrelationship along the theory of change (Befani, Ramalingam, & Stern, 2015). For instance, Williams and Hummelbrunner (2011) describe the use of SD in the evaluation of a microloan project targeting sex workers in a broader HIV/AIDS prevention program in a mining area of West Africa. SD was used to explain why the popularity of the microloan schemes would swing dramatically among the target population of sex workers depending on a range of factors (Williams & Hummelbrunner, 2011, p. 49). Three stocks were identified: the popularity of the program, the number of loans, and the financial liability of the funding company. The only flow of the scheme was the rate of new loans issuance. The evaluators proceeded with several simulations that they compared to real-life patterns. A number of lessons were learned, notably that the system will improve with time even if nothing is done; improving the performance of the funder seems more effective than improving the communication with beneficiaries.

e. Dimension 4: Embeddedness and the Nature of the System: Dealing With Embedded Norms and Beliefs

- **Critical Systems Heuristics**

Critical systems heuristics (CSH) is premised on the idea that the decision of what lies within the boundary of a system determines the focus of attention, while

the rest is put aside. This decision is principally an ethical decision involving value judgments. Being reflexive about where one draws the line is the core of critical systems approaches (Hummelbrunner, 2011). The definition of a critical system revolves around four main categories: motivation, control, expertise, and legitimacy. Table 2.2 summarizes the example of the CSH application described by Williams and Hummelbrunner (2011). It is based on the abovementioned intervention aimed at reducing the incidence of HIV in African mining communities where there is a large population of sex workers. As described by the authors, the intervention had three main goals: (1) ensure that sex workers know about sexually transmitted infections

Table 2.2 Example of Critical Systems Heuristic Application

Motivation (value basis)	There is a range of embedded assumptions about the appropriateness of sex work as a commercial activity, the appropriateness of condom use, and so on that might have been patronizing to the beneficiaries.
Purpose of the system	Health education and economic empowerment
Beneficiaries of the system's purpose	Commercial and occasional sex workers in mining towns
Indicators of success	• Less reliance on commercial sex for income • Target group access to health facilities • Consistent use of condoms • Increased awareness of modes of HIV transmissions among sex workers
Control (power basis)	While peer educators and health workers were key actors in the intervention, they had little power over the resources they needed, leading, for example, to shortages of condoms and little uptake of distributed drugs because the health workers were instructed to sell the drugs and not provide them for free, as the donors' assumption was that patients would use them better if they had to pay for them.
Expertise or knowledge (knowledge basis)	There was a clear expert bias in both the design and implementation. This limited beneficiaries' involvement from the beginning and left stringent assumptions unchecked: • that sex workers had no knowledge of STI and HIV transmissions, when in fact they did; beneficiaries considered the educational material irrelevant • that sex workers were better positioned to distribute condoms in the community when it turned out that the men (clients) were better positioned to make the distribution
Legitimacy	Certain boundary decisions were taken too narrowly and uncritically. For instance, the focus was on street-based commercial sex workers, leaving out occasional sex workers and sex workers based in hotels. The normative worldview of the program was reflective of the funder's take on the legitimacy or lack thereof of sex work.

SOURCE: Adapted from Williams & Hummelbrunner (2011, pp. 310–315).

(STIs), HIV/AIDS, and reproductive health services through a peer education project, (2) offer alternative sources of income through a microcredit scheme, and (3) provide access to reproductive health services. However, while the project's theory of change sounded quite plausible, the empirical reality proved otherwise. CSH was used to make sense of what went wrong and to better understand whether the lack of success was mostly due to implementation, or design failure as a result of boundary definition issues.

3.2 Applicability of Complexity Science Approaches in Development Evaluation

Some of the approaches presented above have started to make their way into the methodological repertoire of development evaluators. Systems mapping, for instance, is now routinely used by evaluators faced with complex interventions in lieu of a more traditional theory of change. For instance, the Global Environmental Facility (GEF) has developed a general theory of change for its role and intended impact on the environment that leverages several concepts of systems thinking as presented in Figure 2.2. The map features, among other things, the display of feedback loops. In the area of GEF contribution, the knowledge and information, implementation strategies, and institutional capacity are meant to build on each other in a positive reinforcement loop. Likewise, the larger transformational process can be understood as a causal loop whereby behavioral changes are encouraged by and feed into successes, mainstreaming environmental issues and replicating successful interventions. Finally, the arrow that gets wider and wider illustrates that change is often nonproportional when a positive feedback loop is activated.

