5

Program evaluations can be based on *different* causal models to suit different purposes. This chapter discusses several types of causal relationships, three *causal* models derived from systems *theory,* and the role that *causal* models *play* in *program* theory *evaluation.*

Causal Models in Program Theory Evaluation

Patricia *J.* **Rogers**

Causal models are at the heart of program theory evaluation, yet there has been surprisingly little discussion of the different types of causal relationships that might be operating in a causal model nor of the different types of causal models that might be useful for program evaluation (exceptions being Lipsey and Pollard, 1989; McClintock, 1990). This chapter begins by discussing the different types of causal relationships that we might want to represent in our causal models and how we might develop standardized ways to represent these visually. It then briefly explores how nonlinear causal models from systems theory might be used for program theory evaluation. The chapter finishes by discussing the role that causal models play in program theory evaluation. As these various models each highlight different issues that may be important in understanding how outcomes were achieved (or not achieved), it may sometimes be appropriate to use a series of different causal models over several cycles of evaluation or monitoring.

Chen (1990) distinguished between two different types of causal model — normative (how the program is understood to work) and descriptive (how the program actually works). The issues raised in this chapter can be applied to either normative models or descriptive models,

Note: This chapter has benefited substantially from the helpful comments, questions, and suggestions from members of the Harvard Evaluation Task Force, participants at the 1998 American Evaluation Association meeting, the editors and series editor of this volume, and as always my graduate evaluation students. Initial work on this chapter was supported by a fellowship from the Spencer Foundation.

What Do the Boxes and Arrows Represent?

Program theory usually involves a diagram of boxes linked by arrows representing cause-and-effect relationships. It is perhaps tempting to consider these causal models to be like wiring diagrams, in which, if we flick a switch at the first box in the diagram, it will cause the lights in the other boxes to illuminate. And indeed, sometimes the descriptions of these models, using a series of if-then statements, suggest this imagery (Owen, with Rogers, 1999; Plantz, Greenway, and Hendricks, 1997).

Evaluators who are familiar with social science principles will not be surprised that few program theory models are based on simple causal relationships like this, even if diagrams do not explicitly show it. However, some program theory models do explicitly attempt to show the processes that are "necessary and sufficient" to produce the desired results—for example, Cooley's causal model (1997) of a program designed to increase girls' participation in high school in developing countries.

