Purposeful Program Theory

Effective Use of Theories of Change and Logic Models

SUE C. FUNNELL AND

PATRICIA J. ROGERS

JOSSEY-BASS
A Wiley Imprint
www.josseybass.com

CONTENTS

Tables

Exhibits

ACKNOWLEDGMENTS

MANY PEOPLE HAVE helped us along the journey to this book. Our work builds on the contributions of the pioneers and innovators in program theory evaluation and those whose work underpins key concepts in program theory. The work of Dan Stufflebeam, Carol Weiss, Joseph Wholey, Claude Bennett, Len Bickman, Carol Fitz-Gibbon, and Lynn Morris laid the foundations for program theory. Benjamin Bloom, Thomas Hastings, and George Madaus's taxonomy of educational objectives, when linked with Bennett's Hierarchy, gave rise to the notions of outcome chains, and Tom Hastings's important article, "Curriculum Evaluation: The Why of the *Outcomes*," expanded thinking beyond a focus on outcomes. Bryan Lenne of the New South Wales Program Evaluation Unit contributed significantly to the development of the matrix approach to program theory and the classification of program archetypes. More recently, we have learned much from the work of Michael Patton, Rick Davies, and Boru Douthwaite exploring the use of network theory and systems approaches; the development of realist evaluation and realist synthesis by Ray Pawson and Nick Tilley; and outcome mapping by Barry Kibel, Terry Smutlyo, Fred Carden, and Sarah Earl in terms of using program theory for complicated and complex situations and interventions.

Over more than twenty-five years, we have had the good fortune to work with colleagues and clients who have stretched our thinking about program theory. Sue particularly acknowledges those whose work and insights have contributed to the development of the matrix approach: Carolyn Wells for development of the concept of attributes of success as part of program theory matrices; Steve Baxter for his contribution to thinking about managing external risk factors that affect outcomes and the implications of that for developing a program theory; and Larraine Larri for her assistance with portraying and explaining program theory matrices. Sue also acknowledges the contributions of those who helped identify program theories for particular program

legal systems; natural resources management; occupational health and safety; primary industries; and roads. Funnell, a past president of the Australasian Evaluation Society (AES) and now an AES Fellow, has been awarded the AES Evaluation Training and Services Award for outstanding contributions to the profession of evaluation.

PATRICIA J. ROGERS is Professor of Public Sector Evaluation at the Royal Melbourne Institute of Technology, Melbourne, Australia. She has worked in public sector evaluation and research for more than twenty-five years, with government and nongovernment organizations (international, national, state, and local) across a wide range of program areas, including agriculture, community development, criminal justice, early childhood education, health promotion, Indigenous housing, international development, and legal aid. She has worked on projects with the United Nations Development Programme, World Bank Institute, Network of Networks on Impact Evaluation, U.S. Department of Energy, the Office of the Presidency (South Africa), the Public Service Commission (South Africa), and the Development Bank of Southern Africa. She has presented keynote addresses at conferences of the Australasian, Aotearoa/New Zealand, European, United Kingdom, South African, and Swedish evaluation societies and associations and is on the editorial boards of the journals *Evaluation* and *New Directions for Evaluation.* She has been awarded the American Evaluation Association's Myrdal Award for Evaluation Practice, the Australasian Evaluation Society's Evaluation Training and Services Award for outstanding contributions to the profession of evaluation, the AES Caulley-Tulloch Prize for Pioneering Literature in Evaluation, and (with Sue Funnell) the AES Best Evaluation Study Award.

THE 1920S ENTREPRENEUR Carl Weeks once wrote, "If you can dream it, you can build it." This is the key idea that underpins program theory. Having a vision of where we are going and some clarity about how we plan to get there can help us work together to achieve our goals, and learn from both success and failure.

WHAT PROGRAM THEORY IS

A program theory is an explicit theory or model of how an intervention, such as a project, a program, a strategy, an initiative, or a policy, contributes to a chain of intermediate results and finally to the intended or observed outcomes. A program theory ideally has two components: a theory of change and a theory of action. The theory of change is about the central processes or drivers by which change comes about for individuals, groups, or communities—for example, psychological processes, social processes, physical processes, and economic processes. The theory of change could derive from a formal, research-based theory or an unstated, tacit understanding about how things work. For example, the theory of change underpinning some health promotion programs is that changes in perceived social norms lead to behavior changes. The theory of action explains how programs or other interventions are constructed to activate these theories of change. For example, health promotion programs might use peer mentors, advertisements with survey results, or some other strategy to change perceptions of social norms.

Program theory, under all its various labels, including "theories of change," "logic modeling," and "intervention logic," has grown in popularity over the past twenty years or so. Many government and nongovernment

organizations across the world now encourage or require its use for planning, monitoring, and evaluating.

When done well, program theory can produce many benefits. It can develop agreement among diverse stakeholders about what they are trying to do and how, or identify where there are legitimately different perspectives. It can help to improve plans by highlighting gaps and opportunities for collaboration with partners. It can help to set realistic objectives. It can support the development of meaningful performance indicators to track progress and report achievements. It can be used to identify where and why unsuccessful programs are failing or what makes successful programs work, and how they might be reproduced or adapted elsewhere. It can provide a framework to bring together information from many sites, many projects, or many evaluations so that it is possible to learn from the past to improve the future.

Program theory, however, is not always done well. And when it is done badly, it misrepresents what an intervention does and what it can achieve. It can lead to monitoring systems and evaluations that produce an incomplete or distorted picture of what is happening and mistaken judgments about what is effective or efficient. It can demotivate staff and deflect attention from what is important to only what can be easily measured. It can silence important voices or fail to touch those who can act on it. It can take up time without adding value.

The promise of good program theory and the risk of bad program theory have motivated us to write this book. Over more than twenty years, we have worked with small and large organizations in countries all over the world; with municipal, state, and federal government agencies, and nongovernment organizations; on tiny local projects, multimillion-dollar national programs, and whole-of-government strategies; with service deliverers, policymakers, and funders; and in many sectors, including health, education, agriculture, justice, infrastructure, natural resources, community services, community development, and emergency management. Over this time, we have seen diverse approaches to program theory.

What we have learned from this experience, and from the expanding library of empirical research on program theory, is that program theory should be developed, represented, and used not in a formulaic way, but thoughtfully and strategically, in ways that suit the particular situation. We call this *purposeful program theory*.

PURPOSEFUL PROGRAM THEORY

Greek legend tells of the fearsome hotelier Procrustes who would adjust his guests to match the length of his bed, stretching the short and trimming off the legs of the tall. Guides to program theory that are too prescriptive risk creating such a Procrustean bed. When the same approach to program theory is used for all types of interventions and all types of purposes, the risk is that the interventions will be distorted to fit into a preconceived format. Important aspects may be chopped off and ignored, and other aspects may be stretched to fit into preconceived boxes of a factory model, with inputs, processes, outcomes, and impacts.

Purposeful program theory requires thoughtful assessment of circumstances, asking in particular, "Who is going to use the program theory, and for what purposes?" and, "What is the nature of the intervention and the situation in which it is implemented?" It requires a wide repertoire, not a one-size-fits-all approach to program theory.

Purposeful program theory also requires attention to the limitations of any one program theory, which must necessarily be a simplification of reality, and a willingness to revise it as needed to address emerging issues. As the American evaluator Daniel Stufflebeam (2001) has pointed out, evaluators who continue to use an unsuitable program theory are similarly at risk of creating a Procrustean bed for the evaluation.