Methods narrowly associated with a complexity science perspective are particularly instrumental in pushing evaluators to focus on the nature of relationships and interdependences within systems, to lay out and explain the multidirectional linkages between a given intervention and its outcomes, whether intended or not. Some of these tools are also proving to be useful in dealing with the fact that some interventions are embedded (nested) in broader policy frameworks. One of the main challenges faced by evaluators dealing with complexity is to understand how different components interact the way they do and why some interactions lead to positive change, while others lead to stasis or negative (unintended) change. Tools inspired by systems thinking can help answer these questions. Table 2.3 summarizes the comparative advantages of these various complexity science approaches in dealing with particular dimensions of complexity and how they can usefully be applied to evaluation.

However, while these and other complexity science approaches are quite promising for the field of development evaluation, they are not silver bullets, and as with any other methods, there are notable limitations:

- One of the main issues with some of the systems approaches described above is that they often result in a high level of abstraction, primarily relying on imagined properties of agents and imagined rules describing their behavior that are only loosely connected with real data (Byrne & Callaghan, 2014).

Figure 2.2 The Outcome to Impact Pathway of the Global Environmental Facility
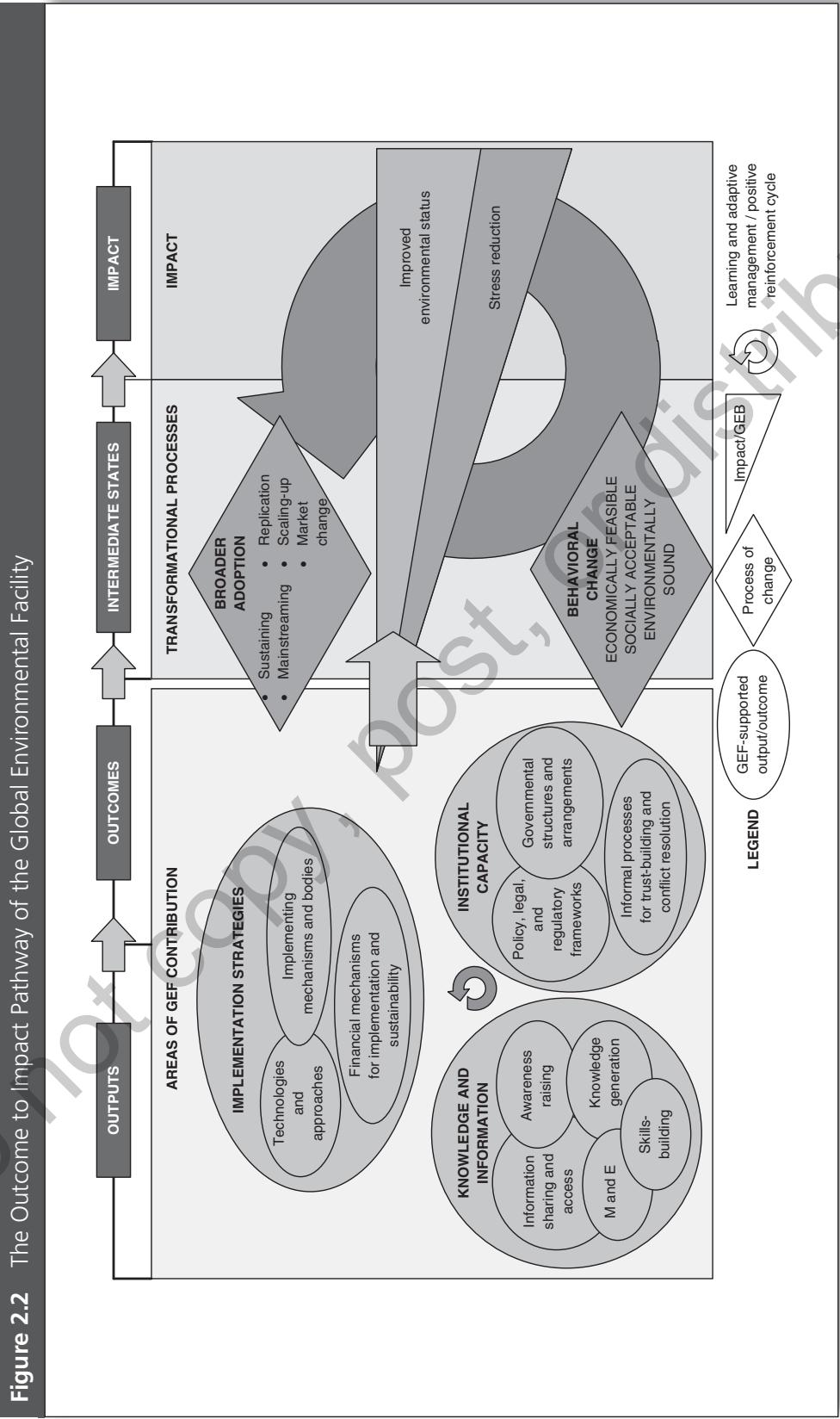
SOURCE: GEF Independent Evaluation Office (2013).

