Importance and Utilization of Theory-Based Evaluations in the Context of Sustainable Development and Social-Ecological Systems

Takaaki Miyaguchi

Abstract

Numerous challenges confront the task of evaluating sustainable development—its complex nature, complementary evaluation criteria, and the difficulty of evaluation at the nexus of human and natural systems. Theory-based evaluation, drawn from critical realism, is well suited to this task. When constructing a program theory/theory of change for evaluating sustainable development, concepts of socioecological systems and coupled human and natural systems are useful. The chapter discusses four modes of inference and the application of different theory-based evaluation approaches. It introduces the CHANS (coupled human and natural systems) framework, a holistic, analytical framework that is useful in evaluating such complex, social-ecological systems and resonates with the challenging elements of sustainable development evaluation.

Since the adoption of the 2030 Agenda for Sustainable Development and its Sustainable Development Goals (SDGs) in 2015, the United Nations member states and various stakeholders all over the world have been galvanizing their efforts to contribute to the achievement of the SDGs. Although the SDGs themselves were the result of international negotiation and consensus among the member states, the breadth of the partnerships and collaborations among non-state actors, including the private sector, nongovernment organizations, and nonprofit organizations, has been unprecedented.

The SDGs, of course, are not without critics. Some argue that these goals are nothing but a wish list (Hickel, 2015), while others point out the inconsistencies and incompatibility among 169 targets and their indicators and question the abilities and capacities of many states, especially those of developing countries, to adequately monitor and evaluate the current and future status toward achieving these SDGs (Leal Filho et al., 2019; Pongiglione, 2015; Stokstad, 2015).

The focus of this chapter is to look at the challenges in evaluating the status of sustainable development, which requires looking into the nexus of human and natural systems, and introduce the utility of theory-based evaluation for such purposes. The chapter introduces a holistic framework called CHANS (coupled human and natural systems), an analytical framework that is useful in evaluating such complex, social-ecological systems.

T. Miyaguchi (✉)

Kyoto University of Foreign Studies, Kyoto, Japan
e-mail: t_miyagu@kufs.ac.jp

Challenges in Evaluating Sustainable Development

We all know that humankind should strive for sustainable development as the concept is declared and promised with the SDGs. However, evaluating the status of and progress toward "sustainable development" is extremely difficult.

Sustainable development is a concept that is not just complicated—with interventions involving multiple components, multiple agencies, and multiple simultaneous and/or alternative causal strands—but also complex, having recursive causality with reinforcing loops, disproportionate relationships with a tipping point, and emergent outcomes (Rogers, 2008). Such characteristics of sustainable development make evaluation practice all the more challenging. Rowe (2012, 2014) identified four types of challenges.

First is the challenge of *attribution*. Because the status of sustainable development is found at the nexus of human and natural systems, achieving sustainability means maintaining the integrity of the combined ecological–societal system (Kay & Boyle, 2008). One can therefore anticipate the difficulties in comparing and matching both human and natural systems against those interventions that take place from the human system (Rowe, 2012; Vaessen & Todd, 2008). Pinning down, let alone quantifying, the level of attribution (or causation) is almost impossible.

The second difficulty is one of *temporal scale*. Although temporal scales for measuring economic activities or wealth being generated can be as brief as quarterly, when we turn our attention to society, a decade or more is required for us to confirm change within any generation in that period. What presents the toughest challenge in evaluating sustainable development is related to ecological time scales. For example, to validate a change of climate through an increase or decrease of greenhouse gas emissions requires 100 years. Even 20 to 30 years is needed to witness any change in climate variability. These scales of ecological systems are beyond our socioeconomic scales.

Temporal scale also has an important subdimension: *spatial frames*. An ecological spatial frame, such as a tropical rainforest, does not respect political or societal boundaries or jurisdictions. Adequate evaluation faces a great challenge due to such ecological spatial characteristics. And our modern history offers ample evidence that such ecological timeframes or spatial frames have been blatantly ignored for short-term benefits to the economy and society.

The third challenging aspect relates to *values*—economic, societal, and environmental. What type of value we adopt is a pivotal question when evaluating progress toward achieving sustainable development. To evaluate such progress, we must identify a common type of value through which we can compare the effectiveness of the efforts toward it. One valuation type that has been overly used in our modern history has been economic, or monetary values. But one can fathom the limitations of relying solely on this dimension of value and trying to apply it to other dimensions, such as ethnic, religious, cultural, and biodiversity. The various methods developed mainly by economists allow us to put an (economic) value on natural resources (such as contingency valuation, hedonic pricing, or cost effectiveness analysis), but these are derived from and based only on the socioeconomic dimension and do not allow us to grasp the complex nature of social-ecological systems.

The fourth type of challenge is one of achieving *use* and *influence*. Numerous knowledge products and evaluation reports address sustainable development, but whether these products have been put to actual use is quite a different matter. Therefore, engaging decision makers and stakeholders in the evaluation process itself is vital so that they will put the results to use toward their decision-making processes.

In addition to these four types of challenges in evaluating sustainable development, we also see an aggregation challenge known as a micromacro paradox (Uitto, 2014; Vaessen & Todd, 2008; Van den Berg & Cando-Noordhuizen, 2017). This refers to lack of coherence or effectiveness when many successes at a micro level do not accumulate accordingly to result in successes at a larger, macro scale. Such paradox stems from reductionism. The shortcomings of reductionism

are made especially apparent when we deal with complex systems for which the whole is more than the sum of the parts (Bhaskar et al., 2010; Kay, 2008a).