OVERVIEW OF THE BOOK

The book is designed to help you assess your particular circumstances and develop, represent, and use program theory in appropriate ways. It has options at every stage and examples to help you decide which options to use and how to adapt them to your circumstances. Throughout the book, we draw on examples from our own work and the work of others. ("Our work" refers to projects we have done together and individually.) Each chapter includes exercises to try out new ideas and techniques.

If you are new to program theory, it will be most useful to read the chapters in sequence. If you have some experience or are coming back to the book during an evaluation, you can select the particular chapter you need.

Key Ideas in Program Theory

Part One sets out the key ideas of program theory and how it has developed over time. We explain in Chapter One the essential features of program theory, using the broad policy objective of eating an apple a day to keep the doctor away as an example of how program theory can be used in different ways to learn from success, failure, and mixed results. Chapter Two describes how program theory has developed over time and sorts out the confusion about the different terms that have been used. And Chapter Three introduces seven widespread myths about program theory and seven common traps to avoid.

Assessing Your Circumstances

A key message of this book is the need to approach program theory in a way that suits your circumstances. Therefore, Part Two examines how to analyze the intended uses of program theory and the nature of the situation and intervention.

We explain in Chapter Four why it is important to be clear about who is going to use program theory and for what purposes. A program theory that is useful for developing internal monitoring systems for incremental correction, for example, could be inappropriate for developing performance measures for external accountability. A theory to guide the design of an impact evaluation might not be sufficient to guide a process evaluation that aims to document an unfolding innovation. Being clear about the intended uses of program theory, reviewing this as circumstances change, and considering this when making decisions is an essential part of purposeful program theory.

Chapter Five discusses how to identify simple, complicated, and complex aspects of the program or policy and the situation in which it is being implemented. Program theory can be used for interventions that are simple; that is, they have a single implementing agency and a well-understood causal process that works pretty much the same everywhere. But most interventions have important complicated or complex aspects that program theory needs to address in order not to misrepresent how it works. The implications of complicated and complex aspects of interventions for developing, representing, and using program theory are addressed throughout the book.

Network theory (Granovetter, 1973) is about how the relationships, networks, and connections among entities, and not just the characteristics of the entities themselves, affect outcomes. The entities could be individuals, organizations, special issues groups, or even whole countries. There are many other research-based theories of change, and the chapter lists some other potentially relevant theories that could be used as the basis for an intervention's specific theory of change.

Chapter Twelve outlines some common program archetypes that can be selected, adapted, and combined for particular situations. These include advisory, information, and education programs that seek to change individual behavior by informing decisions; "sticks and carrots," which work through incentives and sanctions; case management; community capacity development; and direct service delivery.

Chapter Thirteen provides examples of variations on pipeline and outcomes chain logic models.

Using Program Theory for Monitoring and Evaluation

The final part of this book describes how to use program theory specifically for monitoring and evaluation.

Chapter Fourteen explains how to use program theory to identify what aspects of the intervention, the context, and results should be measured and how to use key evaluation questions to focus an evaluation in terms of data collection, analysis, and reporting. Program theory can help to structure a coherent narrative report and a focused analysis, whether reporting the results of a single evaluation or bringing together data from many studies. We provide some suggestions on ways to do this for small and large evaluations.

Even when there is credible evidence that outcomes have occurred, can we be confident that an intervention has caused them or at least contributed to them together with other factors? In recent years there has been a vigorous debate about the suitability of different methods and designs to address the issue of causal analysis. In Chapter Fifteen, we set out a three-part framework for causal analysis when using program theory that can bring to bear the full range of research designs and methods for causal analysis. The starting point is looking for congruence of results with those predicted by program theory. The second part is finding relevant comparisons that indicate the difference

that the intervention has made. These can include creating a control group or a comparison group or making other relevant comparisons. The third part is checking out alternative explanations for the results and exceptions to the patterns.

Chapter Sixteen describes ways to bring together information across the different levels of a program theory, or across several interventions that use the same program theory, and how to report this coherently and effectively.

TAKING A STRATEGIC AND ADAPTIVE APPROACH

Program theory can be developed, represented, and used in many ways. Throughout this book, we invite you to take a purposeful approach to program theory, matching it to your situation, checking how it is going, and adapting it as needed to ensure that it contributes to improved interventions and the outcomes you seek.

Key Ideas in Program Theory

1

The Essence of Program Theory

AN APPLE A DAY KEEPS the doctor away—or does it? Thinking about how we would find out if this is true and how we might use those findings shows the value of program theory. In this chapter, we set out the key ideas in program theory and show how program theory can be used to learn from success, failure, and mixed results to improve planning, management, evaluation, and evidence-based policy.

EVALUATION WITHOUT PROGRAM THEORY

Let us imagine that we have implemented a program based on the broad policy objective of an apple a day in order to keep the doctor away. This program, which we dubbed An Apple a Day, involves distributing seven apples a week to each participant. A representation of this program without program theory would simply show the program followed by the intended outcome of improved health (Figure 1.1).

only at the average effect. If we did see differential effects in different contexts (for example, for men compared to women, or in urban areas rather than rural areas), an evaluation without program theory leaves us in the position of having to do simple pattern matching (for example, using the policy for the groups or sites where it has been shown to work) but with little ability to generalize to other contexts.

EVALUATION WITH PROGRAM THEORY

If we used a program theory approach, we would try to understand the causal processes that occur between delivering apples and improved health. We might start by unpacking the box to show the important intermediate outcome that people actually eat the apples. The logic model diagrams in Figure 1.2 show this: one in the form of a pipeline model and one as an outcomes chain. The pipeline logic model represents the program in terms of inputs, processes, outputs, and outcomes. The outcomes chain model shows a series of results at different stages along a causal chain.

Although these look like many logic models that are used regularly in evaluation, they are not much of a theory; rather, they are more like a two-step

Figure 1.2   Simple Pipeline and Outcomes Chain Logic Models
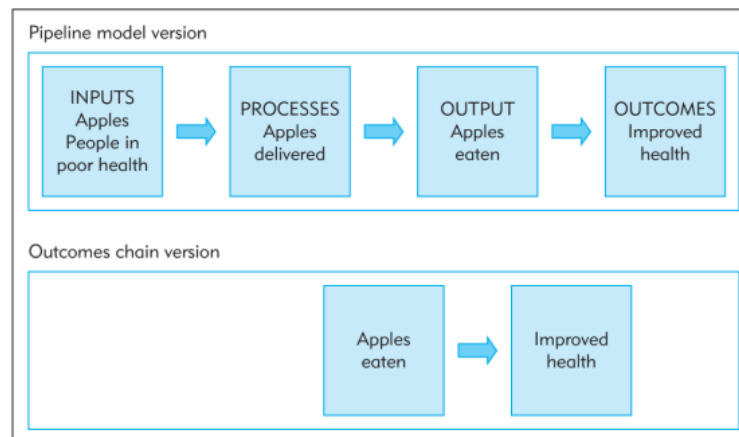
process, as Mark Lipsey and John Pollard (1989) called it, that identifies an intermediate variable without really explaining how it works. These models make it clear that eating the apples is understood to be part of the causal chain (rather than some other variable, such as social interaction with the apple deliverer or physical exercise from playing with the apples). But they do not explain how delivering apples leads to people eating apples or how eating apples improves health.

A plausible explanation would be that delivering apples increases the availability of fresh fruit, which leads to the apples being eaten, which increases the amount of vitamin C in the diet, which improves the physical health of participants. This is only one possible explanation, of course. Figure 1.3 shows this explanation as both a pipeline logic model and an outcomes chain.

The diagrams in Figure 1.3 represent a program theory that articulates the causal mechanisms involved in producing the two changes (changed behavior and changed health status). The first change relates to participants' willingness to act in the way the program intended and the second to the impacts of their actions. For many programs, it can be helpful to articulate both types of changes in the program theory.