Table 2.3 Possible Uses of Complexity Science Approaches in Evaluation		
Methods	**Dimensions of complexity**	**Use in evaluation**
System dynamics	Dealing with interconnectedness and dynamic interrelationships	• Map the program context • Test assumptions underlying a particular causal chain within a theory of change • Assess the change in the state of a system based on different hypotheses about the dynamics of an intervention
Critical systems heuristics	Dealing with embedded norms, beliefs and values	• Elicit particular motivations, values, and perspectives held by a range of program stakeholders • Assess how these different values and perspectives can lead to possible design failures
Systems mapping	Dealing with complex theories of change	• Map the various complexity dimensions of an intervention and its context • As the basis for pre and post comparison (comparing a systems map at the end of an intervention to a systems map prior to the start of an intervention)
Social network analysis	Dealing with a large number of stakeholders and their relationships	• Identify the multiple layers of relationships among various stakeholders • Explore the structures that form or are formed by networks • Assess how different patterns of networks illustrate differential levels of capacity, communication, resource movement, etc.
Agent-based modeling	Dealing with complex causal processes	• Model various levels of reality (e.g., macro level of society, meso-level of an organization, micro-level of a program) • Anticipate the outcome of a situation based on the simulation of interactions, preferences, and characteristics of individual agents

- While complexity science approaches are useful for diagnostic or planning studies and simulations can be used to reconstruct scenarios that can be used as counterfactuals, these techniques need to be combined with more established evaluation techniques in order to address the evaluation questions of what works in what context.
- The statistically based techniques often rely on a narrow definition of complexity and emergence as resulting from interactions among micro-level entities that follow certain rules. Consequently, these techniques are largely unable to deal with social forces that shape the context of the agents' actions (Byrne, 2013; Byrne & Ragin, 2009). Other systems methods such as systemic questioning, critical

systems heuristics, or soft systems methodology are better able to deal with the embeddedness of norms and values (Williams & Hummelbrunner, 2011).

- Many of the methods introduced above are often still the object of experimentation and need refinement. They are not yet ripe for practical applications in real-world evaluation settings.
- Some of the approaches also require a high level of proficiency in quantitative and computational techniques that many real-world evaluators do not possess. Moreover, the types of results that these approaches produce are not easily translatable in a language that decision makers and program managers can readily put to use.
- Most methods require a large amount of data, which are not available (or are too costly to acquire) in many real-world development evaluation settings.[6]

Given the above, one way forward is to promote the use of complexity science approaches and adapt them to development evaluation real-world conditions. For some methods (e.g., social network analysis) this is certainly feasible, yet in other cases (e.g., agent-based modeling) the applicability gap may currently be too wide. A different way forward is to start out from established development evaluation methods and discuss how they might be adapted to include some of the features of the approaches discussed above. A first step toward more complexity-responsive evaluations is discussed below.