Sustainable development maintains the integrity between socioeconomic and ecological systems. But more often than not, measuring, analyzing, and evaluating the status of or movement toward sustainable development has been influenced by social science disciplines rather than natural, biophysical sciences (Rowe, 2012). Such analysis leaves no doubt that all economic and social activities are based on a healthy environment and the finite resources existing on earth. Economic activity is, in effect, the conversion of material and energy from a natural resource pool as input with converted material and used energy as output. As ecological economist Herman Daly (1990) put it, there is no such thing as "sustainable growth" when every single economic activity is based on the natural resources existing on a finite planet. Although natural systems are thus the absolute foundation of all economic activities, the international discourse pertaining to sustainable development until now has been dominated by socioeconomic aspects—the human system side (Rowe, 2012, 2014).

However, the problem is not just over-reliance on social sciences; what matters is the polarization in which attempts to evaluate sustainable development happen only with either social science discipline or with natural science discipline—without their integration or synthesis. The natural ecosystems are diverse, complex, and dynamic; thus, traditional, disciplinary science is "not by itself sufficient for understanding and dealing with ecosystems" (Waltner-Toews et al., 2008, xii). In light of these current situations surrounding sustainable development evaluation efforts, we turn to theory-based evaluation and its approaches.

Theory-Based Evaluation

Before discussing theory-based evaluation and its approaches, we must clarify the term's meaning vis-à-vis other terms used in evaluation literature.

Theory-based evaluation (used by Weiss, 1997a) is, in short, a "plausible and sensible model of how the program is supposed to work" (Bickman, 1987). Other terms are interchangeable, such as logic model (Mathison, 2004), program theory (Bickman, 1990), the theory of action (Patton, 1997), theory of change (Weiss, 1997a), and theory-driven evaluation (Chen & Rossi, 1983). In this chapter, I use Weiss's terms *theory-based evaluation* and *theory of change*, which consists of *implementation theory* and *program theory*.[1]

According to Brousselle and Buregeya (2018), theory-based evaluation has emerged in reaction to current normal evaluation practice. They assert the need for a theory of change, not just for poorly formulated interventions, but especially when evaluating complex interventions. And theory-based evaluation and its approaches are "aimed at reinforcing the explanatory power of evaluations" (Weiss, 1997b).

Theory-based evaluation formulates program elements, rationale, and causal linkages. The atheoretical approach to evaluation has been characterized by "a step-by-step cookbook method of doing evaluations" (Chen, 1990). The atheoretical approach tends to focus on the relationship between inputs and effects without considering the transformational processes, referred to as "black box evaluations" (Norgbey & Spilsbury, 2014). Going beyond such atheoretical approach, theory-based evaluation takes into account the transformational processes inherent in the programs being evaluated (Chen, 1990).

Theory-based evaluation pays close attention to contextual conditions. According to Chen (1990), theory of change consists of two parts, normative theory and causative theory.[2] The causative theory "specifies how the program

---

[1]Funnell and Rogers (2011) reverses these terms so that p*rogram theory* consists of *theory of action* and *theory of change*. Thus, somewhat confusingly, *program theory* by Weiss corresponds to *theory of change* by Funnell and Rogers. Since the terms by Weiss are used more often in international development and its evaluation field, I have adopted her terms in this chapter.

[2]According to the original terms adopted by Chen (1990), it is described as *program theory* (instead of *theory of change*, adopted by Weiss). However as explained in the previous footnote, Weiss's terminology, *theory of change*, is used in this chapter.

works by identifying the conditions under which certain processes will arise and what their likely consequences will be" (Chen, 1990).

With its focus on contextual conditions, theory-driven evaluation also shares three fundamental characteristics: (a) to explicate the theory of treatment by detailing the expected relationships among inputs, mediating processes, and short- and long-term outcomes; (b) to measure all of the constructs postulated in the theory; and (c) to analyze the data to assess the extent to which the postulated relationships actually occurred (Coryn et al., 2011; Shadish et al., 2002).

Several approaches stem from theory-based evaluation, including theory of change, realist evaluation, logic analysis, and contribution analysis. All of these approaches have philosophical and conceptual roots in a philosophy of science known as *critical realism* (Brousselle & Buregeya, 2018). And an origin in critical realism is deemed quite appropriate to evaluating sustainable development, which involves two-evaluand systems.

Critical Realism

Critical realism is a philosophy of science advocated by Roy Bhaskar. It originated as a critique of a deterministic worldview, which took the stance that if some factor X occurred—such as an intervention—then the observed result Y must follow (Forss et al., 2011). This philosophy can be understood through four modes of inference, distinction between open and closed systems, and explanatory power rather than prediction.

First, the four modes of inference are necessary to understanding critical realism. The first two, *deduction* and *induction*, are well known. Through deduction and induction inference, evaluators get to know what works (through deduction by applying a theory, and through induction with observations). The latter two modes of inference, *abduction* and *retroduction*, are less familiar. Abduction combines the deductive and inductive modes of inference and is defined as "working from consequence back to cause or antecedent" (Denzin, 2017, p. 100). In other

words, abduction means "to interpret and recontextualize individual phenomena within a conceptual framework to understand something in a new way" (Danermark et al., 2002, p. 80). In evaluation, this abduction inference is synonymous with constructing a program theory. According to Weiss (1997a), program theory refers to "the mechanisms that mediate between the delivery (and receipt) of the program and the emergence of the outcomes of interest" (p. 57). In other words, program theory is hypothesized causal linkages. In evaluation terms, then, it connotes for whom an intervention may work and, above all, how it works.

