Journal of MultiDisciplinary Evaluation
Volume 19, Issue 45, 2023

JMDE
Journal of MultiDisciplinary Evaluation

ISSN 1556-8180
http://www.jmde.com

A Complexity-Based Plan for Evaluating Transformation

Jonathan A. Morell
4.669… Evaluation and Planning

Background: "Transformation" is becoming an important concept in evaluation.

Purpose: Promote a complexity approach to evaluating transformation.

Setting: Not applicable.

Intervention: Not applicable.

Research Design: Not applicable.

Data Collection and Analysis: Not applicable.

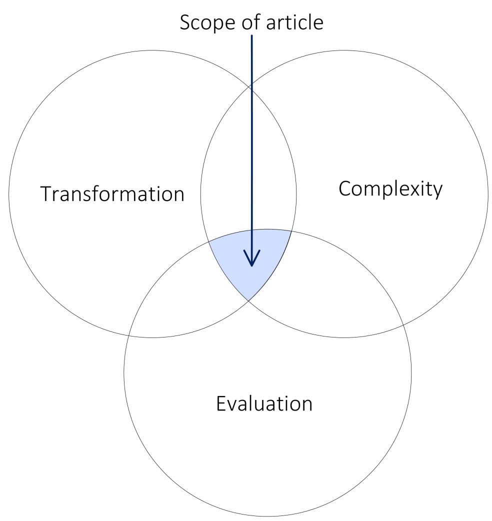
Findings: Not applicable.

Keywords: *transformation; complexity; evaluation*

Introduction

The scope of this paper lies at the intersection of three topics—transformation, complexity, and evaluation (Figure 1). I will argue that complexity science has much to say about the models, methodologies, and data interpretations that make up so much of the evaluation enterprise. I will also argue that in order to evaluate transformation, a working definition of "transformation" must be complementary with what we know about complex behavior. Much of what will be said here can be applicable to applying complexity to any evaluation, but the specific purpose of this article is to shed light on how transformation should be evaluated. Thus, much of the text is specifically about what transformation is, and how complexity can be used to evaluate transformation.

Figure 1. Scope of This Paper at the Intersection of Three Topics



Evaluation's Stance on Transformation

In evaluation the discussion about transformation runs along two lines—the social organization of the field, and intellectual content (Ofir & Rugg, 2021; Patton, 2020; Picciotto, 2021; Van den Berg et al., 2021). Examples of the former include greater gender and geographic diversity in the evaluation workforce; greater emphasis on equity and inclusion; formal organizations (e.g., the International Evaluation Academy); tighter linkages among science, research, policy, evaluation practice and evaluation theory; more emphasis on professionalization; and outreach to young and emerging evaluators. Some examples of the latter are recognition of ecosystem health in evaluation designs, distancing from a neo-liberal mindset, less emphasis on accountability and meeting the needs of donors, more emphasis on sustainability and long-term impact, recognition of bounded rationality in decision-making, more attention to non-Western perspectives, and full-cost accounting. The focus of this paper is strictly on the latter, i.e., technical aspects of evaluating transformation.

Bringing Complexity Into the Evaluation of Transformation. There is a lot of writing and theorizing about how complexity can be applied in evaluation. [1] A much smaller subset of that literature delves specifically into technical issues concerning how complexity science can be applied to the evaluation of transformation. (Morell, 2021; Zazueta, 2019; Zellner et al., 2012). That is the literature I will draw from. Barbrook-Johnson et al. (2020) discovered that policy evaluators are aware

[1] As a gateway into this literature, two sources are recommended. One is a 2021 special issue of *New Directions for Evaluation* edited by Gates, Walton, Vidueira, and McNall. The editors' introduction to the issue, "Introducing Systems- and Complexity-Informed Evaluation" (pp. 13–25), is especially helpful. The second is Barbrook-Johnson, Proctor, Giorgi, and Phillipson's 2020 paper in *Evaluation*, "How Do Policy Evaluators Understand Complexity?" Between the articles and references cited in these publications there is an extensive set of relevant writing to be found.

of complexity, but that they relate to complexity in "terminological and analogical" rather than "literal" terms (p. 327). Presumably, this is true for all evaluators. This paper will seek to move evaluators (and the store of knowledge capital in evaluation) toward a more "literal" understanding of how complexity can be applied to the evaluation of transformation.

Plan for Making the Case

In this paper I will argue that three specific constructs from complexity science provide an understanding of complexity that allows practical and meaningful evaluation of transformation. To do so I will proceed through six topics: (1) What is transformation?, (2) Relevant constructs from the field of complexity, (3) A defensible choice for including or not including complex behavior, (4) Characteristics of models of transformation, (5) Methodologies to evaluate transformation, and (6) How should an evaluation of a specific transformation be constructed?

What Is Transformation?

Transformation is characterized as a new normal that is made up of multiple components that support each other. As progress toward transformation is made, change trajectories are nonlinear, often unpredictable, and subject to tipping points. Markers of transformation are observable and measurable.

Relevant Concepts From the Field of Complexity

There is an extensive literature on the subject of complex phenomena, rich with many terms to describe those phenomena. Only a few of these are needed to provide a set of intellectual tools that will allow transformation to be evaluated with respect to complexity. Those constructs are emergence, sensitive dependence, and attractors.

A Defensible Choice for Including or Not Including Complex Behavior

A scenario will be presented in which a defensible case can be made for using either traditional if→then logic, or a model that recognizes complexity. While it is true that one could make a case that systems are *always* complex, it is also true that drawing on complexity is not always needed to provide useful knowledge. Appreciating this reality has implications for evaluating transformation.

Characteristics of Models of Transformation

Models with varying levels of reliance on complexity are presented, followed by an explanation of complex behaviors that would be useful in understanding programs that are designed to effect transformation.

Methodologies to Evaluate Transformation

Methodologies needed to evaluate transformation in terms of complexity are well known and well practiced in the evaluation community. Less-familiar methodologies should also be considered, but by and large, what we already know is most of what we need.

How Should an Evaluation of a Specific Transformation Be Constructed?

This section will show how the information presented in the article can be applied to an evaluation in a specific case.

Terminology

I will employ three terms that are unfamiliar in common discussions about complexity.

Complex behavior: Instead of "complexity" I will refer to "complex behavior." I do this because if someone told me that a system is complex, I would have no idea what to do about it. But I do know what complex systems *do*, and that is knowledge I can use to develop models, construct methodology, interpret data, and discuss findings with customers.

Model: I use the term "model" instead of the more familiar "logic model" or any of the terms closely associated with it (e.g., "theory of action" or "program theory"). I do this because my focus is on "model" as the term is employed in the scientific enterprise, i.e., as a simplified portrayal of reality that serves to guide inquiry. For a full treatment of the roles of models in inquiry, see Box (1979), Frigg and Hartmann (2018), and Rogers (2012).

Complexity science: No unambiguous definition of "complexity" exists, but it is fair to say that "complexity" is the subject of "complexity science." Complexity science has its own definitional problems, as it has a long history of intellectual roots in many different fields (Bar-Yam, 2021; Castellani & Gerrits, 2021). Precise definition of a field, however, is a sign of maturity, not a precondition for beginning inquiry. [2] What is needed to begin is a group of people who identify with a field, publications and educational programs that claim membership in that field, and research that falls within the boundaries of those publications and programs. In this sense complexity science can be said to be a field in its own right (Phelan, 2001).

What Is Transformation?

Definitions

What does it mean to say that "transformation" has taken place? Various authors involved in evaluation have advanced answers to this question.[3]

- Interventions or series of interventions that support deep, systemic, and sustainable change with the potential for large-scale impact in an area of a major development challenge (International Evaluation Group, 2016).
- There are different kinds of changes. Those that are *transformational* represent significant changes in direction and/or in size. In contrast with changes *a la Gattopardo* (i.e., changes so that everything remains the same), or micro-changes, transformational change of society makes a major and durable difference (Feinstein, 2019).
- Transformational change is a change of systems, not just singular developments, and involves multiple actors at multiple levels. Transformational change constitutes deep, fundamental change that disrupts the status quo, and sustains that change over a long period. Transformational change by itself has no normative connotation; values are added by defining a transformation goal (Initiative for Climate Action Transparency [ICAT], 2020).
- Deep, systemic, and sustainable change with large-scale impact in an area of global environmental concern (Batra et al., 2022).

Running through these definitions is the theme of a "new normal"—a qualitative change in default conditions that cover broad sectors of society. With respect to these new conditions, two possibilities need to be considered. First is ICAT's point about value: "Transformational change by itself has no normative connotation; values are added by defining a transformation goal" (p. 13). Second, transformational change can result from deliberate action or from the natural consequence of activity in the world. This paper is concerned with only one of these possibilities—deliberate action to effect desirable change.

A few examples serve to illustrate the idea of transformation that fits the above definitions. Consider the transition from gas lighting to electric lighting, income tax as a form of taxation in the United States, and capitalism in place of mercantilism. All these transitions can be characterized as deep, systemic, reaching across multiple actors and levels, significant, and as step changes that characterize fundamentally new conditions in which society functions.

A Definition That Can Guide Planning

It is easy to determine in retrospect whether a transformation has taken place. But evaluation is not history. It is prediction. Our customers want to know what will happen if they take action and whether they have made any progress in changing the world according to their designs. (Of course, they know that we cannot make guarantees, and that their efforts may not produce precisely what was intended. But our customers do look to us to increase the probability of success in achieving results that are close to their intentions.) I realize that this statement is a bit extreme because many evaluation approaches deal with a lot more than prediction and others focus on events that cannot be anticipated. Examples include a focus on "merit, worth and significance," theory-based evaluation, developmental evaluation, goal-free evaluation, a focus on unintended consequences, systems thinking, and triple-loop learning. These approaches to evaluation have dimensions that go beyond helping stakeholders increase the probability that deliberate action will affect particular aims. Still, some kind of predictive model usually accompanies any evaluation, no matter how it is conducted. The model will specify a series of actions or events that are related in chronological

[2] Thanks to Roger Brown, author of my first social psychology textbook, for this insight.

[3] It is worth noting that these examples are a minority of cases in which "transformation" is part of the title or abstract. This is because most of the articles that purport to be about transformation are not concerned with a definition of the term. Rather, they assume that large-scale change is involved. To provide a sense of the many cases that fall into this category, consider Greenhalgh et al. (2009), Firmansyah et al. (2021), Ibrahim et al. (2017), and Matz et al. (2018).

order, and which end in specified outcomes. Our stakeholders know full well that to be successful, they need to know more than whether particular outcomes have appeared. Evaluation has responded to this need by developing approaches to evaluation that are designed to provide that knowledge. But as the ubiquity of sequential models indicate, there is almost always an implicit assumption that however evaluation is conducted, it will produce knowledge about the relationship between program action and desired outcome. Accomplishing this task is part of the process of evaluating transformation.

If we do believe in stating outcomes, then we must enter the work with a definition that we can measure. As a basis for the conversation to follow, I offer this generic definition:

> Transformation has occurred if, in geopolitical boundary X, approximately 80 percent of [insert outcome of interest], has maintained itself for about five years, and the new state of affairs is generally accepted as the default condition for how life is lived.