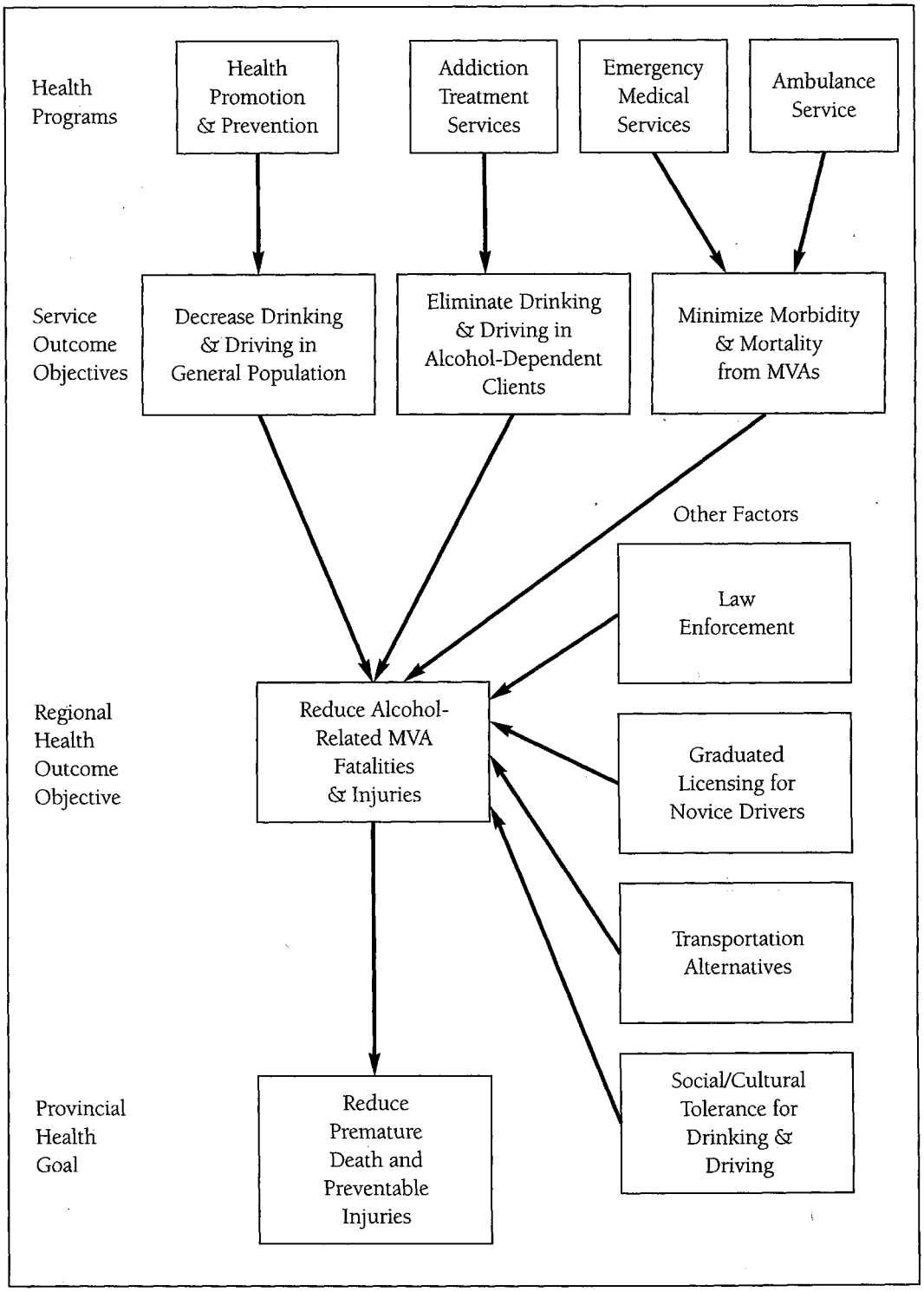
More commonly, program theory models are based on a recognition that other factors may influence the achievement of intermediate and ultimate outcomes. For example, the United Way's generic causal model does not explicitly include other factors, apart from a list of constraints on the program. However, in the instructions provided with this model, it is made clear that the further away from actual program outputs one moves, the weaker the program's influence becomes, and the likelihood of outside forces having an influence increases (Plantz, Greenway, and Hendricks, 1997). They go on to give an example of a program providing prenatal counseling for pregnant teens, pointing out that the program can influence what pregnant teens know about appropriate prenatal practices but cannot influence what the teens' overall health was when they became pregnant. Nor can the programs affect whether teens were using drugs when they became pregnant. The authors recognize that each of these issues, general health and involvement with drugs, can have as much long-term influence on the later health of babies as the program itself.

It is interesting to note that this analysis focuses only on fixed characteristics or events that happen before the client begins in the program—sometimes referred to as *moderators*. Outcomes can also be influenced by factors that occur at the same time as the program and either help or hinder its work.

How can we represent these other factors in our causal models? Funnell's program theory matrix (Chapter Nine) includes other factors explicitly in text associated with each outcome. It is also possible to show them on the program theory diagram, as Halpern (1998, 1999) has done, as seen in Figure 5.1.

We might even make a dramatic move completely away from our program-centric causal model and show the web of client relationships that influence client outcomes, including the influence of family, friends, schools, shops, economy, neighborhood, media, legal system, work, economy, and political system (Bullen, 1995).

Figure 5.1. Representing Other Factors in a Logic Model for Reducing Alcohol-Related Motor Vehicle Accident Injuries and Deaths



Source: Halpern, 1998, 1999.