The fourth mode of inference, retroduction, provides the essence of this philosophy of science. Retroduction means to "reconstruct the basic conditions for these conceptually abstracted phenomena to be what they are" (Danermark et al., 2002, p. 80). It is one thing to talk about hypothesized (abstracted) causal linkages, but it is quite another to pay heed to the conditions under which such generative mechanisms can be triggered. Pawson and Tilley (1997), referring to this notion of critical realism, likened such conditions to a gunpower explosion that does not always take place when flame is applied, but also requires certain conditions, such the gunpower mixture being compacted, the structure not being damp and having sufficient quantity and oxygen, and heat applied long enough. Gunpower explosion functions as a generative mechanism and is synonymous with Weiss's program theory (Blamey & Mackenzie, 2007). In evaluation terms, through this fourth mode of inference, retroduction, evaluators can grasp what may work under what circumstances.

Therefore, through utilizing all four modes of inference described above, evaluators will be able to know what works, for whom, how, and under what circumstances. Theory-based evaluation and its approaches resonate quite well with this statement that is the essence of critical realism, and thus the root of theory-based evaluation.

The second component for understanding critical realism, as described by Bhaskar (2013), is the concept of the world as having three domains:

empirical (observable experiences), actual (a factual event that is generated by mechanisms), and real (the mechanisms that generate an event). These three domains establish a critical perspective in which the reality that scientists study is larger than only the empirical domain (Bhaskar, 2013).

Further understanding this concept requires a grasp of the difference between closed and open systems. A closed system is akin to an experiment in which a certain mechanism is tested in an isolated laboratory setting, allowing the mechanism to operate in isolation, independent of other mechanisms. An open system is akin to society itself, in which social events are the products of many simultaneously existing mechanisms, exemplifying the complex nature of society. Because society is inherently an open system, we must recognize that one cannot isolate a single social mechanism and do an experiment. The above-mentioned modes of inference in social science function as an experiment does in natural science (Danermark et al., 2002).

The third important element in understanding critical realism is the difference between *explanations* and *predictions*. In a closed system, explanations are synonymous with predictions, whereas explanations in an open system indicate tendencies. When attempting to seek external validity in an open system, one should seek explanations, rather than predictions or judgments (Allen, 2008), to reveal the causal mechanism hidden beneath the surface (Brousselle & Buregeya, 2018).

Importance of Theory-Based Evaluation Approaches

The school of theory-based evaluation includes approaches with different implications (Alkin, 2013). When choosing among them to evaluate sustainable development, knowing the strengths and weaknesses of the two theory-based evaluation approaches—realist approach and theory of change—is important. Evaluators need to be aware of these similar but distinct approaches and adopt the one that is appropriate to the purpose of the evaluation.

Realist approach is concerned with promising context-mechanism-outcome configurations (called CMO configurations; Pawson & Tilley, 1997). Utilizing this approach, evaluators can hypothesize various program theories to determine which are effective (or not) under certain circumstances. In other words, realist approach helps to deliver more precise and substantive program learning. At the same time, however, it is less appropriate for dealing with highly complex, multisite interventions with multiple outcomes (Blamey & Mackenzie, 2007). Theory of change, in contrast, is more concerned with overall program outcomes and helps to provide a strategic perspective on a complex program (Blamey & Mackenzie, 2007).

Theory-based evaluation approaches are appropriate for evaluating the status of and progress toward sustainable development, which is both complicated and complex. Based on the characteristics of theory-based evaluation approaches, prudent evaluators adopt appropriate approaches for different purposes. Evaluators should use the theory of change approach, for example, when evaluating the overall status of sustainable development, and choose the realist approach to hypothesize and understand certain program theories that are deemed effective for successful results within each program component. Constructing and analyzing a theory of change is an essential method for resolving the problems inherent in complex interventions (Dubois et al., 2011; Morell, 2010).

But how can we construct theories of change to apply to sustainable development evaluation? How do we assess emergent and anticipated outcomes resulting from relationships that are sometimes non-linear (Morell, 2010; Shiell et al., 2008), and how do we deal with uncertainty created by complex, self-organizing systems (Kay, 2008a)?

According to Funnell and Rogers (2011), theories of change can be constructed in three ways.[3] *Stakeholder mental model* is articulated according to how stakeholders believe a program will achieve what it is designed to do. Through *deductive approach*, a theory of change uses formal and informal documentation and research theories about a program and the needs it is intended to address. And last, *inductive approach* "involves observing the program in action and deriving the theories that are implicit in people's actions when implementing the program" (Funnell & Rogers, 2011, p. 111).

Out of these three techniques, however, there is an over-reliance on the deductive approach for theory development, with as many as 91% of analyzed cases reported to have used this approach, compared to 49% for the stakeholder mental model and 13% for the inductive development approach (Coryn et al., 2011). Predominantly, these theories are derived from social sciences. Scriven (2012) pointed out a strong tendency of professional evaluators to specialize in just one of the many branches of evaluation and only one area of human activity, further narrowing the scope of evaluation and thereby increasing difficulties in evaluating sustainable development.

This discussion of approaches has two important points. First, we find fewer cases of constructing theories of change from a natural science-based standpoint. And second, hardly any theory of change construction integrates both social science and natural science; rather, evaluators have tended one way, using either social science-based or natural science-based approaches (Rowe, 2012).