Note that this definition is agnostic with respect to whether the change is intended as a result of deliberate action or is simply a change that has been observed in nature. Our interest is in bringing about specific change, but the dynamics of transformation are the same whether we are evaluating deliberate efforts at change, or we are simply studying unplanned change.

Of course, the details of geography, implementation, and acceptance may differ radically for different types of transformation. But the point remains. A useful definition of transformation can be expressed in terms of geography, level of implementation, time, and acceptance. Table 1 presents the utility of this definition.

Table 1. Aspects of Definition of Transformation

No specific transformation	The definition says nothing about what is being transformed. As a few of a multitude of examples, it could be green energy, level of literacy in a nation, or incidence of malaria. For different types of transformation, we might want to adjust our approximations for boundaries, level of use, time frame, or acceptance in the population. We might want to add an outcome criterion. But the basic definition would still work—it would indicate a new normal.
Geopolitical boundary	A reasonably good proxy for the combined effect of a multitude of factors.
Level of use / implementation	"Eighty percent" is an approximate level that would be truly different from the old state of affairs.
Time	"About five years" seems like an adequate span of time to indicate that the change will endure.
Culture	"Generally accepted" is a loose term that indicates that there is social support (or at least acceptance) of the new condition.
Measurable but imprecise	Note that each component is stated in imprecise terms—"approximate," "about," "generally accepted." These qualifiers are deliberate because while measurement is critical, we do not know enough to define transformation in specific terms, and in any case, we want to avoid false precision. Suppose we found a smaller geographical area, four years, and ninety-five percent? Would we judge the program efforts a failure? I wouldn't. What matters is that these kinds of numbers indicate a "new normal" that is profoundly different from what went before—a desirable new social attractor.

If we accept this definition, can we use it to measure progress toward transformation? Not easily, because the components of "transformation" are networked, and under those conditions the trajectory of change can be complicated and difficult to predict. As an example, it is not hard to imagine that as the adoption of green technology spreads over geographic terrain, more people will observe that adoption and hence hasten the time it takes to reach eighty percent. Or increasing levels of adoption may spur infrastructure that will accelerate adoption across geographical boundaries. Contrariwise, opposition to geographical spread may increase as those opposed to the change observe its progress. If the change pattern is unknown, there is no meaningful way to interpret data that describes progress toward the goal of transformation.

What Can Be Predicted?

I do not mean to imply that change patterns in networked scenarios are always unpredictable. In fact, a great deal of regularity is predictable and measurable. Some examples of predictable change include innovation adoption, growth driven by preferential attachment, and scaling. Innovation adoption: The familiar "S" pattern for innovation adoption shows up reliably in a very large number of innovation adoption scenarios (Greenhalgh et al., 2004; Rogers, 2003). Preferential attachment: Network structures frequently grow by a process of "preferential attachment," a process by which the connectedness of a node increases as a function of the number of nodes it already has (Barabási & Albert, 1999; Jeong et al., 2003). The result is a fractal structure. Examples of this pattern include Internet and snowflake growth. Scaling in urban settings: Numerous characteristics of urban existence scale in a common, predictable manner. Examples include road length, number of patents, crime rates, and number of restaurants. (In fact, these patterns can be found across a very wide range of domains outside of human-scale phenomena (West, 2017). For many other examples, check out the Models Library in NetLogo.[4] It is full of complexity-based models that are highly predictable.

From the point of view of evaluating transformation, the problem is not being able to predict that a pattern will manifest. Rather, the problem is knowing (at least approximately) the specifics. As an example, it is one thing to know that an "S" curve will develop. It is quite something else to know where the inflection points will be, or to be able to predict the growth patterns between inflection points. There is much we do not know.

What Are the Nodes in the Network? We do not have a theory to tell us what the nodes in the network should be. Consequently, it makes little sense to specify components of transformation, measure change in each, and reach a conclusion about how much transformation has taken place. Take my example of geopolitical boundaries. It may make practical sense to define such a boundary as a state, a province, or an entire country. But that area is made up of subareas—neighborhoods, cities, counties, and the like, each with its own growth pattern. Is it meaningful to look at how change grows in terms of influences between any of the subunits? I have no idea, but I know that it is a question worthy of consideration. Recall that geography is being used as a proxy for many factors, e.g., population, households, businesses, availability of equipment, expertise to install and maintain equipment, acceptable cost, political consensus, regulation, and peer pressure. What geographical size of political unit comprises a meaningful proxy for these variables?

Loci of Change. Suppose we did have a defensible reason to specify network nodes, or at least some of them. Could we use that knowledge to identify high-leverage areas within that network? We could not, because there are several difficulties. One problem is that we cannot answer an important question about the allocation of resources. Would we be better off trying to effect large change in some nodes, or small change in as many of them as possible? One reason we cannot answer this question is because the emergent behavior in evolving networks (driven by sensitive dependence) is so unpredictable. Another reason is that we do not know the network structure. While we may be able to do an adequate job of identifying the nodes in the network, we have very little idea of the edge structure. Finally, we know little about the relative importance of each node.

Shape of Change Over Time. It is safe to assume that degrees of transformation will change nonlinearly. Transformation involves innovation adoption, and an enormous body of research reveals adoption curves with inflection points

[4] NetLogo is a free application that allows one to build agent-based models. It is available at https://ccl.northwestern.edu/netlogo/

(Greenhalgh et al., 2004; Rogers, 2003). Every discussion on innovation adoption that I have ever seen describes two inflection points: one when the adoption rate increases sharply, and the second when the rate asymptotes because such a large percentage of the population has adopted the innovation. However, when it comes to a construct as complicated as "transformation," there is no compelling reason to assume that there will be only one acceleration phase and one deceleration phase. After all, we are dealing with multiple innovations, each with its own adoption curve, and most being networked with many others. In any case, whatever the specifics of transformation, networked behavior is involved, and nonlinear change is a very common characteristic of change in networks. Between change in each variable and emergent effects, who knows what the overall shape of change might be?

Where Is the Tipping Point? Usually when evaluators and planners invoke the term "tipping point" to characterize change, they are referring to a relatively short time span during which a rate of change transitions from linear to nonlinear, or perhaps from somewhat nonlinear to dramatically nonlinear. However, "tipping point" has two different meanings (Lamberson & Page, 2012). One is the common meaning that I have been presenting. But another is a "contextual tip." As the authors put it, "A contextual tip occurs when a gradual change in the value of one variable leads to a discontinuous jump in some other variable of interest" (p.1). They give the example of human rights conditions deteriorating to the point of creating the potential for an uprising. You and I tend to think of measuring the rate of civil unrest, but identifying the contextual tip would require tracking something else entirely—the state of human rights.

As an example, with respect to transformation to green energy use, imagine the commitments that auto makers have been making to phase out the production of internal combustion engine automobiles. To date there has been no phaseout, so the correlation between production volume and fossil fuel use is meaningless. But it is not hard to imagine that when the history is written, those commitments will prove to have been a critical contextual tipping point. When measuring transformation, one might argue that the manufacturers' production commitments were far more meaningful than change in the percentage of internal combustion engines on the road.

Relevant Constructs from the Field of Complexity

What constructs drawn from complexity science can inform evaluation of transformation? There are many candidates. Bicket et al. (2020) suggest adaptation, emergence, self-organization, unexpected indirect effects, feedback, levers and hubs, nonlinearity, domains of stability, tipping points, and path dependency. To this list one might add attractors, stigmergy (Theraulaz & Bonabeau, 1999), and scaling factors (West, 2017).

Many of these concepts overlap. For instance, attractors can be thought of as the "space" in which a system is found. "Self-organization" is a property of a system by which it can maintain order solely from interactions among its internal components following rules that are not dependent on rules that emanate from outside of the system (Comfort, 1994). "Stigmergy" is a concept that was first developed to understand insect behavior. Essentially, the "plan" is embedded in the history of activity that individual actors encounter (Theraulaz & Bonabeau, 1999). "Attractors," "self-organization," and "stigmergy" are not synonyms. Each has unique meaning and unique use. Still, there is enough similarity in the behavior they describe that effective use of complexity might call for only one of these concepts. My contention is that using only three constructs from the field of complexity science will suffice to do an efficient job of designing effective evaluations of transformation. Those three are emergence, sensitive dependence, and attractors.

Emergence and sensitive dependence hold a special place in the population of complex behaviors because they have the potential to change how evaluators think about outcomes and causal change. *Emergence* implies a qualitative change that cannot be explained by changes in constituent parts. This has implications for program theory and for understanding how change happens. It also has implications for constructing or choosing measurements. *Sensitive dependence* implies that because local change can affect the long-term trajectory of a system, familiar patterns of long chains of if→then links may not be accurate portrayals of links between program and outcome. When causal chains do not work, our inclination is to fault ourselves. Sensitive dependence implies that the fault is in the system, not in ourselves. It does not matter how much we know; it is impossible to trust a single path. *Attractors* raise the possibility that even if unique paths cannot be trusted to repeat, there may be a class of paths that lead to the same outcome. Also, visualizing attractor spaces is a useful way of conceptualizing both resistance to change and its reflection, sustainability.

The definitions of the constructs presented in Table 2 are quoted from the glossary at the Santa Fe Institute's Complexity Explorer site. The glossary also has excellent definitions for many other aspects of complexity.

What is it about these three complex behaviors that is relevant to evaluating transformation? The answer to this question appears in Table 3.

Table 2. Complexity Construct Definitions

Emergence	A process by which a system of interacting subunits acquires qualitatively new properties that cannot be understood as the simple addition of their individual contributions (Santa Fe Institute, n.d.-b, para. 1).
Sensitive dependence	A system's sensitivity to initial conditions refers to the role that the starting configuration of that system plays in determining the subsequent states of that system. When this sensitivity is high, slight changes to starting conditions will lead to significantly different conditions in the future. Sensitive dependence on initial conditions is a defining property of chaos in dynamical systems theory (Santa Fe Institute, n.d.-c, para. 1).
Attractor	In dynamical systems, an attractor is a value or set of values for the variables of a system to which they will tend towards over enough time, or enough iterations. Examples include fixed-point attractors, periodic attractors (also called limit cycles), and chaotic (also called "strange") attractors (Santa Fe Institute, n.d.-a, para. 1) While the above is the classic definition that originally came out of complexity science, the term also has meaning in social terms: "Social attractors define a specific subset of states that a social system may take, which corresponds to its normal behavior towards which it will naturally gravitate" (Systems Innovation, 2020b).