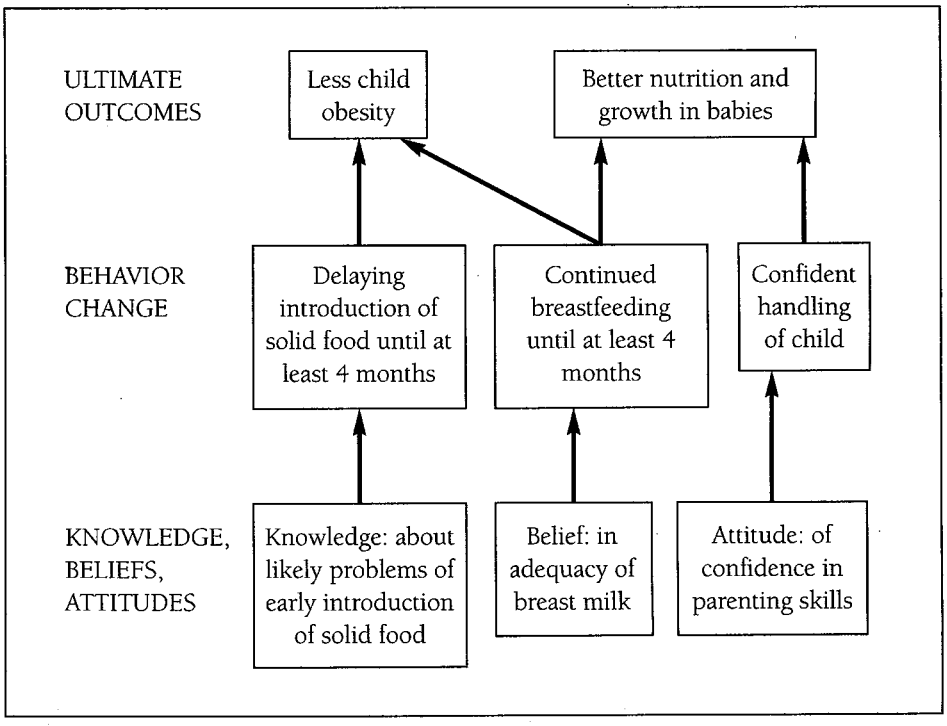
Multiple Strands in Causal Models

Many program theory models portray the program as a single chain of intermediate and ultimate outcomes, where A leads to B and then to C. But it may be helpful to show multiple strands, where A and C both lead to B—either in combination or as alternatives. Ideally, we would be able to distinguish between complementary causal paths and alternative causal paths in a diagram, perhaps by using line arrows for the complementary paths and block arrows for the alternative paths.

If a combination of two causal paths is necessary to achieve the intended results, it is important to make this explicit in order to avoid maximizing only one of them. In many programs, staff must balance competing imperatives like this. When I worked with maternal and child health nurses to develop a causal model of their program to guide the development of performance indicators, they were particularly pleased that they could make visible the balancing they needed to maintain between providing information to parents and supporting parents' confidence in their own abilities. Part of their program model, which used an adaptation of Bennett's hierarchy (Bennett and Rockwell, 1999) to describe their work on infant feeding, showed this clearly, as seen in Figure 5.2.

It was important for the staff to make visible to program managers the competing demands on them and to make sure that performance measures referred to both of these in order to ensure that there were not structural

Figure 5.2. A Partial Program Model Showing Competing Demands



pressures to maximize either information giving or parental support at the expense of the other. When programs are managed by managers without detailed knowledge of program processes or are managed through contractual arrangements, it becomes more important to make explicit competing imperatives such as these. If performance measures only include one of the competing imperatives, then a program may seem to be performing well in terms of its intermediate outcomes because one of these is being maximized at the expense of the other.

Multiple strands in a causal model may instead represent alternative causal paths. For example, Weiss (1998) outlines four possible mechanisms by which higher teacher pay may be linked to increased student achievement. If these are seen as competing explanations for observed outcomes, then a program theory evaluation might focus on testing which of these best explains the evidence (as Weiss discusses in Chapter Four).