4. Bridging the Gap: Principles for Addressing Complexity Issues in Established Evaluation Approaches
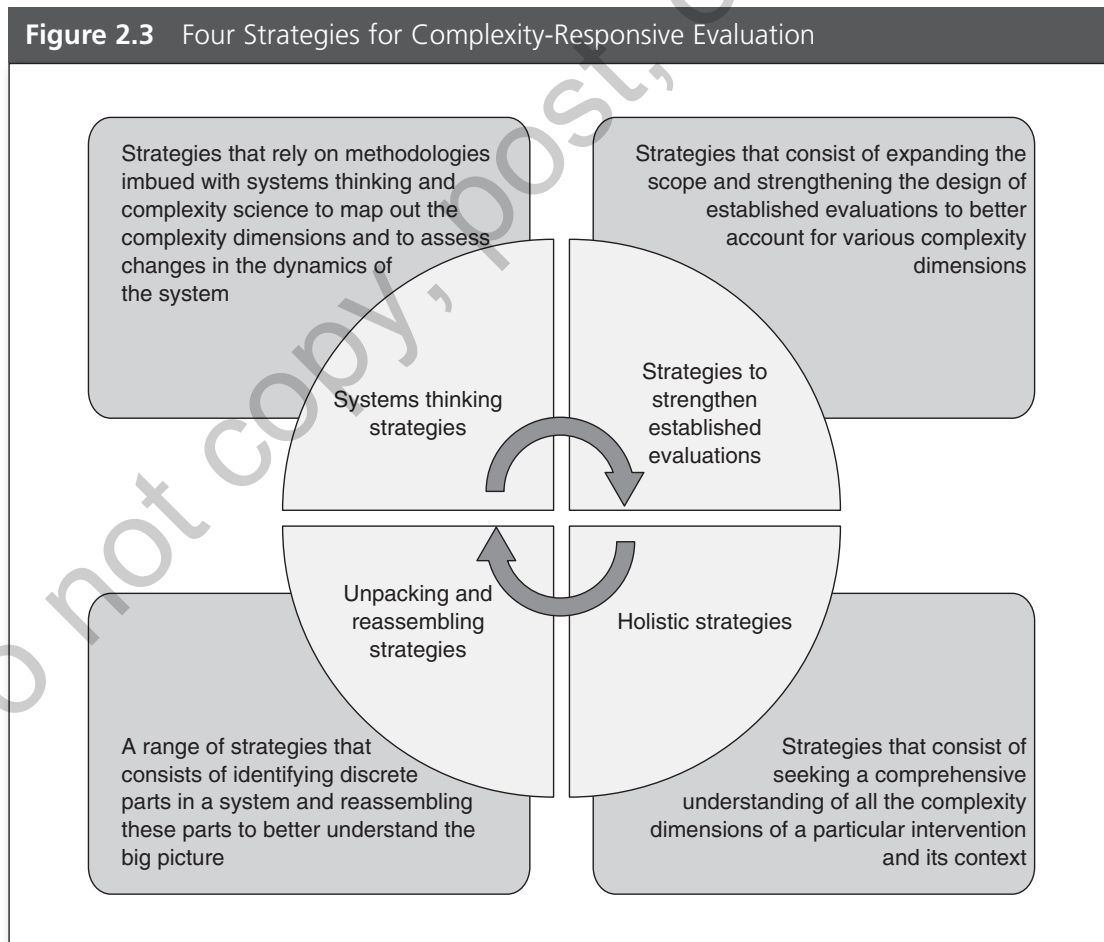
The overarching methodological approach of this book is based on the basic premise, raised in Chapter 1, that all interventions are inherently complex but the nature of complexity and the need for addressing complexity in particular ways (from an evaluation perspective) can widely differ. The practical and methodological implications are as follows: Evaluators always have to keep in mind that some elements of the evaluand, the context, and the processes of change induced by the intervention are complex and may require particular solutions that differ from business as usual (i.e., established approaches) and require a specific (methodological) reflection and response. This book presents a number of alternative and complementary methodological approaches to address different dimensions of complexity. The methodological approaches discussed in subsequent chapters can be defined as complexity-responsive and are based on one or more of the following strategies:

- Systems thinking strategies (some of which were presented in Section 3) to map out the complexity dimensions of an intervention and its context and to assess changes in the dynamics of the system
- Strategies that consist of strengthening established evaluations by expanding their scope and/or introducing novel data collection and analysis methods to better capture various complexity dimensions

- Unpacking and reassembling strategies that consist of breaking down systems into identifiable parts and reassembling them to better understand change in the big picture
- Holistic strategies that consist of seeking a comprehensive understanding of all the complexity dimensions of a particular intervention and its context

As illustrated in Figure 2.3, a complexity-responsive evaluation often builds on several of these strategies. For instance, in Chapter 7 we argue that a complexity-responsive evaluation that opts for unpacking interventions into various parts needs to build on a holistic view of the intervention and its context, both ex ante (when identifying the most relevant unit of analysis) and ex post (when assessing how much of the state of the system has been covered by the evaluation design).

Apart from being grounded in the four aforementioned approaches, complexity-responsive evaluations follow one or more of the following principles.

Figure 2.3 Four Strategies for Complexity-Responsive Evaluation

Strategies that rely on methodologies imbued with systems thinking and complexity science to map out the complexity dimensions and to assess changes in the dynamics of the system

Strategies that consist of expanding the scope and strengthening the design of established evaluations to better account for various complexity dimensions

Systems thinking strategies

Strategies to strengthen established evaluations

Unpacking and reassembling strategies

Holistic strategies

A range of strategies that consists of identifying discrete parts in a system and reassembling these parts to better understand the big picture

Strategies that consist of seeking a comprehensive understanding of all the complexity dimensions of a particular intervention and its context

4.1 Clarifying the Relevant Aspects of the Systems

Mapping the system is an essential aspect of an evaluation dealing with complexity. Looking at the complexity framework presented in Chapter 1, there are at least three major categories of systemic elements that should be articulated in a systems map: the nature of the intervention (e.g., the nature, level, and number of activities); the different stakeholders of the intervention and their interests/demands; the political, economic, cultural, and other factors that shape how individuals and institutions (inter)act, anticipate, and reflect about the intervention. All of these aspects are interrelated. Complex interventions tend to have more components and consequently more interactions among components and are thus exposed to a wider range of contextual factors. An articulation of the organizational structures and the relationships among different agencies and actors is thus warranted. Many evaluations also model power relations and pathways through which different partners can influence how the program is formulated and implemented, and how it achieves outcomes.

4.2 Relying on Theory-Based Principles (see Chapters 4, 5, and 6)

Interventions are theories and evaluations are the test. This well-known proverb points at the heart of evaluative inquiry, the reconstruction of causal assumptions that underlie an intervention. These assumptions can be based on stakeholder expectations, existing documentation, empirical data collection and analysis, or a combination thereof. Theory-based evaluation refers to those evaluations that in some way rely on such sets of assumptions (i.e., the theory of change or the intervention theory) as guiding principles of the evaluation exercise. To avoid bias (e.g., pro-intervention bias, confirmation bias of entrenched beliefs about effectiveness), theories should be perceived as "living" abstractions of reality—with new evidence coming in, theories can be further refined.

Theory-based evaluation approaches that focus on multiple causal pathways are particularly well suited to model the implementation design and the mechanisms through which outcomes are expected to be achieved. Part of the task of mapping complexity also includes a detailed explanation of how the program design is intended to achieve program objectives. In this regard, a fully specified theory of change can also profitably make use of a systems map as mentioned above. The many ways in which theory can be used in evaluation are discussed in subsequent chapters of the book.