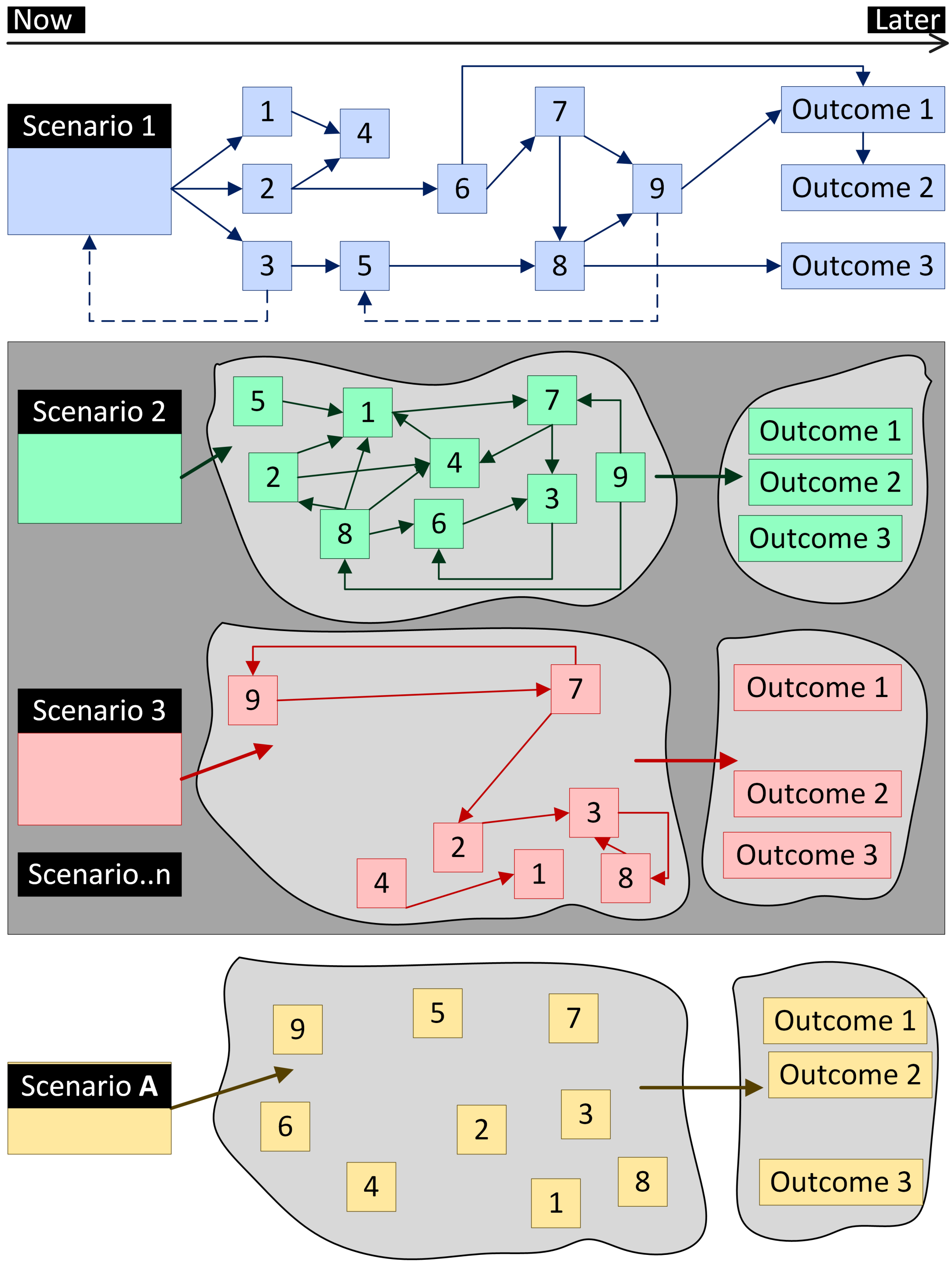
Table 3. Relevance of Complexity Constructs to the Evaluation of Transformation

Emergence	"Emergence" allows us to think of groups of variables that effect change collectively under a condition such that the contribution of each variable is undefinable. An example helps. Imagine an automobile. The automobile is greater than the sum of its parts. (Or to put this in a less familiar but more correct way, the automobile is *qualitatively* different than the collection of its parts.) But now imagine the cylinder. We know what it is, how it is constructed, and how it contributes to the automobile. As the car moves, the cylinder keeps working as designed. Its identity perseveres. The unique contribution of the cylinder to the car's operation endures. Now imagine a traffic jam. Cars are its constituent parts, but the traffic jam cannot be explained by analyzing the behavior of each car. It is also qualitatively different than the collection of cars that make up the jam. The traffic jam must be understood as a holistic manifestation of the overall traffic dynamics. In such an explanation, the unique contribution of each car disappears. Many other examples can be cited, but a few are "vitality" of urban life, an economy, and the intellectual environment of a university. <u>Implications for evaluation</u>: Traditional evaluation reasoning would assume that transformation can be brought about by understanding each of the model's elements and the relationships among them. But what if the cause of transformation was an emergent consequence of all the elements working together (often driven by network relationships)? That would drive a different evaluation methodology and a different approach to interpreting data. It still might be a good idea to measure change in each element of the model, but it would be a mistake to try to tease out the unique contribution of each of those elements.
Sensitive dependence	The concept of sensitive dependence implies that small variations at any point in the model may result in a unique path through the model. <u>Implications for evaluation</u>: If we truly believed that only one path would lead to success, and that sensitive dependence implies that the planned succession of activities is unlikely to proceed as planned, then it would be a mistake to evaluate (not to mention implement) such a fragile program.
Attractor	With respect to evaluating transformation, the idea of an attractor implies that there *may* be a range of paths through the system such that any movement through the model that falls within that range will lead to transformation. I am not aware of any phenomenon like this that is discussed in the evaluation literature, but I do know from personal experience that it happens in simulated social systems that are cast as agent-based models. <u>Implications for evaluation</u>: One can think of multiple paths that end up in (approximately) the same place as defining an attractor space that leads to that outcome. Paths within the attractor boundaries lead to the desired outcome; paths outside the attractor boundaries lead elsewhere. A good way to conceptualize attractors is as a "basin of attraction" that has a topology (Systems Innovation, 2020a). Topology implies steepness, depth, distance, and shape. That can make for useful visualization, and beyond the visuals it can lead to theory about change and stability. Of course, one does not need the term "attractor" to describe this situation. One could just assert that there is more than one way to reach a goal. But thinking in terms of attractors provides a sense that all kinds of movement may take place through a model, and that it is possible to conceptualize a "space" within the model such that all paths within it lead to the same outcome. Related to this concept are tipping points, and state changes. "Tipping point" refers to the region where one attractor transitions to another. "State change" refers to the characteristics of the phenomenon that change upon transition from one tipping point to another.

Implications for Models

Taken together, what are the implications of these complex behaviors for evaluating transformation? To address this question, I present Figure 2., which models three scenarios that progress from a traditional evaluation to a model that fully recognizes the combined consequences of complex behavior. How might each of these scenarios be applied to the evaluation of transformation?

Figure 1. Increasing Recognition of Complex Behavior in Three Models



Scenario 1 is problematic for several reasons: (1) It asserts that all the model elements, and all the relationships among them, can be specified in advance. If it shows up in the model, it plays an important enough role that it must be included in an evaluation. (2) It does not allow for the possibility that different configurations of the elements can lead to transformation. They are all needed, and they are needed in very specific ways. (3) It assumes that each individual element needs to be evaluated for its unique contribution to achieving transformation. It does not allow for understanding that the "real" reason for success is an emergent characteristic of interactions among the individual elements. It follows the "automobile" model of causation rather than the "traffic jam" model.

Scenario 2 moves further toward recognizing complex behavior. The model acknowledges that a transformation activity has specific immediate objectives, but that transformation cannot be attributed to specific elements of the model. By "cannot" I do not mean that we lack an adequate methodology. The message is that emergence is operating, i.e., whatever the "it" is that is bringing about transformation, that "it" needs to be thought of as a phenomenon in which each individual element in the model has lost its unique identity as a relevant factor in effecting change.

Note that Scenario 2 has multiple configurations that lead to success. This indicates the idea of an attractor space within which multiple success paths are possible, that sensitive dependence may be at play to determine which path is followed, and that the paths may lead to similar, but not necessarily the same, outcomes. The problem with Scenario 2 is that it implies that all possible success paths can be identified. Complexity tells us that we cannot do this. In any case, if we thought that we could identify likely paths to success, it would be necessary to implement a design to test each path.

Scenario 3 is the simplest of all because it is modest. It says that we can identify the elements that are needed to lead to success, but that we do not know if they are all needed, or in what configuration they must be arranged. All it says is "these are the elements that probably matter; any number of relationships among them may result in transformation; and however they interact, success cannot be attributed to that particular configuration, but to the emergent effect of all those networked interactions."

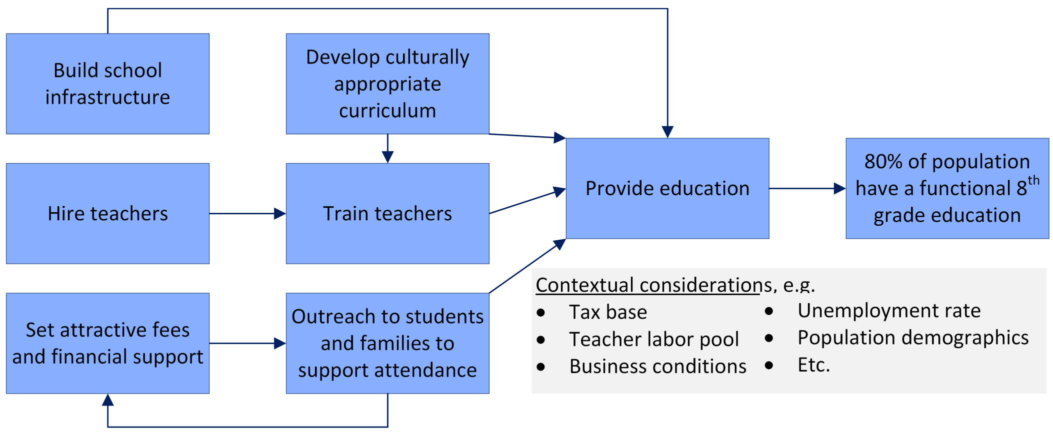
What Can We Know With Respect to Scenario 3? A previous section (What Can Be Predicted?) addresses various aspects of this question, but here it may be worth reiterating two specific issues with respect to Scenario 3. The message in Scenario 3 is not that we are destined to be ignorant about what works and what consequences an intervention may have. On the contrary, we can know a great deal. It is just that what we can know is different from the kind of knowledge we are used to dealing with.

- What, and how reliably, can we predict? It is one thing to say that because of sensitive dependence, we cannot predict the exact path that will be traveled through a system. It is quite something else to say that the outcome of a complex system is unpredictable. On the contrary, although details may be unknowable from one iteration to the next, it is entirely possible that the outcome will reliably appear. We may be able to learn a lot about the attractor space in which successful paths are contained.
- What can we know from retrospective analysis? The essence of a complex system characterized by sensitive dependence is that we cannot know the details of what will happen, even if we can know the final state. The situation changes with retrospective analysis. Once events have occurred, we can look back and see what happened. This is a strong theme in process tracing (Wadeson et al., 2020), it is the main ingredient in accident investigation (U.S. Department of Energy, 2012), and it is a critical element of cross-case comparison (Byrne & Uprichard, 2012). In fact, methods like these can play an important role in defining path-attractor shapes and thus shedding light on predicting outcomes.
- A priori and a posteriori reasoning. Once change has taken place, Scenarios 1 and 2 can both be cast in the form shown in Scenario 1. Doing so can produce much understanding, but the path discovered cannot be assumed to be predictive for future change.