It is also possible to see these alternative causal paths as being true for certain people under certain conditions. Drawing an analogy with gunpowder, which will only fire in favorable conditions, Pawson and Tilley (1997) have suggested that program causal mechanisms only fire within favorable contexts. An evaluation based on this type of causal model will try to understand the circumstances under which particular mechanisms operate. In their reanalysis of a crime prevention program in public housing estates, Pawson and Tilley (1997) demonstrated the importance of understanding the context at different sites, including interactions among various mechanisms (such as improved housing and increased tenant involvement in estate management) and among other coexisting processes.

It is difficult to represent these more complex relationships in a two-dimensional diagram. Pawson and Tilley (1997) instead represent their causal model in a matrix of context-mechanism-outcome configurations, which describes in text the causal mechanism that produces the outcome and the context in which the mechanism is operative.

The characteristics of program clients—their motivations, attitudes, previous knowledge, and skills—are an important part of the context within which causal mechanisms work or fail to work. An iterative series of data collection and analysis activities can be used to identify important ways in which clients vary and the implications of these for program effectiveness (McDonald and Rogers, 1999).

To fully understand the context within which causal mechanisms operate, we may need to develop program models that do more than include program clients simply as passive recipients of treatments that change their lives. If the treatment involves swallowing a pill, we might expect certain physiological effects, regardless of the active involvement of the patient, but even in this example, we know that the patient's expectations about the treatment can influence its reported impacts. It is even less realistic to describe program clients as passive recipients when the program is endeavoring to bring about permanent change in, for example, students' school behavior or communi-

cation strategies of the hearing impaired—changes that require program clients to learn, apply, and maintain new ways of operating.

Pawson and Tilley (1997) have argued that we need "to shake off those conceptual habits which allow us to speak of a program 'producing outcomes' and to replace them with an imagery which sees the program offering chances which may (or may not) be triggered into action via the subject's capacity to make choices. . . . Potential subjects will consider a program (or not), volunteer for it (or not), cooperate closely (or not), stay the course (or not), learn lessons (or not), retain the lessons (or not), apply the lessons (or not)" (p. 38).

This issue does not need to be associated with a philosophical commitment to serving program clients and having their needs and perspectives at the forefront of program planning and evaluation nor with a belief that the personal dignity of clients and staff requires treating them as program partners rather than as passive objects. In fact, the same distinction holds true for programs such as burglary prevention, which intend to change the behavior of potential burglars. Programs can be understood as changing the options available to participants and their capacities to choose and enact these choices. Usually programs seek to increase options and capacities; some, such as burglary prevention, seek to reduce them.

Causal Models from Systems Theory

Systems theory suggests other types of causal models. In this section, I discuss three of these that appear to be potentially useful for program evaluation—virtuous or vicious circles, symptomatic solutions, and feedback delays.

Virtuous or Vicious Circles. Systems thinking suggests that cause and effect might often be connected not in a linear way but in a circular way, through a series of *virtuous circles* (where an initial effect leads to its own reinforcement and magnification) or *vicious circles* (where an initial negative effect is similarly magnified).

Funnell (1997) has discussed how process feedback loops may be important in a program model to show, for example, how attitudes may affect behavior, which may then affect attitudes further through processes of self-attribution. Other feedback loops may exist between the achievement of positive results from a program and its ability to attract clients. Batterham, Dunt, and Disler (1996), in their discussion of the evaluation of rehabilitation programs, discussed a similar process of iteration until a critical level of rehabilitation is reached (for example, independent mobility), at which point the client is able to move onto a further stage.

In terms of achieving ultimate outcomes, the presence of iterative causal mechanisms may not be important, providing that both the program and the evaluation continue for long enough for these effects to take place and be measured. Problems arise, however, when the program is not continued for long enough for the subsequent cycles of improvement to occur

or when the evaluation is terminated before the ultimate outcomes are demonstrated. If a program sets in place an iterative virtuous circle, then it is possible that initially small gains may ultimately become significant. It is worth considering whether the outcomes of the program are likely to decay over time or to become stronger.