4.3 Unpacking Complex Interventions Into Evaluable Parts (see Chapter 7)

Complex programs usually have many different components that are also implemented in different locations or through different partner agencies. This poses a key challenge in evaluation design. An evaluator can try to simplify or standardize

the design by unpacking the program so as to focus on key elements. As we further explain in Chapter 7, a key issue concerns the unit of analysis for deconstruction. Sometimes it makes sense to deconstruct a program into groups of similar activities (e.g., activities relating to the provision of monetary incentives versus activities relating to the transfer of knowledge). Another example is deconstruction into levels of intervention (e.g., local versus national versus international), all of which may exist within a complex intervention.

The other side of the coin of unpacking is reassembling, trying to make sense of the whole by combining information about its parts. Reassembling also refers to making sense of a particular part of a complex intervention (e.g., how does program coordination work in an interagency context? What are the effects of capacity development activities for primary school teachers?) across different intervention contexts. In other words, reassembling can relate to making sense of the complexity of different intervention components at the level of the overall intervention, or it can relate to looking at particular parts across interventions and saying something about how these parts work, for whom, and under what conditions.

4.4 Using a Mixed Methods Approach (see Chapter 8)

Given the many different kinds of intervention strategies, components, actors, and geographical or organizational locations, most complexity-responsive evaluations draw on multiple evidence streams and sources of data. In some cases there is a clearly defined mixed methods strategy in which quantitative and qualitative methods are integrated in a systematic way,[7] but in many cases the approach is more opportunistic, with evaluators drawing on whatever sources of primary and secondary data are available. Given a lack of standardized solutions for dealing with complexity, evaluators can use creative and flexible designs that evolve as the evaluation progresses. For instance, it is not uncommon to have multi-stage evaluations, often with emergent designs that are adapted to changes in how the program actually operates on the ground. This contrasts with many of the established quantitative approaches that tend to have a relatively inflexible design (e.g., with a pretest-posttest comparison group design). Additionally, one particular approach to evaluation is intently developmental and allows the program to evolve as the evaluation is ongoing, rather than requiring the program design to be frozen. The developmental evaluation approach relies on tight information loops that feed back into the program implementation process (Patton, 2011).

4.5 Dealing With Attribution (see Chapters 4, 5, and 6)

Attribution refers to the question of whether and the extent to which changes are due to an intervention, controlling for other factors. It is often used in tandem with the concept of contribution. The latter emphasizes the confluence of multiple factors influencing a particular change (with one of these factors being the intervention). Experimental and quasi-experimental methods are often favored to address the attribution issue. These methods rely on the principle of a counterfactual

(i.e., what would have happened if the intervention had not taken place). For example, through the principle of randomization one can approximate this counterfactual by generating equivalent groups (equivalent on all observable and unobservable variables) but for the intervention (which is specifically assigned to a group). Any changes over time, given this equivalence, can be attributed to the intervention. While this is a powerful principle, it has a number of shortcomings as well. Apart from challenges in applicability, counterfactual designs are not (specifically) equipped to address the emergent nature of complex programs (i.e., both program design and intended outcomes often change in response to contextual, organizational, and political factors); to analyze multiple, nonlinear, and recursive causal pathways; and to identify unintended outcomes. In the case of complex interventions, quantitative counterfactual methods are applied after unpacking an intervention into smaller parts (see discussion above). Alternative methods are more appropriate to address one or more of the challenges described above. Examples are causal chain analysis (using, for example, process tracing) or quantitative comparative analysis (see Chapter 4).

4.6 Making Use of Innovative Data Collection Techniques and Sources of Data (see Chapters 10, 11, and 12)

Evaluation design and the validity of its findings can be strengthened by making optimal use of innovative techniques of data collection and sources of data, such as data on mobile phone use, geographic information systems, and biometric data. The increasing availability of very large data sets ("big data") is creating more opportunities for using new mathematical and statistical approaches to data mining and data analytics in complex program evaluation (UN Global Pulse, 2012).