Defensible Choice for Including or Not Including Complex Behavior

Imagine a program whose purpose is to increase literacy in some geographical area. An outcome for a program like this might be stated as "Eighty percent of the population will have a functioning eighth-grade education." A model for such a program might look like Figure 3 (In this paper I am dealing with models as they are commonly used in our field. We would do well, however, to incorporate more rigorous reasoning about models in our work.[5])

Figure 3. If→Then Model with Discrete Relationships



Of course, other people would draw this model differently with respect to layout and color, elements that would be included, and relationships among elements. They might also include a different set of contextual variables. But for all those differences, the models would be qualitatively the same. The models would specify elements and relationships among them, and contextual characteristics that would be enlightening when it came to data interpretation.

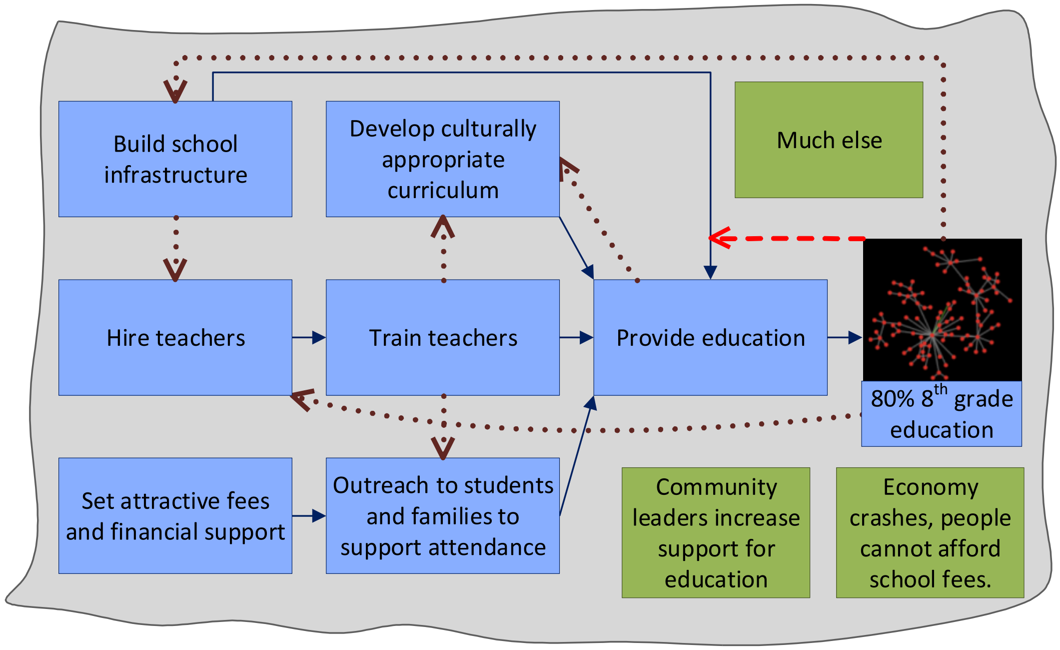
Figure 3 shows relationships that form a plausible model of the important program elements, their relationships, and how they affect education. We chose these because we thought that they are what is needed to devise and deploy a successful program. But complicating the model is the real possibility that many other connections either exist at the beginning or can develop over time. These are depicted in Figure 4. Figure 3 and Figure 4 differ in four consequential ways.

[5] This section (and in fact this entire paper) is based on a treatment of models that is commonly found in the evaluation literature. Those models usually contain arrows among elements but contain no information about necessary and sufficient conditions. Nor do discussions of methodology or findings reflect on this question. In the absence of such information, a reasonable assumption is that when we develop our models, we assume that all inputs are always necessary. This is extremely problematic because models with too many "and" inputs are likely to freeze. One might even say that most of the models we use portray programs that are almost certain to fail. Adding "and/or" specifications is not trivial because doing so requires a lot more thinking about program theory and theories of action than we usually invest. Still, it is worth the effort, at least over critical sections of our models. We should also appreciate the fact that the "and/or" question scratches the surface of an entire field of study known as the logic of conditionals (Egré & Rott, 2021). But even without recourse to this literature, addressing the "and/or" question in our models would be spur a much better understanding of the model relationships that we study.

Another important domain that we would do well to take seriously is some of the work coming out about new algebraic notations that provide a perspective on causality that is not reflected in our traditional understanding of experimental and quasi-experimental designs. The most accessible source for this material is Pearl's *Causality: Models, Reasoning and Inference* (2009).

It is also important to realize that "model" can take many forms in different fields. As Wikipedia puts it, quoting Kenneth Bollen, "A statistical model is usually specified as a mathematical relationship between one or more random variables and other nonrandom variables. As such, a statistical model is 'a formal representation of a theory' " ("Statistical Model," 2023, para. 2). No boxes or arrow. All equations. This paper, however, is about how evaluators go about their business.

Figure 4. Discrete Relationship Model Embedded in Context



1) Connections Among Elements Are Missing

The brown dotted arrows show a few of the many connections that are possible. I have no idea how these relationships would affect the model, but I can conjure a few possibilities. (1) New schools might make it easier to recruit teachers. (2) Experience training teachers may provide input to make the curriculum more culturally appropriate. (3) Program success may make the education profession more appealing and thus make it easier to hire teachers. (4) Program success may contribute to the case for investing in infrastructure. (5) Because teachers are likely to live in the communities where they work, training may support student recruiting. All these connections are reasonable surmises, all or none may occur, any number of interactions among them are possible, and none is direct enough or strong enough to deserve investment of evaluation resources.

2) The Program Exists in an Environment

Green boxes in Figure 4 show some of the many environmental events that might influence the outcome. Most notable is "Much else." A great deal might be going on, some of it obvious and some of it powerful but subtle. Environmental scanning might detect these events, but doing it well is no small task. One needs to know how to look and where to look (Gordon & Glenn, 2009).

3) Outcome Patterns Are Not Obvious

Change in outcome may not be linear. One example is innovation adoption, which exhibits the familiar "S"-shaped adoption curve (Greenhalgh et al., 2004; Rogers, 2003). Another example is the fractal growth pattern that I put into the "outcome" element of Figure 4. Such a pattern is plausible because it characterizes preferential attachment growth (Barabási & Albert, 1999). It makes sense because one might expect that a person's probability of going to school will be affected by the number of people that he or she knows who also attend school.

Change patterns matter because they speak to methodology and data interpretation. No change over a long time may say nothing about a program's potential to succeed. What appears to be a dramatic inflection point may asymptote at a depressingly low level of change.

4) There Are Linkages Among Model Elements, Environmental Events, and Outcome Patterns

Complicating the model further is the possibility of linkages among network relationships, environmental events, and outcome patterns. The red dotted arrow in Figure 4 illustrates one of many possibilities. In that example, the visibility of graduates' earning power mediates the relationship between building infrastructure and providing education.

A pithy summary of all the above is that Figure 4 recognizes complex behavior.

Choosing Between the Complexity-Based and the Traditional Model

I have presented two models for a literacy program. Figure 3 is a traditional if→then model that can be justified as a practical explanation of how the program works and what it will accomplish. Figure 4 sets the program within a context comprised of environmental influences, unknown networked relationships, and a fractal growth pattern for outcome change. We know that all models are wrong, but that some are useful (Box, 1979). So, what are we to make of the models in Figures 3 and 4? Here are some considerations:

1. *Some* elements in Figure 4 that are missing in Figure 3 may or may not be consequential, but we cannot specify which.
2. Consequential relationships that are not present at the beginning of the program may develop over time. And if this is so, we must also allow for the possibility that relationships that do exist may atrophy over time.
3. Evaluating with respect to Figure 4 will require a much more complicated and expensive methodology.
4. Figure 3 contains elements and relationships that stand a good chance of leading to useful knowledge about the program, but by using that model, much worthwhile knowledge will be missed.

Table 4 summarizes the advantages and disadvantages of using each model to guide the evaluation.

Table 4. Advantages and Disadvantages of Traditional and Complexity-Based Models

	Traditional		Complex	
	Yes	No	Yes	No
Program environment considered?		•	•	
Cost easily accommodated?	•			•
Growth patterns recognized?		•	•	
Data requirements manageable?	•			•
Easily understandable to stakeholders?	•			•
Important elements and connections missing?		•		•
High % of findings provide actionable information?	•			•

So, which model to choose? Personally, I would go with the traditional model. It is practical. It stands a good chance of producing actionable knowledge. Customers will understand it and will be able to use it to interpret findings. I am by no means arguing that if an if→then logic is employed, an evaluation should ignore the possibility that other factors may be in operation. Other factors are always operating, and sometimes they may make a difference. We should always make the effort to keep an eye out for them. I am claiming that there are times when it makes practical and intellectual sense to have a narrow focus. Of course, others might choose the complex model, and they would certainly be able to make a strong case for their choice. *What matters is that either choice is defensible.*

But what if the choice of the simple model turned out to not work? Evaluation can go wrong for many reasons, but a prominent candidate for failure would be that our choice to go simple was wrong. The modeling methodology we chose (if→then modeling) was not up to the task. The evaluator's response to this eventuality is simple: "We made a reasonable guess based on the advantages and disadvantages of each approach, and we got it wrong." That won't be the first time we did our best and implemented a suboptimal methodology. We do our best to salvage the evaluation.

The overall message about choosing between these models is that there are many evaluation scenarios where it is perfectly reasonable to ignore complex behavior. Complexity is no more a universal requirement than is Bayesian reasoning or content analysis of interview data. As we shall see, much evaluation dealing with transformation should rely on simple if→then logic. But to evaluate transformation writ large, data from if→then evaluation must be combined with, and interpreted in terms of, complex behavior. I will elaborate on the reasons for this in the Methodologies to Evaluate Transformation section.

A Complex Behavior Justification for Using If→Then Logic

The discussion above has taken a commonsense approach to justifying the use of traditional logic. The argument was that given the practicalities, it makes sense to use the simpler method. It assumes that any unforeseen dynamic (e.g., all the additions added to Figure 3 to generate Figure 4) will be inconsequential. However, the phenomenon of sensitive dependence also implies that evaluation based on if→then logic can be appropriate. To see why, think in terms of probabilities. What is the probability that seemingly insignificant changes will build on each other in such a way as to have profound consequences? The default answer to this question is "low." The fact is that our world has evolved to be stable. Most mutations do not confer a competitive advantage. Most perturbations to systems do not have more than transitory effects. To be sure, small change can represent a tipping point—a phase transition from one state to another. If such a tipping point is expected, then if→then logic will suffice. It is only when tipping points lead to unforeseen conditions that we can get into trouble. That can happen. But does it happen often enough that we should forfeit all the advantages of simpler evaluation methodologies?