If we are to evaluate sustainable development at the nexus between human and natural systems, evaluators should integrate both social and natural sciences in constructing and hypothesizing theories of change, especially when the status of

sustainable development is about maintaining the integrity among society, economy, and environment.

Coupled Human and Natural Systems (CHANS)

Just as social sector problems and their evaluations have been dominated by the social sciences and their theories, the aspect of sustainability—especially within the context of ecological sustainability—has been equally dominated by natural, biophysical scientists. However, dealing with both social and ecological systems requires analyses that involve several components from each system, such as research on energy-water nexus and food-energy-water nexus. Despite this, studies on nexuses with three and four nodes are still very rare (Liu, Hull, Yang, Viña, Chen, et al., 2016).

One promising theoretical framework for understanding the mutual interactions and feedback mechanisms between human and natural systems has been advocated and advanced by Nobel laureate Elinor Ostrom in her pioneering work on social-ecological systems. Her research was concerned mainly with natural resources, especially common pooled resources, and provided a strong foundation to further understand the governance for successfully managing the *commons*, once considered impossible for an economic, rational, decision-maker worldview (Folke, 2007; Liu et al., 2007; McGinnis & Ostrom, 2014; Ostrom, 1990).

The essence of this so-called adaptive management and governance is about two-way interactions and feedback loops found between social-ecological systems (Evans, 2012). What Ostrom's work demonstrated was that socioeconomic entities such as fishing villages could change their way of governing themselves, adapting their decision-making rules and procedures in reaction to a situation such as a change in the ecological status of their surroundings. The related research has resulted in a general framework for analyzing sustainability of social-ecological sys-

[3]The original text of Funnell and Rogers (2011) used the term *program theory* instead of *theory of change*, but I have used theory of change, an interchangeable term by Weiss, to be consistent with the selection of evaluation terms and concepts in this chapter.

tems, fully taking into account both human and natural systems (Ostrom, 2009).

Stemming from Ostrom's work on adaptive management is another insightful analytical framework for understanding social-ecological systems, called the coupled human and natural systems (CHANS) framework. The primary focus of Ostrom's research was on common-pool resources in which the ecological system was either unowned or ownership was shared. However, the CHANS analytical framework goes well beyond the scale of common-pool resources and can thus provide helpful new insights that apply to the evaluation of sustainable development.

According to Liu, Hull, Carter, et al. (2016), the major barrier to effective implementation of sustainable development is the lack of sufficient knowledge about the complex relationships between humans and nature. The CHANS approach is intended "to serve as a pragmatic, heuristic tool for analyzing into relationships between people and the environment." The CHANS framework emphasizes that the human and natural components are coupled rather than separate (Carter et al., 2014, para. 6).

Among many other scholars, Ostrom has emphasized that context (i.e., not interventions themselves but the systems and subsystems that surround them, such as societal, political, and economic situations) does matter in analyzing the intricate interactions between human and natural systems. What is distinctive about CHANS is that it does not treat such contextual factors as external but as intrinsic elements within the framework. Researchers used a CHANS framework to conduct a 20-year-long study of social-ecological interactions that surround the biodiversity hot spot of the Wolong National Park of China, home to an endangered species of panda (*Ailuropoda melanoleuc*). These researchers proposed a framework that incorporates the human subsystem components such as communities and local residents, and the natural subsystem components such as wildlife and the land cover characterizing their habitat (Carter et al., 2014). The variety in the study's analyses was truly transdisciplinary. They included dedicated research on the influence and relationships within this coupled system

surrounding *Ailuropoda melanoleuc*, such as demography at household level and by distance and elevation level, education, energy transition, government policies, human dependence on ecosystem, infrastructure, livestock and livestock-panda interactions, payment for ecosystem services, scenario analysis and modeling, and spatial and tree distribution (Liu, Hull, Yang, Viña, Chen, et al., 2016).

Resonating well with the characteristics of sustainable development—complex systems involving both human and natural systems—and social and natural science disciplines, the CHANS framework "provides a platform for natural and social scientists to work together to quantify and integrate human-nature relationships at multiple organizational levels across space and over time" (Liu, Hull, Carter, et al., 2016).

Another characteristic of this framework is that it considers and treats the focal coupled system as an open system, rather than a closed system, placing the focal coupled system under specific social, economic, and political settings (Ostrom, 2009).

Why We Need a Framework Like CHANS

Especially when evaluating the complex systems of sustainable development, evaluators should consider adopting theory-based evaluation and its approaches instead of an oversimplified, one-size-fits-all, black box approach.

Among the seven traps[4] in constructing a theory of change proposed by Funnell and Rogers (2011), having "no actual theory" is on top of the list. In evaluating sustainable development, we especially need to avoid this trap by developing theories of change that are: (a) based on both social and natural sciences, (b) able to recognize

---

[4]They are: (1) no actual theory; (2) having a poor theory of change; (3) poorly specifying intended results; (4) ignoring unintended results; (5) oversimplifying; (6) not using the program theory for evaluation; and (7) taking a one-size-fits-all approach (Funnell & Rogers, 2011, p. 42).

the interactions between human and natural systems, and (c) capable of describing nonlinearity and emerging traits of complex systems and incorporating ecological temporal scale and spatial frames. Moreover, because theory-based evaluation is method neutral and suited to quantitative or qualitative methods, or both (Chen, 2005; Donaldson, 2007), the CHANS framework also offers flexibility for evaluators. CHANS can systematically guide researchers in analyzing complex sustainability issues surrounding socio-ecological systems.