5. Practical Applications

- Established evaluation approaches have several limitations with regard to addressing complexity issues. In practice, this often results in overlooking and/or oversimplifying essential complexity issues that one may wish to address in evaluations because of their bearing on intervention processes and outcomes. To deal with this issue, this chapter has discussed two options:
  - Using methods from complexity science to address particular complexity dimensions (see Chapter 1). The comparative advantages and limitations of these methods are discussed in this chapter.
  - Applying a number of principles to strengthen established evaluation approaches: clarifying elements of the system; relying on theory-based principles (see Chapters 5 and 6), unpacking complex interventions into evaluable parts (see Chapter 7), using a mixed methods approach (see Chapter 8), dealing with attribution (see Chapter 4), and using innovative techniques and sources of data (see Chapters 10, 11, and 12).

References

Bamberger, M. (2015). *Measuring resilience: Lessons from the design and implementation of monitoring and evaluation methodologies of the Horn of Africa Risk Transfer for Adaptation (HARITA)/R4 Rural Resilience Initiative: A case study.* Boston: Oxfam USA.

Bamberger, M., Rao, V., & Woolcock, M. (2010). Using mixed methods in monitoring and evaluation: Experiences from international development. In A. Tashakkori & C. Teddie (Eds.), *Sage handbook of mixed methods in social and behavioral research* (2nd ed., pp. 613–641). Thousand Oaks, CA: Sage.

Banerjee, A., Cole, S., Duflo, E., & Linden, L. (2007). Remedying education: Evidence from two randomized experiments in India. *Quarterly Journal of Economics, 122,* 1235–1264.

Barron, P., Diprose, P., & Woolcock, M. (2011). *Contesting development: Participatory projects and local conflict dynamics in Indonesia*. New Haven, CT: Yale University Press.

Befani, B., Ramalingam, B., & Stern, E. (2015). Introduction: Towards systemic approaches to evaluation and impact. *IDS Bulletin, 46*(1), 1–6.

Brown, D. G., Page, S. E., Riolo, R., & Rand, W. (2004). Agent-based and analytical modeling to evaluate the effectiveness of greenbelts. *Environmental Modelling & Software, 19,* 1097–1109.

Byrne, D. (2013). Evaluating complex social interventions in a complex world. *Evaluation, 19,* 217–228.

Byrne, D., & Callaghan, G. (2014). *Complexity theory and the social sciences: The state of the art.* Milton Park, UK: Routledge.

Byrne, D. & Ragin, C. (2009). *The Sage handbook of case-based methods.* Thousand Oaks, CA: Sage.

Chambers, R. (2007). *Who counts? The quiet revolution of participation and numbers* (Working Paper 296). Brighton, UK: Institute of Development Studies. Retrieved from https://www.ids.ac.uk/files/Wp296.pdf

Drew, R., Aggleton, P., Chalmers, H., & Wood, K. (2011). Using social network analysis to evaluate a complex policy network. *Evaluation, 17,* 383–394.

Estrella, M., & Gaventa, J. (1998). *Who counts reality? Participatory monitoring and evaluation: A literature review* (IDS Working Paper No. 70). Brighton, UK: IDS.

Forss, K., & Schwartz, R. (2011). Introduction. In K. Forss, M. Marra, & R. Schwartz (Eds.), *Evaluating the complex: Attribution, contribution and beyond* (pp. 1–32). New Brunswick, NJ: Transaction.

Funnell, S. C., & Rogers, P. J. (2011). *Purposeful program theory*. San Francisco, CA: Jossey-Bass.

GEF Independent Evaluation Office. (2013). *Fifth overall performance study: Impact of the GEF OPS5* (Technical Document #2). Retrieved from http://www.thegef.org/gef/sites/thegef.org/files/EO/TD2_Impact%20of%20the%20GEF.pdf

Hermans, L., Cunningham, S., & Slinger, J. (2014). The usefulness of game theory as a method for policy evaluation. *Evaluation, 20,* 10–15.

Hummelbrunner, R. (2011). Systems thinking and evaluation. *Evaluation, 17,* 395–403.

Innovation Network. (n.d.). *System mapping: A case example.* Retrieved from http://www.innonet.org/client_docs/File/advocacy/CARE_systems_map.pdf

Karlan, D., & Zinman, J. (2007). *Expanding credit access: Using randomized supply decisions to estimate the impacts* (Working Paper 108). Washington, DC: Center for Global Development.