Characteristics of Models of Transformation

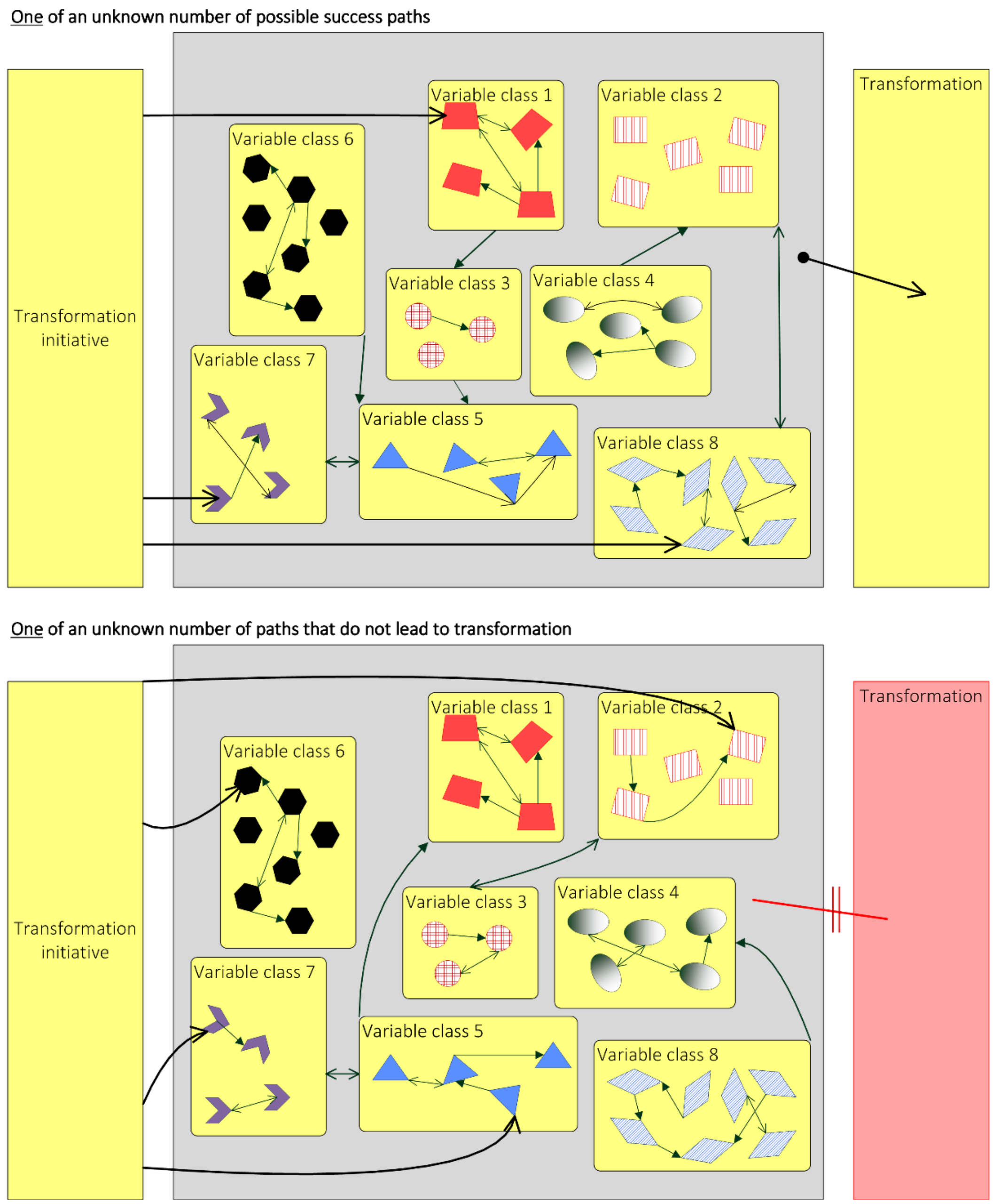
Any model used to guide an evaluation of transformation must recognize that:

- Whatever "transformation" is, a great number of diverse, richly connected actions are needed to bring it about.
- Many consequential connections cannot be known in advance. They can only be detected in retrospect.
- An enumeration of the necessary actions needed to bring about transformation often eludes our best social science knowledge. Seldom do we know enough to identify the important activities and connections among them.

Because of the above, we cannot construct a model along the lines of Scenario 1 Figure 2, or of Figure 3—i.e., an unambiguous model—that is good enough to predict transformation as an outcome. We need a model that recognizes that evaluating transformation involves evaluating a phenomenon that demonstrates complex behavior.

This section contains a generalized model for evaluating transformation that can be adapted to assess specific transformation scenarios. The model is presented in Figure 5. Note that the model lacks specifics. There is no information on what the "transformation initiative" is, what the causal elements are that lead to transformation, or what, precisely, the outcome "transformation" means.

Figure 5. Generic Evaluation Model



Filling in this detail is part of the process of moving from the generic form of the model to a model that can guide a setting-specific transformation. Salient characteristics of the model are:

- While many configurations may lead to transformation (top panel), many others will not (bottom panel). Sequences of change may appear to be trajectories toward transformation, but whether they are truly trajectories towards transformation is an empirical question.
- Many classes of change may affect transformation, e.g., tax incentives, R and D for energy storage, or public advocacy.
- Each class of change effort has discrete components, e.g., tax credits for installing solar panels in commercial establishments, tax deductions for lowered domestic energy use,

and tax write-offs for energy related R and D. These are indicated by the small shapes within each class of events. Those shapes are different across classes to indicate qualitative differences, e.g., there are many different tax incentives, but tax credits for equipment purchases and accelerated equipment write-offs are fundamentally different from, for example, the installed base of service technicians.

- Transformation efforts target specific change. For instance, there may be a belief that tax regulations need to change, but a discrete choice will be made as to what kinds of tax credits. This is why there is a direct connection between "transformation initiative" and discrete elements within each class.
- Networks abound within and across categories. Because of these connections, it is plausible that changing one internal element will affect others within and across the sets. Links in the network can be directed or undirected.
- Transformation is not linked to specifics, because it is collective change that brings about transformational change.
- Multiple scenarios may bring about transformation, but there is no theory that will point to what specific configuration will succeed. The theory depicted in the model is that various configurations of elements will combine in unpredictable ways to effect transformation.

What Role Do Complex Behaviors Play in the General Evaluation Model?

Figure 5 was developed in recognition of the fact that complex behaviors drive transformation. This is not a novel idea in the community of evaluators. For instance, elsewhere (Morell, 2021) I have cited the work of Zazueta (2017), Fisher and Roehrer (2020), and SDG Transformation Forum (2020). In the aggregate, these authors recognize the importance of complexity constructs such as networks, agents, scale, uncertainty about what specific elements are important, and emergent behavior. What is not evident in the existing literature, however, is a grounding of these concepts in the research and theory that is found in complexity science. Such grounding is needed to develop powerful theories of transformation, to construct models to guide the evaluation of transformation, and to explain to concerned parties how transformation can come about, how programs should be evaluated, and what can be drawn from the evaluation data. As noted in Table 1 and Table 2, three constructs are particularly useful: sensitive dependence, emergence, and attractors. The first two have obvious practical value. The third is more speculative, but I believe, important.

Sensitive Dependence. An important part of my argument is that many successful configurations are possible, and that one cannot predict which configuration will be activated. Why is this so? The answer is that success requires numerous influences among networked components and that at each opportunity for action across network edges, small changes can determine the path. Once that determination is made, subsequent choice points are susceptible to the same influence of small perturbations. Another way to put this is that local change can influence the long-term evolution of the system. Local change matters, not just means, variances, and the shapes of distributions.

This is not to say that deliberate choices should not be made with respect to the elements within the model. We may not know how all the elements interact, but we can identify elements that are likely to matter, e.g., tax incentives to lower investment risk, support for businesses that build and maintain infrastructure, or community action to lobby for change. We may not know how they combine, but each can be evaluated without recourse to the measurement of transformation. Each is a discrete change that has specific first-order objectives and thus can be evaluated with common if→then logic, the choice of such logic justified by the decision process discussed with respect to Figure 3 and Figure 4.

Emergence. Figure 5 shows direct connections on the input side, from an effort to bring about transformation, to specific components of transformation. It also shows an ambiguous output from a group of specific changes to transformation. The input logic is our familiar if→then approach. We have a belief about what needs to change. We can devise a plan to bring about that change. The output logic is ambiguous. It implies that there is something unspecifiable about the collective changes that somehow come together to effect transformation. That is a causal model that recognizes both traditional evaluation reasoning and emergent behavior.

What are the implications for evaluation? One implication is that it makes sense to identify critical components. It may not be possible to determine why they are critical, but it does matter whether they are operating. Another implication is that it is reasonable to employ if→then logic on the input side. We do care if a specific action will result in a discrete change. On the output side, however, it is impossible to draw a specific link between constituent parts, relationships among constituent parts, and transformation. The best we can do is to know which parts must be functioning for the emergent phenomenon of transformation to manifest. Finally, there is the matter of research design and data interpretation. On the input side we should insist on designs that will provide as unambiguous a message as possible about discrete act and discrete effect. That is knowledge that decision makers and stakeholders can use.

On the output side, however, we must look for outcome measurement that can be attributable to the consequences of collective change but that is not directly traceable to elements in the change model. As an example, consider the case of "urban vitality." Whatever "urban vitality" is, it is an emergent consequence of some combination of factors such as number and diversity of people, entertainment possibilities, restaurants, cultural attractions, safe and walkable streets, public transportation, intellectual capital among the population, mix of business and nonprofit organizations, population density, and connections among all of these factors, to name but some. How, then, to measure the attraction that a city has? Glaeser (2012) suggests "real income," which is a metric that captures the amount of goods and services that money can buy. The same salary in New York City will buy a lot less than in Detroit. Why, then, are so many people willing to afford less to live in New York? The answer is that food, clothing, rent and entertainment cost more, but the money one does have buys lots of a highly desirable good—"urban vitality." There is no guarantee that such measures can be found for different types of transformation, but it is certainly worth the effort to look.

It is hard to give up on the idea that we cannot trace an outcome to a discrete set of variables. Even when we think in terms of feedback processes, environmental influences, surprise events, and uncertainties, it is hard to escape the belief that there is a specific set of factors that result in specific outcomes that could be identified, if only we could get the data and implement the right methodology. If emergent behavior is operating, we must abandon that belief. That is difficult to do. It is even more difficult to convince funders, planners, and stakeholders that they too must abandon that belief.