Figure 1.3   A Logic Model Showing a Simple Program Theory for An Apple a Day Based on Improved Vitamin Intake
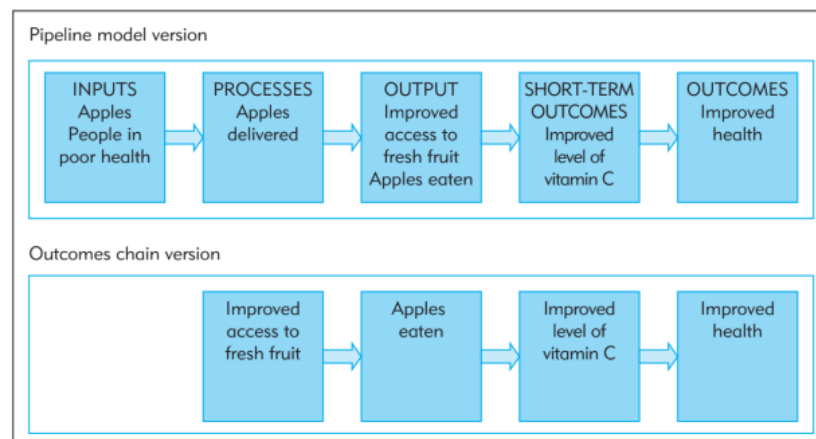
Learning from Failure

An evaluation based on this program theory would collect data about changes in access to fresh fruit, apple eating behavior, and nutritional status, as well as overall health. If the intended outcomes have not been achieved, we could work through the causal chain to see where it has broken down. If the apples were not even delivered, there is obvious implementation failure; if they were delivered but not eaten, then our theory of how to engage people in changing their behavior seems not to work. Similarly, if the expected health benefits had not been achieved, we would start by seeing if it was because the apples had not been eaten. If the apples had been delivered and had been eaten but without producing health improvements, then we have a problem with the theory of change that underpins the program. Based on these results, one option would be to reject the theory and look at other ways of improving health. Another would be to look at dosage: maybe vitamin C levels increased, but not enough to make a difference.

Learning from Partial Success

Developing a program theory also helps clarify differential effects, learning from those participants for whom the program was effective. The simple program theory is based on the assumption that the apples are both necessary and sufficient—that is, the apples will lead to good health in all circumstances and without contributions from other factors. Developing a more complicated logic model would focus on the differential effects we might expect for different types of participants, and we would collect and analyze data to examine these. Disaggregating the data would investigate whether the theory works in some contexts but not in others.

This review might show that the program works only for certain types of participants—for example, those who are affected by diseases related to inadequate nutrition. For people affected by infectious diseases, apples by themselves might not be enough to improve health. Based on these results, we might target the program to people most likely to benefit: those with nutrition-related diseases. Given the importance of the interaction between the intervention and the characteristics of clients, it would be helpful to revise the theory of change and its logic model to show this complicated causal path.

If the program works for some groups but not for others or at some sites but not others, it is important to try to understand why by identifying possible explanations and then checking these out empirically. For example, if the program worked for men but not for women, it might be because of differences in labor force patterns which affected access to fresh fruit or to differences in nutritional needs related to pregnancy. Finding exceptions to the pattern (the men who did not improve and the women who did) would provide more evidence to test these emerging program theories.

Learning from Success

Program theory has another benefit when an evaluation finds that something works: it helps in adapting the intervention to new situations. To be useful for evidence-based policy and practice, a program theory evaluation needs to identify the causal mechanism by which it works and determine whether this is different for different people and in different implementation contexts.

To explore this use, imagine that the evaluation has found that the program theory works: people are healthier when they eat an apple a day. Now the job is to implement a new program based on this evidence. In this case, the goal is not to understand failure but to understand success. Apples might produce these effects through quite different theories of change, which would lead us to quite different intervention theories and different program activities to suit the context. We would immediately have many questions about the statement. Does it work for everyone? Does it have to be a particular variety of apple (Granny Smiths? crab apples?), or does it apply to all varieties? What if apples are not available? Can we substitute other fruit, or apple juice, or vegetables? Would red onions work as well as red apples? An evaluation without program theory would reveal only that it works, with no guidance for how to translate the findings to a particular situation. Without this guidance, we can only blindly copy everything. With this guidance, we can understand how we might adapt it and still achieve the intended results.

We previously sketched out a program theory with a theory of change of providing a good source of vitamins in diets that are otherwise deficient. To test this out if we were implementing it would require data about people's nutritional status through either direct measures or relevant indicators so we

could see if there was any change and also to identify the people we would expect to get the most benefit from the program. We would want to check that they actually ate the apples. And we would want to rule out alternative explanations by finding out if there had been other changes in their diets that might have contributed to changes in their nutrition. If this is the case, then other types of fruit are likely to be equally effective. In a country where apples are hard to obtain or expensive, distribution or subsidization of local fruit is likely to be an effective program, at least for people at risk of nutritional deficiency, if it is implemented well.

But maybe this is not how it works at all. Maybe it is not about the flesh or juice of the apples but their skin. The skin of apples contains a plant-based chemical called quercetin. Some research studies have suggested quercetin may help to prevent cancer, heart disease, and inflammation of the prostate. An evaluation would look at the intake of quercetin from various sources and outcomes in terms of these specific diseases, focusing on outcomes for people at risk of these diseases. If apples were not available, another source of quercetin could be used. Red onions, a rich source of quercetin, might be an effective substitute—an adaptation of the program that would not be immediately obvious if we were thinking only about fruit.

Another possible explanation focuses on apples as a substitute for high-calorie, low-nutrition snacks. Perhaps apples improve health by helping to reduce obesity as people stop eating potato chips and doughnuts and choose apples instead. An evaluation of this possibility would look at what people were eating in addition to apples and whether there had been a decline in their consumption of junk food. It also might measure short-term outcomes such as body mass index (BMI) and percentage fat, which have been linked to subsequent longer-term health outcomes. The evaluation would have to take into account criticisms that have been made of BMI as an indicator and predictor of health. Making other low-calorie snacks such as carrots and celery readily available might be equally effective. Figure 1.4 shows how these three different change theories might plausibly explain why the policy works.

Other possible explanations, involving different theories of change, would lead to different critical features in implementation that should be ensured. For example, if health improvements came about through increased

SUMMARY

This chapter has used a hypothetical example to explore how articulating a program theory—an explicit statement of how change will occur and how an intervention will produce these causal processes—can make evaluations more useful. Throughout the rest of the book, we use examples from actual evaluations to show how to develop, represent, and use program theory for evaluation and other purposes.

EXERCISES

1. If a social marketing campaign was used instead of direct delivery of apples for the Apple a Day program, what would implementation failure look like? What would theory failure look like? What would partial theory failure look like, where it works only in particular contexts?

2. Consider a policy that aims to increase student performance by increasing teachers' salaries. What might be some alternative causal mechanisms that would produce the intended outcomes?

2

Variations of Program Theory over Time

DESPITE OCCASIONAL STATEMENTS that program theory is a new approach, its roots go back more than fifty years. Over time, there have been important variations and different emphases. This chapter reviews these developments, including the different terms that have been used inconsistently over time. We provide a guide for translating the different labels that have been used for concepts and the different ways these have been used.

A SHORT HISTORY OF PROGRAM THEORY

The history of program theory evaluation is not one of a steady increase in understanding. Instead, many of the key ideas have been well articulated and then ignored or forgotten in descriptions of the approach. It is not unusual to have statements that demonstrate a lack of knowledge of previous empirical and theoretical developments, such as a call for proposals from the Agency

for Healthcare Research and Quality (2008) that claimed that "'theory-based evaluation' is a relatively new approach" (p. 14).

