

# Chapter 3

## Carefully articulate the theories linking interventions to outcomes

---

When evaluators talk about the black box “problem,” they are usually referring to the practice of viewing interventions primarily in terms of effects, with little attention paid to how and why those effects are produced. The common thread underlying the various versions of theory-based evaluation is the argument that “interventions are theories incarnate” and evaluation constitutes a test of intervention theory or theories.

### 3.1. Seeing interventions as theories: The black box and the contribution problem

Interventions are embodiments of theories in at least two ways. First, they comprise an expectation that the introduction of a program or policy intervention will help ameliorate a recurring social problem. Second, they involve an assumption or set of assumptions about how and why program activities and resources will bring about changes for the better. The underlying theory of a program often remains hidden, typically in the minds of policy architects and staff. Policies—be they relatively small-scale direct interventions like information campaigns, training programs, or subsidization; meso-level interventions such as public-private partnerships and social funds, or macro-level interventions such as “general budget support”—rest on social, behavioral and institutional assumptions indicating why “this” policy intervention will work, which at first view are difficult to uncover.

By seeing interventions as theories and by using insights from theory-based evaluations, it is

possible to open up the *black box*. Development policies and interventions, in one way or another, have to do with changing behavior/intentions/knowledge of households, individuals, and organizations (grass roots, private, and public sector). Crucial for understanding what can change behavior is information on *behavioral and social mechanisms*. An important insight from theory-based evaluations is that policy interventions are (often) believed to address and trigger certain social and behavioral responses among people and organizations; in reality this may not be the case.

### 3.2. Articulating intervention theories on impact

Program theory (or intervention theory) can be identified (articulated) and expressed in many ways—a graphic display of boxes and arrows, a table, a narrative description, and so on. The methodology for constructing intervention theory, as well as the level of detail and complexity, also varies significantly (e.g., Connell et al., 1995; Leeuw, 2003; Lipsey, 1993; McClintock, 1990; Rogers et al., 2000; Trochim, 1989; Wholey, 1987).

Too often the role of methodology is neglected, and it is assumed that “intervention theories” are like *manna* falling out of the sky. That is not the case. Often the underlying theory has to be dug up. Moreover, much of what passes as theory-based evaluation today is simply a form of “analytic evaluation [which] involves no theory in anything like a proper use of that term” (Scriven, 1998: 59).

The intervention theory provides an overall framework for making sense of potential processes of change induced by an intervention. Several pieces of evidence can be used for articulating the intervention theory:

- An intervention’s existing logical framework as a starting point for mapping causal assumptions linked to objectives and other written documents produced within the framework of an intervention
- Insights provided by and expectations harbored by policy makers and staff (and other stakeholders) on how they think the intervention will affect/is affecting/has affected target groups
- (Written) evidence on past experiences of similar interventions (including those implemented by other organizations)
- Literature on mechanisms and processes of change in certain institutional contexts, for particular social problems, in specific sectors, etc.

Sometimes stakeholders have contrasting assumptions and expectations about an intervention’s impact that has implications for reconstructing the intervention theory. Basically, there are two ways to address this issue. The first is to try to combine the perspectives of different people (for example, program managers and target group members) into an overarching intervention theory that consists of (parts of) arguments from these different sources. The overall theory might be created through an iterative process of dialogue and refinement and as such might contribute to a shared vision among stakeholders (see, e.g., Pawson and Tilley, 1997). Second, when differences are substantial, several competing intervention theories have to be reconstructed. Carvalho and White (2004) give an example of a “theory” and an “anti-theory” dealing with the assumed impact of social funds (see box 3.1).

For an example of what an impact theory might look like, consider the case of a small business development project that provides training to young managers who have started a business. The direct goal is to help make small businesses financially sustainable and the indirect goal is to generate more employment in the region. Closer scrutiny reveals that the project might have a positive influence on the viability of small businesses in two ways: First, by training young

### Box 3.1: Social funds and government capacity: Competing theories

Proponents of social funds argue they will develop government capacity in several ways. Principle among these are that the social fund will develop superior means of resource allocation and monitoring, which will be transferred to the government either directly through collaborative work or indirectly by copying the procedures shown to be successful by the social fund. But critics argue that social funds bypass normal government channels and so undermine government capacity, an effect reinforced by drawing away the government’s best people by paying a project premium. Hence, these are rather different theories of how

social funds affect government capacity. Carvalho and White (2004) refer to both sets of assumptions in terms of “theory” and “anti-theory.” Their study found that well-functioning, decentralized social funds, such as the Zambia Social Investment Fund in Zambia, worked through—rather than parallel to—existing structures and that the social fund procedures were indeed adopted more generally by district staff. But at national level there was generally little evidence of either positive or negative effects on capacity—with some exceptions, such as the promotion of poverty mapping in some countries.