Attractors. Figure 2 and Figure 5 portray a world in which multiple scenarios can successfully lead to a specific state of affairs, while many others will not. Is there any reason to believe that attractors like these exist in the social world in which evaluators operate? Yes. One obvious example is "resistance to change." Evaluators are all too well aware of institutions that remain more or less the same no matter how much evaluation data, or information from other sources, suggests that change is desirable. The scenarios may differ with respect to institutions, environments, the change under consideration, and communication methods between stakeholder and evaluator. But the end point would be qualitatively the same; the situation before the evaluation will be more or less the situation after the evaluation. These kinds of attractor states can be immensely powerful. A good example is the "iron law of oligarchy," which describes the preservation of an extractive, corrupt society after a momentous change such as a revolution or independence from a colonial power. The iron law has been shown to manifest itself over a wide variety of settings (Acemoglu & Robinson, 2012). These social attractor phenomena are also present in social system simulations that show that out of a very large number of potential paths through a system, a relatively few (but many more than one) define an attractor that leads from beginning to end (Parunak, 2022).

If we acknowledge the possibility (or even the likelihood) of attractors, what are the implications for evaluation? Those implications are profound. Meaningful questions include: For what we are evaluating, are there multiple paths that lead to success? If so, what must happen (and how much if it needs to happen) before that path leaves the set of success-oriented paths? Thinking in terms of attractors prompts us to reason about the shape of the "space" in which successful paths dwell, and it influences our thinking about how we should analyze and interpret data across different evaluation efforts.

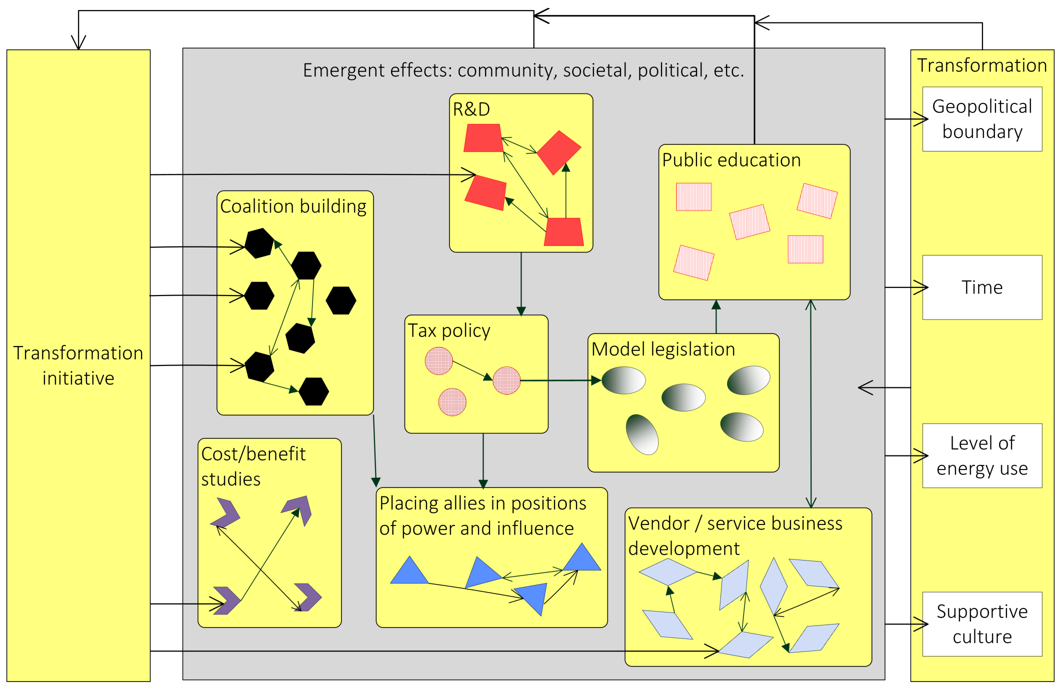
Methodologies to Evaluate Transformation

The model depicted in Figure 5 seems unsatisfying. After all, with all that ambiguity and lack of specificity, how good can the model be to guide methodology and analysis? Further, what does it even mean to collect data that can test the model? In fact, the model is a powerful guide to evaluating transformation because by incorporating complex behavior, it reflects how transformation takes place. Thus, it facilitates the discovery of knowledge that can make a difference to planning. Further, familiar methodology is eminently capable of providing the necessary data. Complex behavior, plebeian methodology.

Choosing Specific Methodologies

Consider how the model in Figure 5 might look once it is fleshed out for a specific transformation evaluation exercise. It might look like Figure 6 (Note how the "variable classes" in Figure 5 are replaced with specifics, e.g., coalition building, tax policy.)

Figure 6. Model to Guide Methodology and Data Interpretation



Although the logic in Figure 5 and Figure 6 is firmly rooted in complex behavior, almost all the methodology needed to test this model is routine and familiar to most of the evaluation community. What would be needed? Archival records of numerous variables, e.g., energy use, labor force makeup, community college offerings, implementation schedules, and so on. Interviews, surveys, and observational methods to capture knowledge such as commitment to whatever transformation is of interest, coalition formation and dissolution, participation in civic organizations, and the like. To observe emergent community-level change, one might monitor social media, track voting patterns, content-analyze real estate ads, or assess demographic change due to people moving in and out of neighborhoods.

Two complexity-related methodologies are missing from the previous paragraph. One is agent-based modeling. The other is formal network analysis. Both of these are important, and both can play an important role in understanding complex change. But neither of these is required. For the most part, familiar methodology will suffice.

I do not mean to imply that methodologies that are unfamiliar to evaluators have no use or should not be sought out. They should (Bamberger & York, 2020). Specific examples include machine learning (Leo et al., 2020), big data and artificial intelligence (Okpe, 2020), geodata (Anand, 2020), and mobile phone data (Anguko, 2020). But my central point remains: Methodologies that are well known to evaluators can go a long way in evaluating transformation.

Of course, it is also important to consider the role of research design. Multiple measures of the same construct are always desirable, as are multiple methodologies. Longer-term data is better than shorter-term data. Long-term data across comparison sites is better still. Multiple case comparisons are enlightening. And so on. (For an in-depth discussion of complexity-based causation, and the importance of case comparisons to understand the characteristics of paths within a "transformation attractor," see Byrne & Uprichard (2012).

Interpreting the Data with Respect to Traditional and Complex Logic

How might traditional and complexity-based reasoning work together? With respect to the model (Figure 6), begin with questions that a traditional approach can address:

- Which elements appeared, and in what order?
- Of all the elements that were deemed consequential, which were implemented?
- Did (and when did) networking relationships among elements develop?
- As the networking structure is revealed, it becomes possible to observe proximate relationships. For instance: Did knowledge from cost-benefit studies play a role in model legislation or coalition formation? Did output of an R and D effort affect the vendor/service community? Did the tax policy changes spur investment?

The next step is to move from these questions to a complexity-based perspective on the relationship between program action and transformation. If we believe the model in Figure 6, then none of the above questions can be thought of as having a unique causal relationship to the outcome of transformation. What we can say is that all of those elements generate an emergent effect that has a causal impact on transformation. Combining the data from traditional evaluation designs with a complex behavior perspective gives us Figure 7. Figure 6 shows the logic of the transformation effort. Figure 7 represents the knowledge that derives from answering questions about that logic.

Figure 7. Evaluation Data Set Within Model

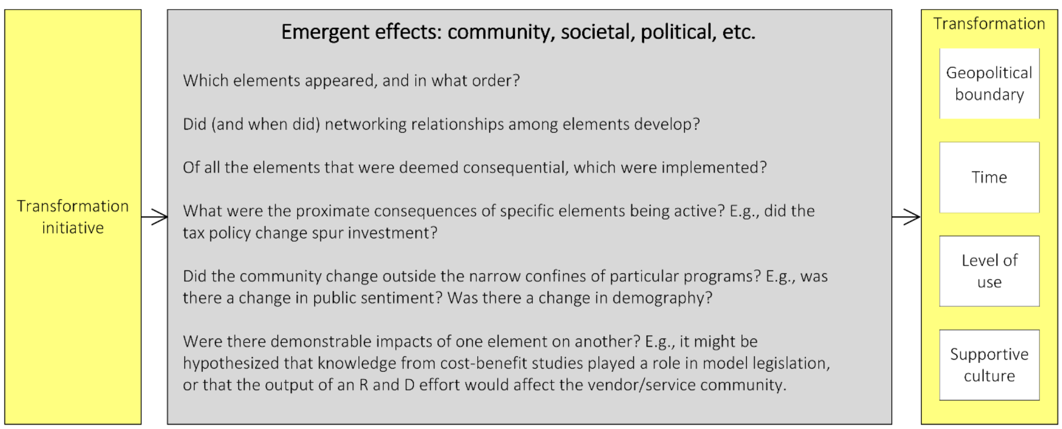


Figure 7 depicts a situation where traditional if→then logic and descriptive data come together to provide knowledge about the success of individual efforts, and about the combined consequences of those efforts for transformation. The model still shows an if→then relationship, but it is not the kind of if→then relationship we are used to. Here the logic is *if an emergent effect occurs, then transformation will take place*. Will an emergent effect be generated? Complexity tells us that many combinations of factors may work, that we may be able to describe that class of successful combinations, and that we can employ traditional logic to each of the constituent parts that we believe are important.

The above echoes Patton's (2019) criteria for transformation. Three of his four transformation principles are:

- <u>Transformation Principle 1: Global–Local Dynamic Interconnection Principle</u>. For evaluation, this means applying a *multilevel connectivity criterion*: Assess global–local interactions and interconnections. This likely will involve documenting contextual variations locally within a coherent global pattern of transformation.
- <u>Transformation Principle 2: Cross-Sector Principle</u>. Integrate and coordinate interventions across sectors and traditional program areas (cutting across silos). Transformational interventions work across sector divisions and program specializations.
- <u>Transformation Principle 3: Multiple Intervention Strategies</u>. Target mixed and multiple types of changes. For evaluators, this

means applying a strategic integration criterion: Track and analyze the interactions and synergies of multiple and diverse interventions and initiatives.

The complexity model depicted in Figure 6 recognizes these principles, but with a twist. To say that global–local connections make a difference is to state a hypothesis about the characteristics of a network that will result in specific emergent effects. It suggests that models at one level of granularity can be understood in terms of lower levels of granularity. That sounds like a reasonable hypothesis (and it is good social policy), but it cannot be assumed correct. Many network patterns are possible. There is no a priori reason to believe that a global–local theory that links network development to transformation is correct.

The second principle calls for integrating and coordinating interventions across sectors. If we believe in evolving networks characterized by sensitive dependence, "integrating" and "coordinating" is not feasible across more than a very limited range. (The coming discussion of Figure 8 will say a lot more about this.) It is because of this limitation that the advice in the third principle is so important, it argues for tracking and analyzing interactions and synergies of multiple and diverse interventions and initiatives (Patton, 2019).

Sustainability

Each of the four components of transformation (geography, time, use, and culture—Table 1) can be seen as a node in a four-node network. It is the nature of that network that makes the state of transformation stable or unstable, or in more familiar evaluation parlance, sustainable or unsustainable. In complexity terms, then, "sustainability" is an emergent property of the network behavior. An important question, then, is what range of values can each node take before the emergent behavior disappears? Or, to draw on another complex behavior, how deep is the attractor in which the emergent outcome resides? Short of historical or cross-site comparisons, it is hard to imagine being able to answer this question in any but an intuitive way. On the other hand, we probably do have some knowledge of similar efforts at transformation, and we do have intimate knowledge of the data and of how the transformation process has fared over time. We also have methodologies to derive predictions about future states and the paths that may lead to those states. It seems worthwhile to bring whatever wisdom and insight we can to do that forecasting as an exercise in discerning the depth of the attractor in which our observed transformation lies.