Another valuable element of the CHANS framework is that it recognizes the importance of the participatory approach, or "putting researchers in the local residents' shoes" (Liu, Hull, Yang, Viña, An, et al., 2016). Many studies of social-ecological systems adopt "participatory approaches to identify, characterize, and solve management-related problems" (Norberg & Cumming, 2008, p. 238). The importance of such an approach goes beyond a specific set of rules of one method. Participatory approach is vital because complex systems cannot be captured by any single perspective and require a plurality of perspectives. Such plurality requires a variety of "forms of inquiry, inclusion of, and dialogue with persons representing different interests and different world views" (Waltner-Toews & Wall, 1997, p. 30). Because all coupled systems in question develop out of historical and cultural conditions, the future of such a system cannot have one single preferred state. As Kay (2008b) poignantly stated, researchers, if left to decide, will inquire into those aspects of the system that they themselves deem important; therefore, it is "crucial that the values, concerns, and knowledge of local stakeholders and actors be central to any inquiry" (p. 30).

Of course, this is not to claim that CHANS is the only framework through which we can evaluate sustainable development at the nexus of environment and development. However, evaluators should seek to use a framework that: (a) can encompass the complicated and complex nature of sustainable development; (b) is holistic, multilayered, and multiscaled; and (c) draws from both social and natural sciences, so that program theories develop using perspectives from both disciplines.

Appropriate Methodologies

CHANS appears to provide a useful framework for evaluating sustainable development. What can then be the appropriate methodologies and approaches for capturing such coupled systems? Evaluators have four types of methodologies to consider. First is *triangulation*, "the process of gathering scientific evidence about a system through a combination of laboratory, field, modeling, and historical investigations, facilitated by iterative and cross-disciplinary collaboration among research groups" (Plowright et al., 2008). When investigating the dynamism of complex systems, we cannot predict or reach a correct answer, because such is only possible based on a linear (irreversible, one-way) cause-and-effect worldview that excludes all influencing factors under a simple, laboratory-like system. To narrow the level of uncertainty and describe complex systems with more explanatory power, we need to shed light on the triangulation method. This method has been well practiced and its importance widely acknowledged among many evaluators (Carugi, 2016; Forss et al., 2011; Morra-Imas & Rist, 2009; Patton, 2002; Uitto, 2016).

The second type of methodology is *cross-scale/cross-layer comparison*. Complex social-ecological systems are nonlinear with reversible feedback loops, in requirement for multiple perspectives, and are multiscaled and multilayered (Waltner-Toews & Kay, 2008). Therefore, the ability to pursue several different lines of exploration at several different scales is necessary (Norberg & Cumming, 2008). For one example, analyzing or constructing simulation models only at a large, global scale (e.g., greenhouse gas emission modeling) would be inadequate; instead, the evaluator must compare different scales or layers within the systems. A local landscape is applied to a sub-watershed, which is made up of the ecological communities such as woodlots, wetlands, open fields, etc., each of which then is made up of individual species (Kay & Boyle, 2008).

To understand why certain social-ecological systems have not succeeded, we can conduct cross-scale/cross-layer comparisons and analyses

at different spatial and temporal scales (Cumming, 2007; Ostrom, 2009). Evaluation already has a method that encompasses such nested nature models, called nested theories of change (Mayne, 2015; Richards, 2019; Riley et al., 2018). Although almost all the applied cases of nested theories of change in evaluation literature are found within the human (social) systems, evaluators in natural (ecological) systems can also adopt this method.

The third methodology type is *causal inference*. Even though the field of evaluation has been dominated by social scientists and their theories, the use of causal inference within natural science domains has begun to attract attention, notably in the cases of emerging infectious disease (Plowright et al., 2008) and global biodiversity scenarios and landscape ecology (Cumming, 2007). Thus, we see the utility of theory-based evaluation approaches even in the realm of natural science. Incorporating both natural science and social science perspectives in constructing theories of change is a prerequisite for starting to evaluate sustainable development; therefore, and an analytical framework like CHANS that enables such integration is necessary.

The final methodology is *cross-site synthesis* and *meta-analysis.* Because social-ecological systems are both complicated and complex, trying to identify a one-size-fits-all strategy will be in vain. At the same time, treating every single social-ecological system as a completely different and local incidence will not likely generate any externally valid insights that are generalizable to other parts of the world. Rather, to do so, "different ecological, socioeconomic, political, demographic, and/or cultural settings need to be synthesized" (Carter et al., 2014). Liu, Hull, Carter, et al. (2016) stressed the importance of seeking external validity and generalizability despite highly localized situations in each social-ecological system. They also advocated the importance of "model (social-ecological) systems," i.e., those that contain the core and essence of CHANS. By conducting cross-site syntheses or meta-analyses, CHANS researchers have been already able to identify some common aspects of

social-ecological complex systems that are applicable and spread across the globe (Carter et al., 2016).

Several CHANS sites have shared these common characteristics:

- Organizational—restoring reciprocal effects and feedbacks with nested hierarchies, indirect effects, emergent properties, vulnerability, and thresholds and resilience
- Spatial coupling—coupling across spatial scales, couplings beyond boundaries, and heterogeneity
- Temporal couplings—human impacts on natural systems, rising natural impacts on humans, legacy effects, time lags, increased scales and pace, and escalating indirect effects[5]

Evaluators are encouraged to start paying close attention to this research field on social-ecological systems and coalesce the previously separated efforts and research results from social and natural science into one, holistic framework such as CHANS.