Source: Carvalho and White (2004).

people in basic management and accounting skills, the project intends to have a positive effect on financial viability and ultimately on the growth and sustainability of the business; second, by supporting the writing of a business plan, the project aims to increase the number of successful applications for credit with the local bank, which previously excluded the project’s target group because of the small loan sizes (high transaction costs) and high risks involved. Following this second causal strand, efficient and effective spending of the loan is also expected to contribute to the strength of the business. Outputs are measured in terms of the number of people trained by the project and the number of loans the bank extends (see figure 3.1.).

Any further empirical analysis of the impact of the project requires insight into the different factors—besides the project itself—that affect small business development and employment generation. Even in this rather simple example, the number of external variables that affect the impact variables either directly or by moderating the causal relations specified in figure 3.1. is manifold. Some examples are the following:

- Short-term demands on the labor efforts of business owners in other activities may lead to suboptimal strategic choices, jeopardizing the sustainability of the business.
- Inefficient or ineffective use of loans because of short-term demands for cash for other expenditures might jeopardize repayment and the financial viability of the business.

- Deteriorating market conditions (in input or output markets) may jeopardize the future of the business.
- The availability and quality of infrastructure or skilled labor at any point may become constraining factors on business development prospects.
- The efforts of other institutions promoting small business development or any particular aspect of it might positively (or negatively) affect businesses.

Methods for reconstructing the underlying assumptions of project/program/policy theories are the following (see Leeuw, 2003):

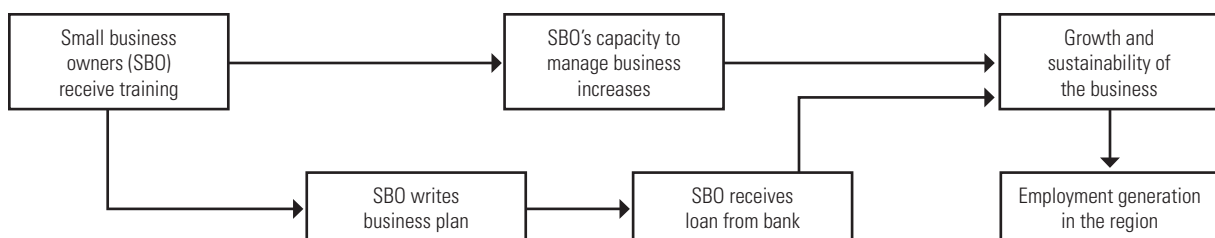
- A policy-scientific method, which focuses on interviews, documents, and argumentation analysis
- A strategic assessment method, which focuses on group dynamics and dialogue
- An elicitation method, which focuses on cognitive and organizational psychology.

Central in all three approaches is the search for mechanisms that are believed to be “at work” when a policy is implemented. Box 3.2 discusses social and behavioral mechanisms for understanding impact.

### 3.3. Testing intervention theories on impact

After articulating the assumptions on how an intervention is expected to affect outcomes and impacts, the question arises as to what extent these assumptions are valid. In practice,

**Figure 3.1: Basic intervention theory of a fictitious small business support project**



### Box 3.2: Social and behavioral mechanisms as heuristics for understanding processes of change and impact

Hedström (2005: 25) has defined the concept of social mechanisms as “a constellation of entities and activities that are organized such that they regularly bring about a particular type of outcome.” Mechanisms form the “nuts and bolts” (Elster, 1989) or the “engines” (Leeuw, 2003) of interventions (policies and programs), making them work, given certain contexts (Pawson and Tilley, 1997). Hedström and Swedberg (1998: 296–98), building on the work of Coleman (1990), discuss three types of mechanisms: situational mechanisms, action formation mechanisms, and transformational mechanisms.

Examples of *situational mechanisms* are self-fulfilling and self-denying prophecies and crowding-out (e.g., by striving to force people who are already largely compliant with laws and regulations into full compliance, the opposite is realized, because due to the extra focus on laws and regulation, the internal motivation of people to comply is reduced).

*Action-formation mechanisms* are the heuristics that people develop to deal with their bounded rationality, such as—

- Framing and the endowment effect—“The fact that people often demand much more to give up an object than they would be willing to pay to acquire it,” but also the tendency for people to have a stronger preference for more immediate payoffs than for later payoffs, the closer to the present both payoffs are
- Types of learning (social learning, vicarious learning)
- “Game-theoretical” mechanisms, such as the “grim strategy” (to repeatedly refuse to cooperate with another party as a punishment for the other party’s failure to cooperate previously) and the shadow of the future /shadow of the past mechanism
- Mechanisms such as the “fight-or-flight-response” to stress and the “tend-and-befriend mechanism” are other examples.