Philosophical Roots

The value of intentional action has long been recognized in philosophy. The Greek Stoic philosopher Epictetus wrote nearly two thousand years ago, "First say to yourself what you would be; and then do what you have to do." However, early planning and evaluation focused on specific objectives rather than on articulating the links between activities and these objectives.

Early Examples

Probably the first published use of what we would recognize as program theory was a series of four articles on evaluating training by Don Kirkpatrick (1959a, 1959b, 1960a, 1960b) published in the *Journal of the American Society for Training and Development*. Kirkpatrick looked at reactions of participants to training; learning in terms of new knowledge, skills, and attitudes; behavior in terms of applying the learning back in the work environment; and results in terms of targeted outcomes. Over time these four categories became known as Kirkpatrick's Four Levels of Learning Evaluation.

Kirkpatrick argued that this sequence of results should first be used for planning purposes, beginning with the ultimate targeted outcomes, working back to the behavior needed to produce these; the new knowledge, skills, and attitudes needed to be able to engage in the behavior; and the training experiences that would be needed to produce a positive reaction from participants. Then this sequence of results could be applied as a framework for planning evaluation, where evidence from each level could be used to build an argument about the contribution of the training.

This idea was more broadly applied by Edward Suchman, who argued in 1967 that program evaluation would benefit from explicitly examining the achievement of a "chain of objectives" (p. 55). He drew attention to the need to identify and examine an intervening process in between an activity and its objective. "The evaluation study," he wrote, "tests some hypothesis that activity A will attain objective B because it is able to influence process C which affects the occurrence of this objective. An understanding of all three

factors—program, objective and intervening process—is essential to the conduct of evaluative research" (p. 177).

In the same year, Daniel Stufflebeam (1967), in a discussion of the limitations of experimental approaches to evaluating education programs, outlined a new evaluation model. It was in the form of a generic program theory, although it was not labeled as such. The CIPP model (Stufflebeam, 1967) set out an intervention in terms of four boxes—context, input, processes, and product—and asked a series of questions about each of them. This format put all the results together in one category, which avoids debates about whether a result is best classified as an output, an outcome, or an impact. It was one of the few models that incorporated context as an intrinsic part of the description of an intervention.

In 1969 a version of program theory, the logical framework approach, commonly referred to as the logframe, was developed for Practical Concepts Incorporated (1979). The title of the report outlining the approach, *The Logical Framework: A Manager's Guide to a Scientific Approach to Design and Evaluation,* shows that it was also intended to be used for both planning and evaluation. Subsequently the logframe was further developed for use by U.N. agencies by Gesellschaft für Technische Zusammenarbeit, the German international development agency. In the logframe, the causal chain was standardized into four components: activities, outputs, purpose (the rationale for producing the outputs), and goal (a higher-level objective to which this program and others contributed). For each component, four aspects were articulated: a narrative description, objectively verifiable indicators, means of verification, and assumptions (factors outside the control of the program on which the success of achieving that component depended). Because the logframe is still widely used in international development, we look at it in detail in Chapter Thirteen.

The idea of identifying and measuring intermediate variables was used in three evaluations related to safety: a speed control program, a motor vehicle inspection program, and an improved ambulance system (Hall and O'Day, 1971). They argued that this "causal chain" approach to evaluation provided more realistic indicators of the success of an intervention than evaluations that solely measured final results. In 1972, in one of the earliest books on

New Directions for Program Evaluation, this time focusing on advances in program theory (Bickman, 1990), describing how to develop program theory (Chen, 1990b) through path analysis (Smith, 1990), social science theory (Riggin, 1990), and pattern matching (Marquart, 1990); how to use program theory to understand program quality (Bickman and Peterson, 1990) and program types (Conrad and Buelow, 1990); and different ways of testing program theories (Mark, 1990; McClintock, 1990).

During the 1980s, in response to economic challenges and concerns about accountability, many countries introduced reforms within the public sector to focus on managing for results rather than only for compliance with processes, sparking the first boom in the use of program theory and logic models (U.S. General Accounting Office, 1995). In Australia, a system of program management and budgeting was introduced in 1983 requiring departments to articulate specific goals, explain how they planned to meet them, and report on performance. In Canada, the federal government implemented a new expenditure management system that incorporated a shift in focus from activities and outputs to impacts and results, and requiring departmental business plans to articulate priorities, set goals, and report on performance.

Some organizations adopting program theory focused on the Suchman notion of a chain of objectives. In the 1980s, Bryan Lenne led a team in the New South Wales Public Service in Australia that worked with departments across the state to use program theory in the form of outcomes hierarchies (Lenne and Cleland, 1987; Funnell and Lenne, 1990). Program activities were not necessarily at the front end of these chains, but they could contribute directly to later outcomes. External factors were also to be included in the program theory (in Chapters Seven and Eight, we discuss this approach in more detail). This approach was subsequently adopted by state and federal governments across Australia (Funnell, 1990; Milne, 1993). The Roundtable on Comprehensive Community Development, supported by the Aspen Institute, published an influential paper by Carol Weiss (1995) showing how theory-based evaluation, in the form of an outcomes chain, could be useful for programs where classic experimental and quasi-experimental approaches were not possible.

Other organizations focused on a pipeline approach to program theory. The United Way (1996), a nonprofit organization that works in a coalition of charitable organizations, developed a guide to developing and using logic models for outcome measurement that was widely accessed both through the Internet and in hard copy (United Way, 1996). It set out a four-box logic model of inputs (both resources and constraints), activities, outputs (direct services or products), and outcomes for participants. The W. K. Kellogg Foundation produced the *Logic Model Development Guide* (2004), which was also in the form of a linear template of five components: inputs, activities, outputs, outcomes, and impact. This template defined impacts as results for the broader community, organization, or system, beyond the individuals who participated in the intervention. NORAD, the Norwegian government international development agency, published a widely used guide to the logical framework approach that has now gone through four editions (NORAD, 1999).

More evaluation books and guides began to include discussion of program theory and logic models. For example, a widely used textbook, *Program Evaluation: A Systematic Approach,* added a chapter in its sixth edition (Rossi, Freeman, and Lipsey, 1999). Program theory was added to the repertoire in the revised edition of *Evaluation Models*, edited by Daniel Stufflebeam, George Madaus, and Tom Kellaghan (Rogers, 2000a). The British government included discussion of program theory in its *Magenta Book,* which provided a guide to planning and evaluation (Government Chief Social Researcher's Office, 2003).

A third issue of *New Directions for Evaluation* focusing on program theory was published (Rogers, Petrosino, Hacsi, and Huebner, 2000), discussing the promise of program theory for replication (Hacsi, 2000), meta-analysis (Petrosino, 2000), performance monitoring (Funnell, 2000), and developing a shared understanding (Hubener, 2000), and the challenges in terms of addressing causal inferences (Weiss, 2000; Davidson, 2000; Cook, 2000; Rogers, 2000b).

Innovations

A new conceptualization of program theory was introduced by Ray Pawson and Nick Tilley. In a paper in the *British Journal of Criminology* (Pawson and

Tilley, 1994) and their subsequent book, *Realistic Evaluation* (Pawson and Tilley, 1997), they set out a realist approach to evaluation, where program theory was understood as configurations of context-mechanism-outcome. Context was seen as a critically important part of program theory because causal mechanisms fire only in favorable contexts, that is, in particular implementation environments or for particular types of participants. We discuss the importance of differential program theories in Chapter Five and how to represent these in realist matrices in Chapter Nine.