How Should an Evaluation of a Specific Transformation be Conducted?

Figure 5 is presented as a generic evaluation model. How should this model be applied to an actual evaluation? One dimension of this question is how models and methodologies should be chosen. A second dimension deals with how data can be interpreted in terms of complex behavior.

Models and Methodologies

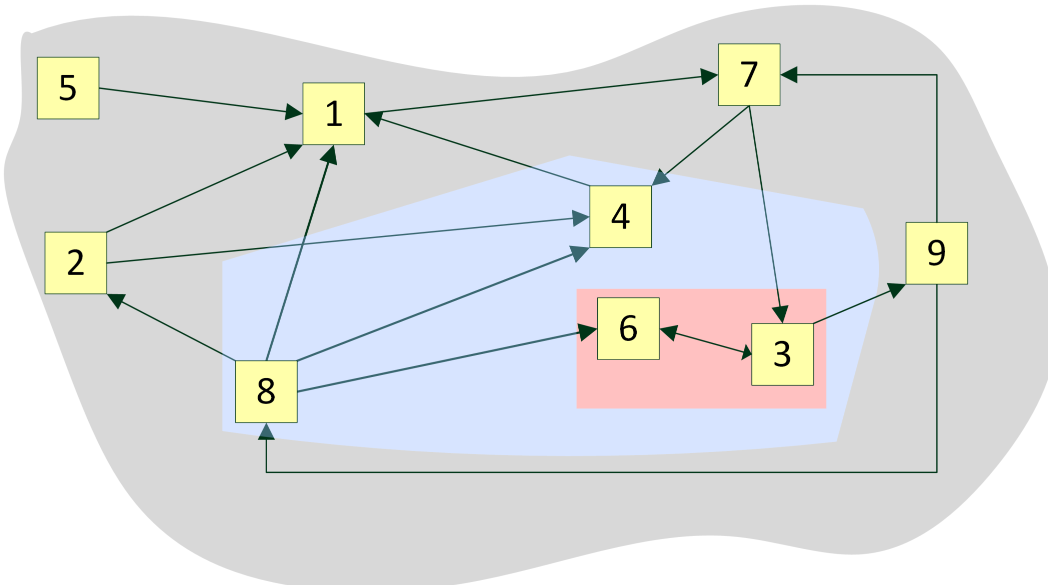
Recast the Generic Model. The generic model depicted in Figure 5 must be recast in a form that applies to the transformation that is being pursued. For instance, the model for transformation to renewable energy would be different from that for transforming health care or education. The specific model should capture both model elements that are the targets of change efforts, and other factors that are known to be important. If there is confidence in relationships among model elements, they too should be depicted. Once the model is developed, the rest of these activities can be pursued.

Construct a Definition of "Transformation." An appropriate multidimensional definition of "transformation" needs to be developed, as per the example in Table 1.

Employ Available Research and Knowledge to Identify Patterns of Change. Theories for a specific transformation and data collection plans require some sense of patterns of change, e.g., likely inflection points, or the possibility of contextual tipping points that would not be detected just by observing change over time. Without attention to patterns of change, data on the relationship between action and outcome cannot be properly interpreted.

Choose Appropriate If→Then Models and Accompanying Contextual Descriptors. If→then logic can prevail in local regions of an entire model, particularly if there is a short time lag between the elements, the connections among them are sparse, and there are few feedback loops. This situation is illustrated in Figure 8. Consider the 6→3 relationship. There is a short time lag between them and only a single direct connection. While even this simple configuration may behave unexpectedly (note the 3 → 9 → 7 →3 relationships), it seems reasonable to claim that if we were only interested in whether change in 6 resulted in a change in 3, a traditional deterministic model would suffice.

Figure 8. Model Range Reasoning



Now consider the blue region. It encompasses four elements (3, 4, 6, and 8). Would traditional logic work for an evaluation of that relationship? That is a tougher call because there is a lot more going on and the time lags are longer. Depending on how much one knew about the specifics, it might be possible to argue either way. In general, decisions about methodology for Figure 8 are rife with judgment calls. What constitutes a short time lag, a sparse model, or "few" elements? How much change has each element demonstrated in the past?

Evaluate Individual Components of the Outcome and Their Networked Relationships. Table 1 provides a generic definition of transformation. Imagine evaluating a transformation effort to increase the literacy rate in a country, as illustrated in Figure 3 and Figure 4. To flesh out that definition, we might define "transformation" as a situation in which (1) *Eighty percent* of the population who reach the age of fourteen in *ten of a country's fifteen administrative units* have a functioning *eighth-grade education*, (2) This level of literacy maintains itself for *four successive generations of children*, and (3) There is *significant social and political support* for maintaining local and national literacy efforts.

Note the measurements, indicated by *italics*, in the definition. All of these are observable and measurable. (It would take some effort to operationalize "social and political support" in both qualitative and quantitative terms, but it could be done.) Changes in each of these could be observed over time. It would also be possible to observe developing network relationships among these elements. For instance, local civic action may play a role in keeping students in school, i.e., social support affects the number of students being educated.

Interpreting Data with Respect to Complexity

Three examples illustrate how a complexity lens can be applied to the data.

Example 1: Change Within the Model. In this example, an appreciation of change requires consideration of network relationships, sensitive dependence implications for model scope, and emergence. Imagine a scenario in which evaluation shows that multiple discrete innovations are achieving their immediate objectives and that some elements of transformation seem to be developing. How might one interpret that data? First, we know that network relationships among local changes matter. Thus, in addition to knowing about the immediate impact of each program, it is also important to determine whether network effects among those changes are developing. If they are not, it would be difficult to make the case that interventions are affecting transformation. In essence our model says that developing networks are an intermediate step between program action and transformation. We may not be able to show that the networks caused transformation, but we can be confident that if the networks are absent, all the first-order outcomes we are demonstrating are probably not on the causal path to transformation.

Second, to make the case, we need evaluation that will show that networks are developing, *not* that *particular* networks are developing. This is because multiple paths through the system may bring about transformation, but because of sensitive dependence, observed causal paths cannot be expected to repeat. What we might be able to do is to use the data to detect commonalities, to plug that data into simulation software, and to try to gain a bit of insight on the characteristics of social attractors that lead to transformation. If we could compare patterns across cases, so much the better.

Third, we need to exercise restraint in drawing conclusions based on our observations of network behavior in small regions of the model. Caution is needed because as knowledge about connections develops, we may be tempted to draw conclusions about connections across larger and larger regions. Those are dangerous conclusions because local perturbations can affect relationships in the model as a whole.

Fourth, at some point we need to give up the comforting belief that whatever specifics we are measuring can be identified as causal factors in transformation. What we can do, however is to try to determine what collections of factors are required in order for the emergent construct to form. We can also search for measures that are not related to any elements in our model, but which do capture the emergent consequences of those elements. For instance, consider the “urban vitality” example presented earlier (in the subsection Emergence in the section What Role do Complex Behaviors Play in the General Model?). There, real income was proposed as a measure of the emergent construct that derived from a host of characteristics of urban life.

Example 2: Transformation as an Outcome. In this example, an appreciation of change requires consideration of network relationships, sensitive dependence, uncertainty as a function of model scope, and emergence. Outcome models can be complicated because they can be characterized by nonlinear patterns and contextual tipping points that are not measured by change in the outcomes of interest. This has implications for various elements in a model of outcome change.

For instance, the importance we attribute to contextual tipping points may affect our beliefs about the change elements we put into the outcome model. As an example, consider measurement of conversion from gasoline-powered to electric-powered vehicles. As of this writing the percentage of electric vehicle use is miniscule. But manufacturers’ commitments to those vehicles, and the fact that many countries have mandated their use, may be tipping points on the path to transformation. An outcome model that recognized the possibility of such changes, and hence included them as an element of transformation outcome, might not look like a model that ignored that possibility. Different models shape methodology and data interpretation.

Or consider the challenge of time horizons for prediction and our theory as to when an inflection point might take place. If we believed that a tipping point will occur early in the change process, we would have early confidence in any statements we might make about whether our efforts were succeeding to bring about transformation.

Example 3: A Priori and a Posteriori Reasoning. Once change has been observed, the specific if→then reasoning in the model can be discovered. (Assuming, of course, that one has been astute enough to know where the data are and has implemented a methodology capable of exploiting that data.) Or, put differently, history will tell us what path was followed. While the implications of sensitive dependence are that the discovered path may not repeat, wisdom can be gained, and judgment sharpened, by understanding what happened, and why. Further, such analysis across several transformations may provide a sense of the attractor that leads to successful outcomes.

Conclusion

This article has been an effort to bridge two approaches to evaluating transformation. One entails a prescriptive stance that tries to predict causal relationships. The traditional approach uses the kind of if→then reasoning that has always been present in evaluation. The second approach recognizes that three complex behaviors—emergence, sensitive dependence, and attractors—imply a predictive stance that does not recognize relationships between specific actions and outcomes but does accept that a collection of activities can lead to specific outcomes.

The importance of complexity is well recognized in the community of evaluators who are involved in the evaluation of transformation. This article builds on those efforts by taking a deep dive into how theory and research in complexity science can inform our work. It argues that a more rigorous draw on that body of knowledge would allow us to devise a definition of "transformation" that captures the nonlinear and multidimensional nature of transformation. It also argues that much evaluation of transformation can be accomplished by predictive if→then designs, but as the scope of inquiry expands, the logic of complexity must be brought to bear. Finally, this article argues that evaluating complexity can be done with familiar methodologies that are common knowledge in the evaluation community. Complex behavior. Plebeian methodology.

References

Acemoglu, D., & Robinson, J. A. (2012). *Why nations fail*. Currency / Random House. https://www.penguinrandomhouse.com/books/205014/why-nations-fail-by-daron-acemoglu-and-james-a-robinson/

Anand, A. (2020, second quarter). Integrating big geodata and technology in evaluation: What do we need to know? *Evaluation Matters Magazine*. https://idev.afdb.org/en/document/preparing-evaluation-future-big-data-modern-technologies-and-shifts-global-development

Anguko, A. (2020). Application of SurveyCTO mobile data collection technology in household surveys: The case of an impact evaluation of the Community Based Integrated Water Resource Management Project in Niger *Evaluation Matters Magazine*. https://idev.afdb.org/en/document/preparing-evaluation-future-big-data-modern-technologies-and-shifts-global-development

Bamberger, M., & York, P. (2020, second quarter). Transforming evaluation in the 4th industrial revolution: Exciting opportunities and new challenges. *Evaluation Matters Magazine*. https://idev.afdb.org/en/document/preparing-evaluation-future-big-data-modern-technologies-and-shifts-global-development

Bar-Yam, Y. (2021). *Concept map*. New England Complex Systems Institute. https://necsi.edu/concept-map

Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*(5439), 509. https://doi.org/10.1126/science.286.5439.509

Barbrook-Johnson, P., Proctor, A., Giorgi, S., & Phillipson, J. (2020). How do policy evaluators understand complexity? *Evaluation*, *26*(3), 315–332. https://doi.org/10.1177/1356389020930053

Batra, G., Garcia, J., & Temnenko, K. (2022). Transformational change for achieving scale: Lessons for a greener recovery. In J. I. Uitto & G. Battra (Eds.), *Transformational change for people and the planet: Evaluating environment and development* (pp. 27–38). Springer.