Conclusion

Since the adoption of the 2030 Agenda for Sustainable Development in 2015, the concept and its goals have spread globally, with an increasing level of awareness and with inspiring, collaborative, multistakeholder implementation initiatives all over the world. At the outset, with 17 SDGs, the objectives seemed clear. However, beyond the political rhetoric of these goals and targets, we realize that we cannot declare achievement of sustainable development when all 169 targets are met *separately*. The essence of sustainable development is to acquire and maintain integrity among the three pillars—social, economic, and environmental. These three pillars are closely interlinked and interwoven. Accumulating each block or project successes from the micro

---

[5]For more details, refer to Liu, Hull, Yang, Viña, Chen, et al. (2016).

level will not lead to the macro-level integrity that these goals are seeking overall. Evaluators face a formidable task in evaluating sustainable development, homing in on the nexus between human and natural systems.

We face four types of challenges in evaluating sustainable development: the issue of attribution, temporal scale, the values, and achieving use and influence. At the same time, we also face an extra challenge of the micro-macro paradox. Theory-based evaluation and its approaches offer a means well suited to evaluating these complex systems that are multilayered, multiscaled, and span different time scales.

Theory-based evaluation has its roots in critical realism, a philosophy of science that emerged out of criticism against a deterministic worldview. Fully utilizing four modes of inference, critical realism can help reconstruct the basic conditions for certain phenomena to be what they are, by paying special attention to the context in which the specific generative mechanism is triggered.

Even though theory-based evaluation and its approaches are considered appropriate in evaluating complex systems, the theories of change that we develop and use tend to come predominantly from the social science discipline and be deductively constructed, instead of articulated by stakeholders or inductively constructed. When we deal with a social-ecological system, which is both complicated and complex, we need to develop theories of change that are based on well-developed principles from both the natural and social sciences—particularly ecology, economics, and political science—and we must confront this formidable task through comparative analyses of many cases (Walker et al., 2006).

This chapter introduced the useful analytical framework called CHANS (coupled human and natural systems) that is capable of addressing the issues mentioned above. This framework has a strong influence from Ostrom and her work on adaptive management and governance of resources held in common. CHANS emphasizes that human and natural components are coupled, rather than separate, and incorporates political and socioeconomic situations as an integral part

of the framework, rather than merely the external drivers of change.

By closely examining and applying the CHANS framework to ongoing and future programs concerned with achieving sustainable development, evaluators can address the four types of challenges in evaluating sustainable development. Although CHANS is not the only framework that facilitates addressing these issues and challenges, it has particular promise in supporting evaluation of sustainable development.

Knowing about a framework is one thing, but conducting actual analyses is quite another. However, the methodologies discussed here, such as triangulation, cross-scale/cross-layer comparisons, causal inference utilizing both social and natural science, and use of meta-analysis, are considered appropriate in evaluating social-ecological systems.

Although one might argue that no conceptual model exists for evaluating sustainable development with a holistic lens, using a framework like CHANS allows evaluators to construct theories of change and conduct subsequent analyses. At the same time, it supports specific analysis both quantitatively and qualitatively and utilizes both social and natural sciences.

Evaluating outcomes that a program cannot hope to influence may be impossible. However, because the CHANS framework specifically focuses on the interlinkages and mutual influence at the nexus between environment and development, it enables analysis, if not outright attribution, of a level of contribution to long-term outcomes that are seemingly outside of a program's direct scope.

With the recent increase in the level of awareness and attention to the concept of sustainable development and its goals, we should soon see more evaluations of subjects that would traditionally be considered outside the (narrow) scope of a program. Theory-based evaluation and its approaches, with the support of an analytical framework like CHANS, should be a great resource for our continuous and collaborative efforts in evaluating sustainable development.

References

Alkin, M. C. (2013). *Evaluation roots: A wider perspective of theorists' views and influences* (2nd ed.). Sage.

Allen, T. (2008). Scale and type: A requirement for addressing complexity with dynamical quality. In D. Waltner-Toews, J. J. Kay, & N.-M. E. Lister (Eds.), *The ecosystem approach: Complexity, uncertainty, and managing for sustainability* (pp. 37–49). Columbia University.

Bhaskar, R. (2013). *A realist theory of science*. Routledge.

Bhaskar, R., Frank, C., Parker, J., & Høyer, K. G. (2010). *Interdisciplinarity and climate change: Transforming knowledge and practice for our global future*. Taylor & Francis.

Bickman, L. (1987). *Using program theory in evaluation*. Jossey-Bass.

Bickman, L. (1990). Editor's notes: Advances in program theory. *New Directions for Program Evaluation, 1990*(47), 1–4. https://doi.org/10.1002/ev.1550.

Blamey, A., & Mackenzie, M. (2007). Theories of change and realistic evaluation: Peas in a pod or apples and oranges? *Evaluation, 13*(4), 439–455. https://doi.org/10.1177/1356389007082129.

Brousselle, A., & Buregeya, J.-M. (2018). Theory-based evaluations: Framing the existence of a new theory in evaluation and the rise of the 5th generation. *Evaluation, 24*(2), 153–168. https://doi.org/10.1177/1356389018765487.

Carter, N. H., An, L., & Liu, J. (2016). Cross-site synthesis of complexity in coupled human and natural systems. In J. Liu, V. Hull, W. Yang, A. Viña, Z. Ouyang, & H. Zhang (Eds.), *Pandas and people: Coupling human and natural systems for sustainability* (pp. 203–217). Oxford University.