Some different approaches to program theory were developed to address concerns that logic models make it seem that the entire causal process is under the control of program implementers. Sarah Earl, Fred Carden, and Terry Smutylo (2001), working with Barry Kibel, developed outcome mapping, an approach designed for interventions where implementers cannot control impacts but seek to influence these by affecting the behavior of boundary partners. Steve Montague, Gail Young, and Carolyn Montague (2003) described this in terms of circles of influence: the operational environment, the environment of direct influence, and the environment of indirect influence. We explore these ideas in more detail in Chapters Seven and Nine.

Variations on pipeline models were also developed, drawing on Bennett's hierarchy from 1975. In Canada, Steve Montague (1998) argued that logic models needed to include "reach," that is, articulating where and with whom particular results were intended. A generic logic model was developed at the University of Wisconsin that included articulating who was expected to participate, but also showing assumptions, external factors, needs, and priorities on the logic model (University of Wisconsin, 2003). We explore these variations in Chapter Thirteen.

The Current State

In more recent years, the use of program theory has become part of the mainstream of most approaches to evaluation. Many organizations in many countries are referring to program theory and often requiring its use in planning proposals and reporting. For example, the European Commission has included discussion of program theory in its guide to evaluability assessment (European Commission, 2009). More resources have been developed

to support people in learning to develop and use program theory. A special topical interest group of the American Evaluation Association, Program Theory and Theory-Driven Evaluation, has been established. And a number of books focusing on program theory and logic models have been published (Frechtling, 2007; Donaldson, 2007; Knowlton and Phillips, 2008). However, the three issues of concern identified by Weiss in her review of twenty-five years (Weiss, 1997)—not having an articulated theory about how change comes about, or having a poor theory, or not really using it to guide the evaluation—are continuing features of many examples of program theory (Rogers, 2007; Astbury and Leeuw, 2010).

This explosion of activity has produced great diversity in what program theory is called, how it is represented, and how it is used (Coryn, Noakes, Westine, and Schröter, forthcoming). Sometimes organizations use a common template of components (such as a logframe or other pipeline model), and other times, there is scope for a more flexible representation, such as an outcomes chain. Sometimes the expectation is that a single program theory will be developed and used as a reference point throughout planning, implementation, and evaluation, and other times there is scope for developing several versions or revising the program theory throughout the process. Sometimes the program theory is developed by program staff, sometimes by an external evaluator, and sometimes this is done collaboratively. We discuss these options in Chapter Six.

This rich diversity of experience presents a wide range of options at each stage, which can be quite confusing. This book presents the diversity of options and also a way to select, combine, and adapt these to match the situation at hand.

TERMINOLOGY IN PROGRAM THEORY

Over the years, many different terms have been used to describe the approach to evaluation that is based on a "plausible and sensible model of how the program is supposed to work" (Bickman, 1987b):

- *Chains of reasoning* (Torvatn, 1999)
- *Causal chain* (Hall and O'Day, 1971)

- *Causal map* (Montibeller and Belton, 2006)

- *Impact pathway* (Douthwaite et al., 2003)

- *Intervention framework* (Ministry of Health, NZ 2002)

- *Intervention logic* (Nagarajan and Vanheukelen, 1997)

- *Intervention theory* (Argyris, 1970; Fishbein et al., 2001)

- *Logic model* (Rogers, 2004)

- *Logical framework (logframe)* (Practical Concepts, 1979)

- *Mental model* (Senge, 1990)

- *Outcomes hierarchy* (Lenne and Cleland, 1987; Funnell, 1990, 1997)

- *Outcomes line*

- *Performance framework* (Montague, 1998; McDonald and Teather, 1997)

- *Program logic* (Lenne and Cleland, 1987; Funnell, 1990, 1997)

- *Program theory* (Bickman, 1990)

- *Program theory-driven evaluation science* (Donaldson, 2005)

- *Reasoning map*

- *Results chain*

- *Theory of action* (Patton, 1997; Schorr, 1997)

- *Theory of change* (Weiss, 1998)

- *Theory-based evaluation* (Weiss, 1972; Fitz-Gibbon and Morris, 1975)

- *Theory-driven evaluation* (Chen and Rossi, 1983)

The labels and definitions used in this book are not the only ones in use. Given the wide variety of terms used, it is important to be aware of the other labels you might come across. These terms are not always used interchangeably; sometimes they have particular meanings, but these vary widely. In particular, although the terms *program theory* and *program logic* are often used interchangeably, they sometimes focus on particular distinctions. As we have worked on evaluations in different organizations, we have been confronted by very different uses of these terms. Table 2.1

appropriate, and the wording can suggest that earlier results are unimport-
ant. For some people, any label with "program" in it will be problematic
because "program" has a specific meaning, and they are managing or evaluat-
ing something other than a program (such as a project, strategy, initiative, or
policy) and want a term that sounds as if it applies to their work.

Terms for Components in a Pipeline Model

Many pipeline models use the terms *input, processes, outputs, outcomes,* and
impacts. But this is far from a universal practice. Outputs can be defined in
very different ways: as completed activities, tangible products, or services, or
as the first change in the causal chain. The distinction between outcomes and
impacts is sometimes made in terms of sequence, but there is little consis-
tency of approach (sometimes outcomes are before impacts, and sometimes
vice versa) or scope (with outcomes defined as results for individuals and
impacts as results for the broader community or organization). The follow-
ing list shows the labels and definitions used by some major international
organizations:

Australia: AusAID (Australian Agency for International Development) (Logframe)

Outputs: The tangible products or services that the activity will deliver

Component objectives or intermediate results: A level in the objectives or
results hierarchy that can be used to provide a clear link between
outputs and outcomes

Purpose or *outcome:* The medium-term results in terms of benefits to the
target group that the activity aims to achieve

Goal or *impact:* The long-term development impact (policy goal) that
the activity contributes to at a national or sectoral level (http://www
.ausaid.gov.au/ausguide/pdf/ausguideline3.3.pdf)

Canada: Treasury

Activities: Key activities intended to contribute to the achievement of
the outcomes

Outputs: Products or services generated by the activities

Immediate outcomes: Short-term outcomes that stem from the activities and outputs

Intermediate outcomes: Next links in the chain of outcomes that occur after immediate outcomes

Final outcomes: Final outcomes, or why these activities are being engaged in

(http://www.tbs-sct.gc.ca/cee/dpms-esmr/dpms-esmr05-eng.asp#_5.1_Overview_of)

India: Comptroller and Auditor-General

Objectives

Aim—the broad mandate, legislative direction

Objectives—translation of the mandate into well-formed objectives

Targets—well-defined physical goals to achieve the objectives

Inputs: Resources that are to be transformed by program instruments into outputs

Processes: Planning, organization, and implementation of the program operations to produce the outputs

Outputs: Results achieved by the program operations that are within the control of the program manager

Outcome: The broad effects of the program outputs; expected to meet the program objectives and aim and influenced by external factors (http://www.cag.gov.in/publications/peraudrepcent/appendix%20E.pdf)

South Africa: Framework for Managing Programme Performance Information, National Treasury

Inputs: All the resources that contribute to the production and delivery of outputs: "What we use to do the work"

Activities: The processes or actions that use a range of inputs to produce the desired outputs and, ultimately, outcomes: "What we do"

Outputs: The final products, goods, and services produced for delivery: "What we produce or deliver"

Outcomes: The medium-term results for specific beneficiaries that are the consequence of achieving specific outputs: "What we wish to achieve"

Impacts: The results of achieving specific outcomes, such as reducing poverty and creating jobs (http://www.treasury.gov.za/publications/guidelines/FMPI.pdf)