Symptomatic Solutions. *Symptomatic solutions* are solutions that relieve the symptoms but that actually make it harder to solve the problem. It would be like having the flu and taking tablets to reduce the symptoms and then continuing to work excessively, rather than convalescing, thereby making it harder to actually recover.

This problem has implications both for evaluation and for monitoring. For an evaluation, where we are trying to understand how effective a particular program has been in solving a problem, we should design our evaluation so that it can distinguish between temporary reduction of symptoms and sustainable solving of the problem. For monitoring, where we are seeking to simultaneously understand and influence program implementation, we should set up systems that do not encourage people to develop dysfunctional symptomatic solutions.

Owen and Lambert (1995) addressed this issue in their evaluation of a program in which all grade-five students used their individual notebook computers in all subjects. One of the unintended consequences of this program was an increase in teacher stress, as teachers struggled to develop their computer skills and simultaneously adapt their teaching material and processes. Initially, teachers responded to this increased stress by "getting on with the job," avoiding spending time in coordination meetings, or liaising with other teachers, and the administration sought to support teachers by leaving them alone and not making additional calls on their time. If the evaluation had measured teacher stress at this point only, it would have found that in the short-term teacher stress was reduced through this coping mechanism. But over time, this symptomatic solution led to hoarding of equipment, rivalry among groups, and poor attendance at information sessions—consequences that made it harder to implement the fundamental solution, which involved better sharing and coordination of resources and increased support and training for teachers.

Feedback Delays. We have probably all experienced the effects of *feedback delay* when trying to adjust the water temperature in a shower. If there is a delay in response, we tend to overcorrect—first too hot, then too cold, until eventually reaching the desired equilibrium. The Massachusetts Institute of Technology "beer game" simulation (Senge, 1990) has demonstrated the effects of feedback delay on a simple program—a system for producing and distributing a single brand of beer. Once there is a delay built into the system, so that the decision makers do not immediately see the impact of the changes they are making, the orders become more and more excessive and unbalanced.

The reason for using performance measures and indicators as part of ongoing evaluation in the public sector is that they can be used by managers to take corrective action in programs, much like someone monitoring and

adjusting the shower temperature. Unfortunately, few if any of these systems address the problem of feedback delay. In fact, I have been unable to find a single example that does.

How Complex Should Program Theory Models Be?

Although this chapter has focused on more complex models and causal relationships, it is worth remembering that simple models can often be helpful, particularly in programs in which there have previously been few explicit conceptual and empirical connections made between program activities and outcomes. Building a plausible model of how the program is meant to work helps managers identify the most important processes or intermediate outcomes and focus their measurement and attention there. Given that many program evaluations still collect little data about program implementation or intermediate outcomes, there is often considerable value (as Lipsey, 1993; and Petrosino, Chapter Six, point out) in using even a simple two-step program model that simply identifies and measures one mediating variable that is understood to be necessary for the achievement of the ultimate outcomes. And having a common model of how the program is meant to work can help program staff work together and focus on those activities that are most important for program success.

In fact, as Weick (1995) argues, a model might provide a useful heuristic for purposeful action without necessarily being correct. He recounts the story of the reconnaissance unit, lost in the snow in the Swiss Alps for three days in a blizzard, who eventually managed to find their way safely back to camp with the help of a map—a map, they later discovered, of the Pyrenees, not of the Alps. "This incident raises the intriguing possibility that when you are lost, any old map will do. . . . Once people begin to act, . . . they generate tangible outcomes . . . in some context . . . and this helps them discover . . . what is occurring, . . . what needs to be explained, . . . and what should be done next" (pp. 54–55). Weick goes on to quote Sutcliffe, "Having an accurate environmental map may be less important than having some map that brings order to the world and prompts action" (pp. 56–57).

This analysis may well explain the positive responses that program staff often have to program theory evaluation (see, for example, Huebner, Chapter Eight), even when this is based on very simple causal models. But simple causal models can be dysfunctional in high-stakes evaluations that are linked to organizational incentive structures such as performance-pay or to summative evaluations that inform decisions about ending or continuing programs.