King, J., Cousins, B., & Whitmore, E. (2007). Making sense of participatory evaluation: Framing participatory evaluation. *New Directions for Evaluation, 114,* 83–105.

Leroy, J. L., Ruel, M., & Verhofstadt, E. (2011). The impact of conditional cash transfer programmes on child nutrition: A review of evidence using a programme theory framework. *Journal of Development Effectiveness, 1*(2), 103–129.

McKenzie, D., & Woodruff, C. (2008). Experimental evidence on returns to capital and access to finance in Mexico. *World Bank Economic Review, 22,* 457–482.

Miguel, E., & Kramer, M. (2004). Worms: Identifying impacts on education and health in the presence of treatment externalities. *Econometrica, 72*(1), 159–217.

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

Olken, B. (2007). Monitoring corruption: Evidence from a field experiment in Indonesia. *Journal of Political Economy, 115,* 200–249.

Patton, M. Q. (2011). *Developmental evaluation: Applying complexity concepts to enhance innovation and use.* New York, NY: Guilford Press.

Perrin, B., & Wichmand, P. (2011). Evaluating complex strategic development interventions: The challenge of child labor. In K. Forss, M. Marra, & R. Schwartz (Eds.), *Evaluating the complex: Attribution, contribution and beyond* (pp. 243–282). New Brunswick, NJ: Transaction.

Pritchett, L., & Sandefur, J. (2013). *Context matters for size: Why external validity claims and development practice don't mix* (Working Paper 336). Washington, DC: Center for Global Development.

Rao, V., & Ibanez, A. M. (2003). *The social impact of social funds in Jamaica: A mixed methods analysis of participation, targeting and collective action in community-driven development* (Policy Research Working Paper 2970). Washington, DC: World Bank.

Rogers, P. J. (2011). Implications of Complicated and complex characteristics for key tasks in evaluation. In K. Forss, M. Marra, & R. Schwartz (Eds.), *Evaluating the complex: Attribution, contribution and beyond* (pp. 33–52). New Brunswick, NJ: Transaction.

Sastre-Merino, S., Vidueira, P., Diasz-Puente, J., & Fernandez-Mora, J. (2014). Capacity building through empowerment evaluation: An Aymara women artisans organization in Puno, Peru. In D. Fetterman, S. Kaftarian, & A. Wandersman (Eds.), *Empowerment evaluation* (2nd ed., pp. 76–85). Thousand Oaks, CA: Sage.

Schwartz, R., & Garcia, J. (2011). Intervention Path Contribution Analysis (IPCA) for complex strategy evaluation: Evaluating the smoke-free Ontario strategy. In K. Forss, M. Marra, & R. Schwartz (Eds.). *Evaluating the complex: Attribution, contribution and beyond* (pp. 187–208). New Brunswick, NJ: Transaction.

Spinatsch, M. (2011). Evaluating a complex policy in a complex context: The elusive success of the Swiss Smoking Prevention Policy. In K. Forss, M. Marra, & R. Schwartz (Eds.), *Evaluating the complex: Attribution, contribution and beyond* (pp. 170–206). New Brunswick, NJ: Transaction.

Stewart, R., van Rooyen, C., Korth, M., Chereni, A., Rebelo Da Silva, N., & de Wet, T. (2012). *Do micro-credit, micro-savings and micro-leasing serve as effective financial inclusion interventions enabling poor people, and especially women, to engage in meaningful economic opportunities in low- and middle-income countries? A systematic review of the evidence.* London, UK: EPPI-Centre.

Toulemonde, J., Carpenter, D., & Raffier, L. (2011). Coping with the evaluability barrier: Poverty impact of European support at country level. In K. Forss, M. Marra, & R. Schwartz (Eds.), *Evaluating the complex: Attribution, contribution and beyond.* New Brunswick, NJ: Transaction.

UN Global Pulse. (2012). Big data for development: Challenges and opportunities. Retrieved from http://unglobalpulse.org/projects/BigDataforDevelopment