United States: University of Wisconsin Extension Service

Inputs: What we invest

Outputs–Activities: What we do

Outputs–Participation: Who we reach

Short-term outcomes: Short-term results

Medium-term outcomes: Medium-term results

Impact: Ultimate impact

(http://www.uwex.edu/ces/pdande/evaluation/evallogicmodel.html)

United States: Innovation Network (InnoNET)

Resources: What you have to work with

Activities: What you will do with your resources in order to achieve program outcomes and, ultimately, your goal

Outputs: Tangible products of your activities

Outcomes: Changes expected to occur as a result of your work

Goal: Overall purpose of your program

(http://www.innonet.org/client_docs/File/logic_model_workbook.pdf)

United States: W. K. Kellogg Foundation

Resources/inputs: Resources needed to operate program

Program Activities: Processes, tools, events, technology, and action that are an intentional part of the program implementation

Outputs: Types, levels, and targets of service delivered

Outcomes: Specific changes in program participants' behavior, knowledge, skills, status, and level of functioning

Impact: Changes to organizations, communities, or systems as a result of program activities within seven to ten years

(http://www.wkkf.org/~/media/6E35F79692704AA0ADCC8C3017200208 .ashx)

United States: United Way

Inputs: What we invest to make the program happen

Activities: What we do in our program

Outputs: Products and participation—the "how manys"

Outcomes: Benefits for participants during and after program activities

(http://www.yourunitedway.org/media/Guide_for_Logic_Models_and_ Measurements.pdf)

KEY IDEAS IN PROGRAM THEORY

Program theory involves a particular use of the words *program* and *theory.* In this section, we address how we have used terms in this book. If you are working in an organization or with an organization that uses different terms for these concepts, it might be useful to pencil in the local equivalent.

Program

In *program theory,* the term *program* refers not only to something formally labeled as a program (for example, in a corporate management hierarchy of programs, subprograms, and components). It can refer to any intervention: a project, a strategy, a policy, a funding initiative, or an event. It includes interventions that are undertaken by a single organization, such as a direct service delivery project, and those that are undertaken by multiple organizations, such as a whole-of-government policy. It refers to both preplanned and tightly specified interventions, and broadly defined and emergent interventions.

Program Theory

A program theory is an explicit theory or model of how an intervention contributes to a set of specific outcomes through a series of intermediate results. The theory needs to include an explanation of how the program's activities contribute to the results, not simply a list of activities followed by the results, with no explanation of how these are linked, apart from a mysterious arrow. We find it helpful to think of a program theory as having two components: a theory of change theory and a theory of action. One of the benefits of articulating program theory is being able to systematically review the quality of the theory in terms of plausibility, consistency with evidence, and utility.

Theory of Change

This refers to the central mechanism by which change comes about for individuals, groups, and communities. For example, many health promotion programs are based on a theory of change that behavior changes in response to perceived social norms. It is possible to have more than one theory of change. There might be different theories of change at different stages of the program (for example, one about participants becoming engaged with the program and one about their changing their behavior) or for different groups of people (for example, some people might change their behavior in response to new information about risks and benefits, while others might change only in response to tangible incentives).

Programs are usually, but not always, about change. Sometimes a program aims to stop or reduce change or prevent something from happening—for example, maintaining the mobility of a person with a disability who might otherwise develop restricted movement, or maintaining biodiversity despite pressures from agriculture and industry. In these cases, the theory of change explains how pressure to change will be resisted or deflected. Another way to think of this is that the program is changing a situation from what it otherwise would have been.

Theory of Action

This explains how programs or other interventions are constructed to activate their theory of change—for example, what the program does to change

Table 2.2   Definitions Used in This Book

Term	Definition	Your Organization's Label (if applicable)
Program theory	An explicit theory of how an intervention is under-stood to contribute to its intended or observed outcomes; ideally includes a theory of change and a theory of action	
Theory of change	The central processes or drivers by which change comes about for individu-als, groups, or communi-ties. It can be derived from a research-based theory of change or drawn from other sources.	
Theory of action	The ways in which pro-grams or other interven-tions are constructed to activate these theories of change.	
Theory of Change	A research-based theory of change	
Logic model	A representation of a pro-gram theory, usually in the form of a diagram	

SUMMARY

The long and rich history of program theory has produced many different ways of developing, representing, and using it. It has also produced many different terms and definitions. Being aware of this can help you to make the best choices to suit your situation, and communicate effectively with other people who are using different terminology.

EXERCISES

To answer these questions, consider the organization you work in or one in your area of work.

1. What terms are commonly used in the organization to refer to program theory?

2. Does the organization prescribe or recommend a particular form of logic model? If so, what is it?

3. Do partner organizations use the same terminology? If not, what are the differences?

3

Common Myths and Traps

AS PROGRAM THEORY has become more widely used, a number of myths have arisen about how to go about it and how useful it might be. In this chapter, we challenge seven myths that can get in the way of effective use of program theory. We also set out seven common traps for those using program theory and discuss how to avoid them. We provide more detail throughout the book about avoiding these traps.

SOME COMMON MYTHS

As program theory has become increasingly widely used, and even mandated, within organizations, a number of myths have been promulgated. Knowing that these are myths and being able to discuss them with colleagues and partners will reduce your risk of making common mistakes when using program theory. The myths are set out in Table 3.1 and discussed in the sections that follow.

Table 3.1   **Some Common Myths About Program Theory**

Myths	How These Are Sometimes Expressed
1. New approach	Program theory is a new approach to evaluation, and what is contained in a local guide is all there is to know about it.
2. One way to do it	There is one way to draw a logic model.
3. Not a good model	Program theory does not address all necessary aspects of evaluation and therefore is not a good model of evaluation.
4. Too much time	Developing a credible and useful program theory always requires so much time, content knowledge, and research expertise that it is usually impractical.
5. Just draw it	Developing a credible and useful program theory can be done by simply asking people to draw a logic model.
6. Can't really test it	Testing a program theory can be done only by formal experimental methods, which are beyond the scope of most evaluations.
7. Assume causality	Program evaluation should not bother with testing the program theory. If results are consistent with the theory, it can be assumed that the program has caused the results.

Myth 1: A New Approach

A surprising number of introductory guides to evaluation state that this is a new approach. These statements demonstrate an ignorance of previous developments of program theory and of the lessons that can be learned from these. Program theory has been used for over fifty years, and current practice should draw on what has been learned by this.

Myth 2: One Way to Do It

Many guides to program theory show only one way to draw a logic model—usually a version of a pipeline model. In fact, there are many ways to draw logic models, including results chains and causal matrices. Some show activities and some external factors. Some show how the program works differently for different types of participants. Some show how the program works in combination with other programs. Some should be read left to right, some

program theory, drawing on systematic review of previous research literature and expert review. Investing these resources is particularly important when a program theory is being developed for a large-scale or high-risk intervention. Purposeful program theory will find the appropriate investment of time in developing the program theory. Chapter Six describes the range of evidence sources that should be used to inform the development of program theory.

Myth 6: Can't Really Test It

Evaluations can use the whole array of methods for causal attribution, including experimental, quasi-experimental, and nonexperimental methods. It is not realistic to expect that any single evaluation, or any single research project, can provide a definitive proof of a program theory, but evaluations can and should provide insights into its credibility. In Chapter Fifteen, we discuss different ways to undertake causal analysis using program theory, including rigorous systematic approaches that can be used when this is a major focus of the evaluation.

Myth 7: Assume Causality

This is the opposite of the previous myth, and again the truth is somewhere in between. Converting a logic model into a series of indicators and then reporting these as evidence of the impact of the program can be tempting, but it is inadequate for evaluation. In Chapter Fifteen, we discuss different ways to undertake causal analysis using program theory, including approaches that can and should be added into small evaluations.