Carter, N. H., Viña, A., Hull, V., McConnell, W. J., Axinn, W., Ghimire, D., & Liu, J. (2014). Coupled human and natural systems approach to wildlife research and conservation. *Ecology and Society, 19*(3), 43. https://doi.org/10.5751/ES-06881-190343.

Carugi, C. (2016). Experiences with systematic triangulation at the global environment facility. *Evaluation and Program Planning, 55*, 55–66. https://doi.org/10.1016/j.evalprogplan.2015.12.001.

Chen, H.-T. (1990). *Theory-driven evaluations*. Sage.

Chen, H.-T. (2005). *Practical program evaluation: Assessing and improving planning, implementation, and effectiveness*. Sage.

Chen, H.-T., & Rossi, P. H. (1983). Evaluating with sense: The theory-driven approach. *Evaluation Review, 7*(3), 283–302.

Coryn, C. L., Noakes, L. A., Westine, C. D., & Schröter, D. C. (2011). A systematic review of theory-driven evaluation practice from 1990 to 2009. *American Journal of Evaluation, 32*(2), 199–226. https://doi.org/10.1177/1098214010389321.

Cumming, G. S. (2007). Global biodiversity scenarios and landscape ecology. *Landscape Ecology, 22*(5), 671–685.

Daly, H. E. (1990). Sustainable growth: A bad oxymoron. *Journal of Environmental Science & Health Part C, 8*(2), 401–407. https://doi.org/10.1080/10590509009373395.

Danermark, B., Ekström, M., Jakobsen, L., & Karlsson, J. C. (2002). *Explaining society: Critical realism in the social sciences*. Routledge.

Denzin, N. K. (2017). *The research act: A theoretical introduction to sociological methods*. Transaction.

Donaldson, S. I. (2007). *Program theory-driven evaluation science: Strategies and applications*. Routledge.

Dubois, N., Lloyd, S., Houle, J., Mercier, C., Brousselle, A., & Rey, L. (2011). Discussion: Practice-based evaluation as a response to adress intervention complexity. *The Canadian Journal of Program Evaluation/La Revue canadienne d'evaluation de programme, 26*(3), 105–113.

Evans, J. P. (2012). *Environmental governance*. Routledge.

Folke, C. (2007). Social–ecological systems and adaptive governance of the commons. *Ecological Research, 22*(1), 14–15. https://doi.org/10.1007/s11284-006-0074-0.

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

Funnell, S. C., & Rogers, P. J. (2011). *Purposeful program theory: Effective use of theories of change and logic models*. Wiley.

Hickel, J. (2015). The problem with saving the world. *Jacobin Magazine*, 8. https://www.jacobinmag.com/2015/08/global-poverty-climate-change-sdgs/

Kay, J. J. (2008a). An introduction to systems thinking. In D. Waltner-Toews, J. J. Kay, & N.-M. E. Lister (Eds.), *The ecosystem approach: Complexity, uncertainty, and managing for sustainability* (pp. 3–13). Columbia University.

Kay, J. J. (2008b). Framing the situation: Developing a system description. In D. Waltner-Toews, J. J. Kay, & N.-M. E. Lister (Eds.), *The ecosystem approach: Complexity, uncertainty, and managing for sustainability* (pp. 15–34). Columbia University.

Kay, J. J., & Boyle, M. (2008). *Self-organizing, holarchic, open systems (SOHOs)*. Columbia University.

Leal Filho, W., Tripathi, S. K., Andrade Guerra, J., Giné-Garriga, R., Orlovic Lovren, V., & Willats, J. (2019). Using the sustainable development goals towards a better understanding of sustainability challenges. *International Journal of Sustainable Development & World Ecology, 26*(2), 179–190. https://doi.org/10.1080/13504509.2018.1505674.

Liu, J., Dietz, T., Carpenter, S. R., Alberti, M., Folke, C., Moran, E., Pell, A. N., Deadman, P., Kratz, T., Lubchenco, J., Ostrom, E., Ouyang, Z., Provencher, W., Redman, C. L., Schneider, S. H., & Taylor, W. W. (2007). Complexity of coupled human and natural systems. *Science, 317*(5844), 1513–1516. https://doi.org/10.1126/science.1144004.

Liu, J., Hull, V., Carter, N., Viña, A., & Yang, W. (2016). Framing sustainability of coupled human and natural systems. In J. Liu, V. Hull, W. Yang, A. Viña, X. Chen, Z. Ouyang, & H. Zhang (Eds.), *Pandas and people:*

Coupling human and natural systems for sustainability (pp. 15–32). Oxford University.

Liu, J., Hull, V., Yang, W., Viña, A., An, L., Carter, N., Chen, X., Liu, W., Ouyang, Z., & Zhang, H. (2016). Lessons from local studies for global sustainability. In J. Liu, V. Hull, W. Yang, A. Viña, X. Chen, Z. Ouyang, & H. Zhang (Eds.), *Pandas and people: Coupling human and natural systems for sustainability* (pp. 240–252). Oxford University.

Liu, J., Hull, V., Yang, W., Viña, A., Chen, X., Ouyang, Z., & Zhang, H. (Eds.). (2016). *Pandas and people: Coupling human and natural systems for sustainability*. Oxford University Press.

Mathison, S. (2004). *Encyclopedia of evaluation*. Sage.

Mayne, J. (2015). Useful theory of change models. *Canadian Journal of Program Evaluation, 30*(2), 119–142. https://doi.org/10.3138/cjpe.230.