*Transformational mechanisms* illuminate how processes and results of interacting individuals and groups are “transformed” into collective outcomes. Examples are the following:

- Cascading is a process by which people influence one another, so much so that participants ignore their private knowledge and rely instead on the publicly stated judgments of others. The bandwagon phenomenon (the tendency to do [or believe] things because many other people do [or believe] the same)

is related to this, as are group think, the common knowledge effect, and herd behavior.

- “Tipping points,” “where a small additional effort can have a disproportionately large effect, can be created through virtuous circles, or be a result of achieving certain critical levels” (Rogers, 2008: 35).

#### **Relevance of mechanisms for impact evaluations**

Development policies and interventions, in one way or another, have to do with changing behavior/intentions/knowledge of households, individuals, and organizations (grass roots, private, and public sector). Crucial for understanding what can change behavior is information about these mechanisms. *The mechanisms underlying processes of change might not be necessarily those that are assumed to be at work by policy makers, programs designers, and staff.* Creating awareness on the basis of (public) information campaigns does not always lead to behavioral change. Subsidies and other financial incentives run the risk of causing unintended side effects, such as benefit snatching, but also create the “Mitnahme-effect” (people already tended to behave in a way the incentive wanted them to behave before the incentive existed). Mentoring dropouts in education might cause “learned helplessness” and therefore increase dropout rates. Many other examples are available in the literature. The relevance of knowing which social and behavioral mechanisms are believed to do the work increases as the complication and complexity of interventions increases.

A focus on mechanisms helps evaluators and managers open up and test the theory underlying an intervention. Spending time and money on programs based on “pet theories” of policy makers or implementation agents that are not corroborated by relevant research should probably not be high on the agenda. If a policy intervention is based on mechanisms that are known not to work (in a given context or in general), that is a signal that the intervention probably will not be very effective. This can be found out on the basis of desk research as a first test of the relevance and validity of an intervention theory, that is, by confronting the theory with existing knowledge about mechanisms. That knowledge stems from synthesis and review studies (see chapter 6). Further empirical impact evaluation can generate more contextualized and precise tests of the intervention theory.

evaluators have at their disposal a wide range of methods and techniques to test the intervention theory. We can distinguish between two broad approaches. The first is that the theory constitutes the basis for constructing a “causal story” about how and to what extent the intervention has produced results. Usually different methods and sources of evidence are used to further refine the theory in an iterative manner until a credible and reliable causal story has been generated. The second approach is to use the theory as an explicit benchmark for testing (some of) the assumptions in a formal manner. Besides providing a benchmark, the theory provides the template for method choice, variable selection, and other data collection and analysis issues. This approach is typically applied in statistical analysis but is not in any way restricted to this type of method. In short, theory-based methodological designs can be situated anywhere in between “telling the causal story” and “formally testing causal assumptions.”

The systematic development and corroboration of the causal story can be achieved through *causal contribution analysis* (Mayne, 2001), which aims to demonstrate whether the evaluated intervention is one of the causes of observed change. Contribution analysis relies on chains of logical arguments that are verified through careful analysis. Rigor in causal contribution analysis involves systematically identifying and investigating alternative explanations for observed impacts. This includes being able to rule out implementation failure as an explanation for lack of results and developing testable hypotheses and predictions to identify the conditions under which interventions contribute to specific impacts.

The causal story is inferred from the following evidence:

- There is a reasoned theory of change for the intervention: it makes sense, is plausible, and is agreed to by key players.
- The activities of the intervention were implemented.
- The theory of change—or key elements thereof—is verified by evidence: the chain of expected results occurred.
- Other influencing factors have been assessed and either shown not to have made a significant contribution or their relative role in contributing to the desired result has been recognized.

The analysis is best done iteratively, building up a more robust assessment of causal contribution. The overall aim is to reduce the uncertainty about the contribution the intervention is making to the observed results through an increased understanding of why the observed results have occurred (or not) and the roles played by the intervention and other factors. At the impact level this is the most challenging, and a “contribution story” has to be developed for each major strategy that is part of an intervention, at different levels of analysis. They would be linked, as each would treat the other strategies as influencing factors.

One of the key challenges in the foregoing analysis is to pinpoint the exact causal effect from intervention to its impact. Despite the potential strength of the causal argumentation on the links between the intervention and impact, and despite the possible availability of data on indicators, as well as data on contributing factors, etc., there remains uncertainty about the *magnitude* of the impact as well as *the extent* to which the changes in impact variables are really due to the intervention or to other influential variables. This is called the attribution problem and is discussed in chapter 4.

### Key message

Carefully articulate the assumptions behind the theories linking interventions to outcomes. What are the causal pathways linking intervention outputs to processes of change and impact? Be critical if an “intervention theory” appears to assert or assume changes without much explanation. The focus should be on dissecting the causal (social, behavioral, and institutional) mechanisms that make interventions “work.”