Conclusion

This chapter has outlined a wider repertoire of causal models for evaluators to use to guide their evaluations. In using these models, we should remember that *any* causal model is indeed only a model—a simplification of real-

ity to help us understand, predict, make decisions, and act. Rather than searching for the true causal model that underpins a program, evaluators might understand a program more by using a variety of causal models.

References

Batterham, R., Dunt, D., and Disler, P. "Can We Achieve Accountability for Long-Term Outcomes?" Archives of Physical Medicine and *Rehabilitation*, 1996, 77, 1219–1225.

Bennett, C. F., and Rockwell, K. "Targeting Outcomes of Programs (TOP): A Hierarchy for Targeting Outcomes and Evaluating Their Achievement." [http://deal.unl.edu/TOP/synopsis.html]. 1999.

Bullen, P. "Evaluating Human Services: Complexity—Uncertainty—Self-Delusion—Rigour." Evaluation News and Comment, 1995, 3(2), 3–10. [available at http://www.mapl.com.au/A1.htm].

Chen, H. T. *Theory-Driven* Evaluation. Thousand Oaks, Calif.: Sage, 1990.

Cooley, L. Presentation to the Washington Evaluators' Conference, 1997.

Funnell, S. "Program Logic: An Adaptable Tool." Evaluation News and Comment, 1997, 6(1), 5–17.

Halpern, G. "From Hubris to Reality: Evaluating Innovative Programs in Public Institutions." *Innovation Journal*: The *Public* Sector *Innovation Journal*, Nov. 1998, revised Nov. 1999. [available at http://www.innovationcc/rev_arts/halp4.htm].

Lipsey, M. "Theory as Method: Small Theories of Treatments." In L. Sechrest and **A.** Scott (eds.), Understanding Causes and *Generalizing* ***About*** Them. New Directions **€or** Program Evaluation, no, 57. San Francisco: Jossey-Bass, 1993.

Lipsey M., and Pollard, J. "Driving Toward Theory in Program Evaluation: More Models to Choose From." Evaluation and Program Planning, 1989, 12, 317–328.

McClintock, C. "Administrators as Applied Theorists." In L. Bickman (ed.), New Directions for Evaluation: Advances in Program Theory. New Directions for Program Evaluation, no. 47. San Francisco: Jossey-Bass, 1990.

McDonald, B., and Rogers, P. "Market Segmentation as an Analogy for Differentiated Program Theory: An Example from the Dairy Industry." Paper presented at the annual meeting of the American Evaluation Association, Orlando, Fla., 1999.

Owen, J. M., with Rogers, P. J. Program Evaluation: Forms and Approaches. Thousand **Oaks,** Calif.: Sage, 1999.

Owen, J. M., and Lambert, F. C. "Roles for Evaluation in learning Organizations." *Evaluation*, 1995, *1*(2), 237–250.

Pawson, R., and Tilley, N. Realistic Evaluation. Thousand Oaks, Calif.: Sage, 1997.

Plantz, M. C., Greenway, M. T., and Hendricks, **M.** "Outcome Measurement: Showing Results in the Nonprofit Sector." In K. E. Newcomer (ed.), Using *Performance* Mea***surement*** to Improve Public and Nonprofit Programs. New Directions for Evaluation, no 75. San Francisco: Jossey-Bass, 1997.

Senge, P. M. *The* Fifth Discipline: The ***Art*** *and* Practice *of the Learning Organization.* New York: Doubleday, 1990.

Weick, K. E. Sensemaking in *Organizations.* Thousand Oaks, Calif.: Sage, 1995.

Weiss, C. H. Evaluation: Methods for Studying *Programs and* Policies. (2nd ed.) Englewood Cliffs, N.J.: Prentice Hall.

PATRICIA J. ROGERS **is** director *of* the Program for Public Sector Evaluation in the Faculty *of Applied* Science, *Royal Melbourne* Institute of Technology, *Australia.*