McGinnis, M. D., & Ostrom, E. (2014). Social-ecological system framework: Initial changes and continuing challenges. *Ecology and Society, 19*(2), 30. https://doi.org/10.5751/ES-06387-190230.

Morell, J. A. (2010). *Evaluation in the face of uncertainty: Anticipating surprise and responding to the inevitable*. Guilford.

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

Norberg, J., & Cumming, G. (2008). Complexity theory for a sustainable future: Conclusions and outlook. In J. Norberg & G. Cumming (Eds.), *Complexity theory for a sustainable future* (pp. 277–293). Columbia University.

Norgbey, S., & Spilsbury, M. (2014). A programme theory approach to evaluating normative environmental interventions. In J. I. Uitto (Ed.), *Evaluating environment in international development* (pp. 123–149). Routledge.

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge University.

Ostrom, E. (2009). A general framework for analyzing sustainability of social-ecological systems. *Science, 325*(5939), 419–422. https://doi.org/10.1126/science.1172133.

Patton, M. Q. (1997). *Utilization-focused evaluation*. Sage.

Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Sage.

Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. Sage.

Plowright, R. K., Sokolow, S. H., Gorman, M. E., Daszak, P., & Foley, J. E. (2008). Causal inference in disease ecology: Investigating ecological drivers of disease emergence. *Frontiers in Ecology and the Environment, 6*(8), 420–429. https://doi.org/10.1890/070086.

Pongiglione, F. (2015). The need for a priority structure for the Sustainable Development Goals. *Journal of Global Ethics, 11*(1), 37–42. https://doi.org/10.1080/17449626.2014.1001912.

Richards, R. (2019). *The value of theory of change in large-scale projects and complex interventions* (K4D Helpdesk Report). Institute of Development Studies.

Riley, B. L., Kernoghan, A., Stockton, L., Montague, S., Yessis, J., & Willis, C. D. (2018). Using contribution analysis to evaluate the impacts of research on policy: Getting to 'good enough'. *Research Evaluation, 27*(1), 16–27. https://doi.org/10.1093/reseval/rvx037.

Rogers, P. J. (2008). Using programme theory to evaluate complicated and complex aspects of interventions. *Evaluation, 14*(1), 29–48. https://doi.org/10.1177/1356389007084674.

Rowe, A. (2012). Evaluation of natural resource interventions. *American Journal of Evaluation, 33*(3), 384–394. https://doi.org/10.1177/1098214012440026.

Rowe, A. (2014). Evaluation at the nexus: Principles for evaluating sustainable development interventions. In J. I. Uitto (Ed.), *Evaluating environment in international development* (pp. 69–85). Routledge.

Scriven, M. (2012). Conceptual revolutions in evaluation: Past, present, and future. In M. C. Alkin (Ed.), *Evaluation roots* (pp. 167–179). Sage.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

Shiell, A., Hawe, P., & Gold, L. (2008). Complex interventions or complex systems? Implications for health economic evaluation. *BMJ, 336*(7656), 1281–1283. https://doi.org/10.1136/bmj.39569.510521.AD.

Stokstad, E. (2015). Sustainable goals from UN under fire. *Science, 347*(6223), 702–703. https://doi.org/10.1126/science.347.6223.702.

Uitto, J. I. (2014). Evaluating environment and development: Lessons from international cooperation. *Evaluation, 20*(1), 44–57. https://doi.org/10.1177/1356389013517443.

Uitto, J. I. (2016). Evaluating the environment as a global public good. *Evaluation, 22*(1), 108–115. https://doi.org/10.1177/1356389015623135.

Vaessen, J., & Todd, D. (2008). Methodological challenges of evaluating the impact of the Global Environment Facility's biodiversity program. *Evaluation and Program Planning, 31*(3), 231–240. https://doi.org/10.1016/j.evalprogplan.2008.03.002.

Van den Berg, R. D., & Cando-Noordhuizen, L. (2017). Action on climate change: What does it mean and where does it lead to? In J. I. Uitto, J. Puri, & R. D. van den Berg (Eds.), *Evaluating climate change action for sustainable development* (pp. 13–34). Springer.

Walker, B. H., Anderies, J. M., Kinzig, A. P., & Ryan, P. (2006). Exploring resilience in social-ecological systems through comparative studies and theory development: Introduction to the special issue. *Ecology and Society, 11*(1).

Waltner-Toews, D., & Kay, J. J. (2008). Implementing an ecosystem approach: The diamond, AMESH, and their siblings. In D. Waltner-Toews, J. J. Kay, & N.-M. E. Lister (Eds.), *The ecosystem approach: Complexity, uncertainty, and managing for sustainability* (pp. 239–255). Columbia University.

Waltner-Toews, D., Lister, N., & Bocking, S. (2008). A preface. In D. Waltner-Toews, J. J. Kay, & N.-M. E. Lister (Eds.), *The ecosystem approach: Complexity, uncertainty, and managing for sustainability* (pp. ix–xviii). Columbia University.

Waltner-Toews, D., & Wall, E. (1997). Emergent perplexity: In search of post-normal questions for community and agroecosystem health. *Social Science & Medicine, 45*(11), 1741–1749.

Weiss, C. H. (1997a). *Evaluation: Methods for studying programs and policies* (2nd ed.). Prentice Hall.

Weiss, C. H. (1997b). Theory-based evaluation: Past, present, and future. *New Directions for Evaluation, 1997*(76), 41–55. https://doi.org/10.1002/ev.1086.