TRAPS TO AVOID WHEN DEVELOPING AND USING PROGRAM THEORY

Being able to draw a logic model and produce an evaluation plan does not mean that program theory has been developed and used well. This book provides advice on avoiding seven common traps, which we set out in Table 3.2 and discuss in the sections that follow.

Table 3.2   **Seven Traps to Avoid When Developing and Using Program Theory**

1. No actual theory
2. Having a poor theory of change
3. Poorly specifying intended results
4. Ignoring unintended results
5. Oversimplifying
6. Not using the program theory for evaluation
7. Taking a one-size-fits-all approach

Trap 1: No Actual Theory

Many versions of program theory, particularly those using a pipeline approach to logic modeling, fall short of having an actual theory. They simply display boxes of activities and boxes of outcomes without demonstrating logical and defensible relationships between them and the various items listed in the boxes. This setup can make it difficult to understand the causal chain and identify what measurement and evaluation would be appropriate for purposes of causal attribution.

Program activities and resources are what staff are most familiar with on a day-to-day basis. So it is not surprising that the program theories that they construct can be more preoccupied with what the program does than what it achieves in terms of outcomes. Renger and Titcomb (2002) introduced the notion of activity traps: well-intended activities that appear to address particular problems but on closer inspection do not address any of the conditions that underlie the problems. Flowing on from these activity traps, a program monitoring system may expend much effort on measuring activities of little consequence in producing outcomes and resolving the problems or issues that the program was established to address.

A frequent criticism of monitoring systems and, perhaps to a lesser extent, evaluation studies is that they place too much emphasis on measures of busyness—measures of what the program does on the assumption that what it does will produce outcomes. In other words, the measures of activity come to be used as proxy measures of outcomes. It is particularly important in these situations to be confident that what the program does is relevant

to the outcomes it wishes to achieve and that the program has not become caught in activity traps.

A program theory should be able to show that each activity is relevant to achieving one or more outcomes and that each outcome is addressed by one or more program activities unless there is reason to believe that activities that achieve lower-level outcomes will propel clients to higher levels of outcomes with little further program intervention. Even if the activities are relevant, a program theory that gives excessive attention to the detail of activities and insufficient attention to outcomes can lead to monitoring and evaluation that focuses on aspects of program delivery rather than outcomes.

Strategies to Avoid This Trap

- Instead of having a single box labeled "outcomes" in a pipeline logic model, present it as an outcome chain to show the assumed relationships among the various outcomes (see Chapter Seven).

- Identify program and nonprogram factors that affect each outcome and how the program activities and resources address those factors (see Chapter Eight).

- Identify all significant program activities and resources and show their relevance to one or more outcomes and to factors that will affect outcomes, and identify activity traps and gaps (see Chapter Eight).

Trap 2: Having a Poor Theory of Change

Just because you can represent a program theory in a logic model does not make it credible or relevant. Program theories may fail to provide a credible explanation of why one would expect a higher-level outcome to flow from a lower-level one. For example, it is not unusual to see a logic model that explains that a program to change health behaviors is understood to work by providing people with information about the health consequences of their choices of diet, exercise, and smoking. Knowledge of these consequences (achieved, for example, by improving food labeling about fat and calories) might be a necessary part of changing choices, but it is rarely sufficient to achieve behavior change. A program that works only through this mechanism is unlikely to be successful. More problematic, a monitoring system

that looks only at these processes will be ignoring other important processes and intermediate outcomes and either directly or indirectly discouraging them.

From some program theories, it can be difficult to envisage how the outcomes that are the main focus of a program (and are also likely to be the focus of measurement and evaluation) will make a difference to the overall problem that gave rise to the program. There may be missing links and large gaps between the intermediate outcomes that the program is designed to achieve and the ultimate outcomes to which the program is to contribute in order to reduce or resolve the problem. For example, the outcomes chain for a mass media campaign to improve awareness of HIV/AIDS would have a significant gap if the chain jumped straight from increased awareness to reduction in HIV/AIDS. Sometimes these gaps in the outcomes chain occur because the situation that gave rise to the program (the nature and extent of the problem, its causes and consequences) has been poorly conceptualized and analyzed, measured, or poorly documented and explained.

Another version of this trap is to develop a credible solution but to the wrong problem. For example, people's failure to consume fresh fruit and vegetables might not be a deliberate choice, and therefore amenable to strategies aimed at changing individuals' choices. Rather, it might be a consequence of constrained options in their local neighborhood or budget, which would require a solution at the supply end, not the demand end, or it may be a combination of both. Moreover, further analysis or social marketing may show that even if one identified impediment is removed, compelling reasons, such as personal preferences, may exist not to consume fresh fruit and vegetables. When behavior change is an objective, good evidence about what motivates different people is essential.

Failure to undertake a good situation analysis can lead to developing a solution to the wrong problem and producing a program theory that does not reflect the situation (needs, resources, problems, and opportunities) or its causes and consequences or that has large gaps in the causal chain. Inadequate formulation can lead to failure to collect baseline data for program evaluation and to routinely monitor the problem addressed by program causes and consequences.

Strategies to Avoid This Trap

- Involve the right mix of people to develop the program theory to ensure that adequate knowledge is brought together to develop a plausible and defensible program theory that can be directly related to the situation analysis (see Chapter Six).

- Draw on research and previous evaluations to identify previous theories that may be relevant (see Chapter Six).

- Undertake a situation analysis that identifies problems and opportunities to be addressed, causes and consequences, including baseline data and other information about the problem and its causes and consequences (see Chapter Seven).

- Systematically critique the quality of the program theory (see Chapter Ten).

- Draw on wider theories of change and theories about how different types of program work (see Chapters Eleven and Twelve).

- Include an evaluation of the program theory as part of an empirical program evaluation (see Chapter Fourteen).

Trap 3: Poorly Specifying Intended Results

This trap has two variations. One is that the program theory gets stuck in the direct, tangible products of the program and does not include the longer-term outcomes and impacts that form its rationale. For example, the outcomes are expressed in terms of completed cases rather than anything about the results for clients. This can come about when people believe they will be held accountable for achieving everything in the program theory, and they know that they cannot totally control the results for clients. The other variation is that the longer-term outcomes and impacts are included but expressed in narrow ways that reflect what is readily measurable rather than what is actually the intended result. For example, instead of including a hard-to-measure outcome such as well-being, the program theory shows "percentage of clients satisfied," which is an incomplete indicator of the actual outcomes sought.

Strategies to Avoid This Trap

- Specifically include consideration of possible unintended effects when articulating the program theory, including using negative program theory—a model of the program that represents how the intervention strategy might lead to negative results (see Chapter Six).

- Specifically include unintended results in the evaluation plan, drawing on the negative program theory and also data collection strategies that will capture unintended and unanticipated results (see Chapter Fourteen).

- Include consideration of unintended results when developing an overall judgment about the effectiveness of a program (see Chapter Sixteen).

Trap 5: Oversimplifying

Some aspects of programs are simple, with clear causal sequences under the control of the program, which can be identified in advance and managed tightly. However, not all aspects of programs are like this.

Many interventions have complicated aspects. A program might be just one piece of the jigsaw needed to produce the intended results, and it will work only if other components are in place—for example, other interventions, favorable implementation environment, or particular participant characteristics. A program theory needs to identify the outcomes that others need to achieve, as well as those to be achieved by the program.

Many interventions have complex aspects where the program cannot be specified in advance. In such cases, the program is appropriately emergent and adaptive in response to needs and opportunities that arise and as understandings of what is effective develop over time. As open systems, programs need to be scanning their environment, looking for warning signs, and picking up on opportunities to influence those out-of-scope conditions that are critical to the program's success. Complex programs, in particular, need to adopt approaches that involve working with other systems and subsystems.