Bicket, M., Christie, I., Gilbert, N., Hills, D., Penn, A., & Wilkinson, H. (2020). *Supplementary guide: Handling complexity in policy evaluation* (HM Treasury Ed.). CECAN—the Centre for Evaluating Complexity Across the Nexus. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/879437/Magenta_Book_supplementary_guide._Handling_Complexity_in_policy_evaluation.pdf

Box, G. E. P. (1979). Robustness in the strategy of scientific model building. In R. L. Launer & G. N. Wilkinson (Eds.), *Robustness in Statistics* (pp. 201–236). Academic Press.

Byrne, D., & Uprichard, E. (2012). Useful complex causality. In H. Kinkard (Ed.), *Oxford handbook of philosophy of social science* (pp. 109–129). Oxford University Press. https://doi.org/https://doi.org/10.1093/oxfordhb/9780195392753.013.0006

Castellani, B., & Gerrits, L. (2021). *2021 map of the complexity sciences*. Art & Science Factory. https://www.art-sciencefactory.com/complexity-map_feb09.html

Comfort, L. K. (1994). Self-Organization in complex systems. *Journal of Public Administration Research and Theory*, *4*(3), 393–410. http://www.jstor.org/stable/1181895

Egré, P., & Rott, H. (2021). The logic of conditionals. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Stanford University.

Feinstein, O. (2019). Dynamic evaluation for transformational change. In R. D. van den Berg, C. Magro, & S. S. Mulder (Eds.), *Evaluation for transformational change: Opportunities and challenges for the Sustainable Development Goals* (pp. 17–32). International Development Evaluation Association (IDEAS). https://ideas-global.org/wp-content/uploads/2019/11/2019-11-05-

Final_IDEAS_EvaluationForTransformational Change.pdf

Firmansyah, R., Putri, D., Wicaksono, M., Putri, S., Widianto, A., & Palil, M. (2021). Educational transformation: An evaluation of online learning due To COVID-19. *International Journal of Emerging Technologies in Learning (iJET)*, *16*(7), 61–76.

Fisher, S., & Roehrer, C. (2020). *Understanding progress towards transformation, in time*. EIT Climate-KIC.

Frigg, R., & Hartmann, S. (2018). Models in Science. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Stanford University. https://plato.stanford.edu/archives/sum2018/entries/models-science/

Gates, E. F., Walton, M., Vidueira, P., & McNall, M. (2021). Introducing systems- and complexity-informed evaluation. *New Directions for Evaluation*, *170*, 13–25. https://doi.org/https://doi.org/10.1002/ev.20466

Glaeser, E. (2012). *Triumph of the city: How our greatest invention makes us richer, smarter, greener, healthier, and happier*. Penguin Group.

Gordon, T. J., & Glenn, J. C. (2009). Environmental scanning. In T. J. Gordon & J. C. Glenn (Eds.), *Futures research methodology — version 3.0*. The Millennium Project. https://www.researchgate.net/profile/Jerome-Glenn/publication/286279000_Environmental_Scanning/links/5667954a08aea62726ee9541/Environmental-Scanning.pdf

Greenhalgh, T., Humphrey, C., Hughes, J., Macfarlane, F., Butler, C., & Pawson, R. (2009). How do you modernize a health service? A realist evaluation of whole-scale transformation in London. *The Milbank Quarterly*, *87*(2), 391–416. https://doi.org/https://doi.org/10.1111/j.1468-0009.2009.00562.x

Greenhalgh, T., Robert, G., Macfarlane, F., Bate, P., & Kyriakidou, O. (2004). Diffusion of innovations in service organizations: Systematic review and recommendations. *The Milbank Quarterly*, *82*(4), 581–629. https://doi.org/10.1111/j.0887-378X.2004.00325.x

Ibrahim, M., El-Zaart, A., & Adams, C. (2017, September 12–14). *Theory of change for the transformation towards smart sustainable cities* [Paper presentation]. Sensors Networks Smart and Emerging Technologies (SENSET). Beiriut, Lebanon

Initiative for Climate Action Transparency (ICAT). (2020). *Transformational change methodology: Assessing the transformational impacts of policies and actions*. https://www.iges.or.jp/en/pub/transformational-change-methodology/en

International Evaluation Group. (2016). *Supporting transformational change for poverty reduction and shared prosperity*. World Bank. https://ieg.worldbankgroup.org/evaluations/supporting-transformational-change-poverty-reduction-and-shared-prosperity

Jeong, H., Néda, Z., & Barabási, A. L. (2003). Measuring preferential attachment in evolving networks. *Europhysics Letters (EPL)*, *61*(4), 567–572. https://doi.org/10.1209/epl/i2003-00166-9

Lamberson, P. J., & Page, S. E. (2012). *Tipping points* (Paper #12-02-002). Santa Fe Institute. https://www.santafe.edu/research/results/working-papers/tipping-points

Leo, B., Pattni, S., Winn, C., Lewis, Q., Paton, C., & Fraym, M. P. (2020, second quarter). Using machine learning for climate related impact evaluations. *Evaluation Matters Magazine*. https://idev.afdb.org/en/document/preparing-evaluation-future-big-data-modern-technologies-and-shifts-global-development

Matz, R. L., Fata-Hartley, C. L., Posey, L. A., Laverty, J. T., Underwood, S. M., Carmel, J. H., Herrington, D. G., Stowe, R. L., Caballero, M. D., Ebert-May, D., & Cooper, M. M. (2018). Evaluating the extent of a large-scale transformation in gateway science courses. *Science Advances, 4*(10), eaau0554. https://doi.org/doi:10.1126/sciadv.aau0554

Morell, J. A. (2021). A complexity-based metatheory of action for transformation to a green energy future. In R. D. Van den Berg, C. Maro, & M.-H. Adrien (Eds.), *Transformational evaluation for the global crises of our times* (pp. 345–370). IDEAS. https://ideas-global.org/ideas-book-transformational-evaluation-for-the-global-crises-of-our-times/

Ofir, Z., & Rugg, D. (2021). International developments in evaluation: Transforming evaluation [special section]. *American Journal of Evaluation*, *42*(1), 47–138.

Okpe, B. (2020, second quarter). Integrating big data analytics and artificial intelligence into monitoring and evaluation in a fast-changing development landscape. *Evaluation Matters Magazine*. https://idev.afdb.org/en/document/preparing

-evaluation-future-big-data-modern-technologies-and-shifts-global-development

Parunak, H. V. D. (2023). Analyzing the Behavioral Attractors of a Social Simulation. In P. A. Research, Model Mechanisms and Behavioral Attractors. Social Simulation Conference (SSC2022), Milan. https://www.abcresearch.org/abc/papers/SSC2022Attractors.pdf.

Patton, M. Q. (2019). Transformation to global sustainability: Implications for evaluation and evaluators. *New Directions for Evaluation*, *162*, 103–117. https://doi.org/10.1002/ev.20362

Patton, M. Q. (2020). *Blue marble evaluation: Premises and principles*. Guilford Press.

Pearl, J. (2009). *Causality: Models, reasoning and inference* (2nd ed.). Cambridge University Press.

Phelan, S. E. (2001). What is complexity science, really? *Emergence*, *3*(1), 120–136. https://doi.org/https://doi.org/10.1207/S15327000EM0301_08

Picciotto, R. (2021). Toward a complexity framework for transformative evaluation. *Journal.of MultiDisciplinary Evaluation*, *16*(35), 54–76.

Rogers, E. (2003). *Diffusion of innovations* (5th ed.). Free Press.

Rogers, K. (2012). Scientific modeling. *Encyclopædia Britannica*. https://www.britannica.com/science/scientific-modeling

Santa Fe Institute. (n.d.-a). *Attractor*. Complexity Explorer. https://www.complexityexplorer.org/explore/glossary/6-attractor

Santa Fe Institute. (n.d.-b). *Emergence*. Complexity Explorer. https://www.complexityexplorer.org/explore/glossary/414-emergence

Santa Fe Institute. (n.d.-c). *Sensitive dependence*. Complexity Explorer. https://www.complexityexplorer.org/explore/glossary/44-sensitive-dependence-on-initial-conditions

SDG Transformation Forum. (2020). *Our theory of transformational change*. SDG Transformation Forum.

Statistical model. (2023, June 30). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Statistical_model&oldid=1161602898

Systems Innovation. (2020a). Attractor. YouTube. https://www.youtube.com/watch?v=C9Xl6BNu-kM

Systems Innovation. (2020b). *Social attractors & chaos* [Video]. YouTube. https://www.youtube.com/watch?v=C9Xl6BNu-kM

Theraulaz, G., & Bonabeau, E. (1999). A Brief history of stigmergy. *Artificial Life*, *5*, 97–116.

U.S. Department of Energy. (2012). *DOE handbook: Accident and operational safety analysis: Vol. 1. Accident analysis techniques* (DOE-HDBK-1208-2012). https://www.standards.doe.gov/standards-documents/1200/1208-bhdbk-2012-v1/@@images/file

Van den Berg, R. D., Maro, C., & Adrien, M.-H. (Eds.). (2021). *Transformational evaluation for the global crises of our times*. International Development Evaluation Association (IDEA). https://ideas-global.org/wp-content/uploads/2021/07/2021-IDEAS-book-Transformational-Evaluation.pdf.

Wadeson, A., Monzani, B., & Aston, T. (2020). *Process tracing as a practical evaluation method: Comparative learning from six evaluations*. https://mande.co.uk/wp-content/uploads/2020/03/Process-Tracing-as-a-Practical-Evaluation-Method_23March-Final.pdf

West, G. (2017). *Scale: The universal laws of growth, innovation, sustainability, and the pace of life in organisms, cities, economies, and companies*. Penguin.

Zazueta, A. (2017). *Principles for the development of integrated transformational projects in climate change and chemicals & waste*. Global Environment Facility. http://www.stapgef.org/sites/default/files/publications/Final_Principles%20for%20%20Development%20of%20Integrated%20CCM%20and%20ChW%20Projects_October%202017.pdf

Zazueta, A. (2019). *Using theories of change to steer transformational development* [Paper presentation]. IDEAS Global Assembly / 3rd International Conference on Evaluating Environment and Development, Prague, Czech Republic.

Zellner, M. L., Lyons, L. B., Hoch, C. J., Weizeorick, J., Kunda, C., & Milz, D. C. (2012). Modeling, learning, and planning together: An application of participatory agent-based modeling to environmental planning. *URISA Journal*, *24*(1), 77–92. https://www.academia.edu/download/47465368/Zellner_2012.pdf