Treating a program as if it were a simple, closed system, when in fact it has important aspects that are complicated or complex, can lead to insensitivity

to the inherent unpredictability of the contexts within which programs function and the ways in which they are implemented, the need for programs to adapt, and the likelihood, and in many cases desirability, of emergent outcomes. Although it is important to define the scope and boundaries of a program, it is also important to recognize that boundaries are often moving and that what is on both sides of the boundary, in scope and out of scope, need to be considered in the program theory. These considerations include how what is out of scope interacts with and affects what is in scope and the extent to which there is fluid movement back and forth across the boundary. It is important to monitor and evaluate the impacts on program success of out-of-scope outcomes of other programs and other factors or conditions that affect program outcomes and its capacity to make a contribution to resolving the problem. If this is not done, there is a risk of overclaiming or underclaiming outcomes. When external factors are an element in producing outcomes, then attributing the outcomes to the program may be overclaiming. However, external factors may run counter to and dilute the apparent effects of a program. In this case, ignoring them can lead to underclaiming results. The situation might very well have been worse without the program. Counterfactuals (what would have happened without the program), even if difficult to apply, should be considered a key part of developing a program theory.

Although recognizing complicated and complex aspects is important, a risk is that stakeholders may see their program as operating so much in an open system that they believe it is powerless to affect anything but that which is within their total control. As a result, they may limit their program theory to quite instrumental and sometimes banal achievements (for example, delivery of outputs). Such an approach encourages measurement of only outcomes that are relatively easy to attribute to the program and loss of perspective on the big picture context within which the program operates. The links between what the program does and solving the problem are lost.

Strategies to Avoid This Trap

- Undertake a situation analysis that identifies as many of the important causes as possible—those that the program will and will not directly address (see Chapter Seven).

- Construct an outcomes chain that culminates in impacts that address the main situation that gave rise to the program, even if the program will not be held accountable directly or fully for those ultimate outcomes (see Chapter Seven).

- Scope the program to identify both what lies within and what lies outside the program's boundaries. As part of the scoping, show which items in the full outcomes chain will be the direct and primary responsibility of the program. Scoping is not intended to narrowly define a program in ways that prevent it from seizing opportunities as they arise. The more complex the program is, the more fluid the boundaries may need to be (see Chapter Seven).

- Develop success criteria for each of the results in the outcomes chain—for example, quality, quantity, timeliness, equity, different target audiences and other features, and comparisons. Knowledge of these criteria helps with identifying factors (program and external) that will affect the achievement of various aspects of the outcomes. Different factors will be relevant to achieving different attributes. For example, achieving an outcome with many participants (a success criterion relating to quantity) will require consideration of different factors from those that would need to be considered for achieving an outcome to a particular intensity or in a sustainable manner (a success criterion relating to quality). (See Chapter Eight on how to identify success criteria.)

- Identify both program factors and external factors that are likely to affect each intended outcome and the extent to which the desired features of each outcome are successfully achieved. (In Chapter Eight, we discuss how to identify factors that will affect success.)

Trap 6: Not Using the Program Theory for Evaluation

Although one of the main reasons for developing program theory is to use it in evaluation, one of the common traps is then not actually using it to conduct the evaluation. A program theory might have identified intermediate outcomes, for example, but the evaluation gathers data only about ultimate

impacts, in the same way that a black box evaluation would. More commonly, program theory is used to identify intermediate outcomes, and the evaluation gathers data about these, but the analysis of the data does not go beyond simply reporting whether the outcomes were achieved. One of the most important ingredients of a program theory and its evaluation is not just measuring the building blocks but exploring the relationships among the building blocks.

Strategies to Avoid This Trap

- Ensure that the intended purposes of program theory are clearly articulated and agreed on (see Chapter Four).

- Rather than getting bogged down in endlessly revising the program theory, use an efficient process for articulating it (see Chapter Six).

- Make sure the evaluation uses the program theory to guide data collection (see Chapters Fourteen and Fifteen).

Trap 7: Taking a One-Size-Fits-All Approach

There are two versions of this trap: one at the program level and one at the level of an organization.

For a program, the trap is to develop a program theory, draw a logic model, and then present it as if it can meet all possible needs for all time. It is unlikely that any one model will be able to simultaneously provide an overview of all important aspects and important details. It is much more likely that different versions might be useful, each highlighting a particular aspect—a stage of the program, or how it works in a particular context, or how it is viewed from a particular perspective, for example.

For an organization or an evaluation practice, the trap is to use the same method to develop the program theory, the same way of representing it, and the same way of using it to guide evaluation. This is not to deny some value in streamlining processes and making it easier to develop and use program theory. A balance needs to be found, however, to ensure that the program theories and logic models developed represent the particular features of interventions adequately and suit the particular purposes.

Strategies to Avoid This Trap

- Carefully consider alternative ways of developing and representing program theory. Use different types of representation for different purposes—for example, big picture overviews for communication and program or organizational cohesion and nested program theories for parts of the big picture (see Chapter Nine).

- Periodically review the program theories and logic models in use, and decide whether they need revising (see Chapter Ten).

SUMMARY

The myths and traps identified in this chapter frequently reduce the usefulness of program theory. Remain alert to them, especially when introducing program theory to an organization, and be prepared with strategies for myth busting and trap avoiding.

EXERCISES

1. Are any of the seven myths set out in this chapter alive and well in your organization? If so, is this likely to be a problem? How might you address it?

2. Review an evaluation report or article that has used program theory. Is there evidence of falling into any of the traps discussed in this chapter? If not, are there descriptions of the processes they used to avoid them?

4

Scoping Intended Uses

PURPOSEFUL PROGRAM theory begins by identifying its intended users—who will be using it, how, and when. These have implications for who should be involved in developing program theory, what sources should be used, and how it should be represented.

WHY INTENDED USE MATTERS

Different types of program theory are needed for different uses. As a communication device to explain an intervention broadly to outsiders, it should be brief and clear. As a framework for developing accountability performance indicators, it needs to be much more detailed and show external factors that influence results. In addition, program theory developed for one use might not be suitable for another use.

Considering Intended Users

Early in our careers, we worked on a project to develop performance indicators for reporting to senior management and funders. We worked closely with staff and their supervisors to develop a program theory that they thought clearly represented important aspects of the program. Unfortunately, soon after this, a new manager was appointed who had not been involved in the development of the program theory. She did not understand what had been represented and felt no commitment to revising it to address the issues she had raised. The program theory that the group had developed was set aside and ignored when performance indicators were developed, leading to disengagement and cynicism among the staff.

Those who have contributed to developing a program theory often feel much more committed to it than those who have not been involved. Identifying who will use it, and seeking to engage them in some way, is important for successful use of program theory. Since turnover will always occur, a strategy is needed for briefing and involving new staff, managers, funders, and policymakers.

Processes that can be used to develop program theory with a small group of people who share knowledge and values will not work for a large group of people with different understandings about how an intervention could, should, and does work. In the latter case, time might be needed to negotiate the different perspectives or elaborate the details of intermediate steps and standards of performance that are not commonly understood or accepted. If intended users are numerous, dispersed, or time poor, participatory processes

---

THE IKEA EFFECT IN PROGRAM THEORY

The IKEA effect, according to Wikipedia, is the disproportionate sense of attachment felt for a piece of furniture, such as a bookcase, that one has assembled. This effect is often evident in program theory when those who have been involved in developing a logic model both understand and value it, while other people, who also need to use it, are reluctant to engage with it and are inclined to dismiss it.