Hyunjung Kim
California State University, Chico
Roderick H. MacDonald
David F. Andersen
University at Albany

Simulation and Managerial Decision Making: A Double-Loop Learning Framework

Systems approaches present opportunities for public managers and policy makers to view policies and programs in a broader context. This article presents a framework to explain how simulation modeling promotes double-loop learning in management teams by building and exploring collective mental models as well as by enhancing accuracy of the mental models. The authors discuss the types of problems that may benefit from simulation modeling and illustrate how double-loop learning occurs in the process of dynamic hypothesis testing. Using a case from New York State's Division of Disability Determination, the article shows how simulation modeling built confidence in a management team's decision by providing the team with tools to share and examine multiple hypotheses about a declining trend in initial disability recipients in the state between 1998 and 2004.

On a day-to-day basis, public administrators manage complex cross-agency and interjurisdictional problems, making decisions that blend knowledge of the hard facts with deep intuitions gained from years of experience. On the other hand, researchers who investigate these complex decision-making behaviors—or, perhaps more precisely, those researchers who describe research methods to support these decision tasks—can maintain a focus either on positivist methods that rely on statistics and quantifiable "hard" data or on interpretive methods that use verbal arguments and "soft" qualitative data. However, the advancement of information and computing techniques in recent decades has given rise to a new wave of methods that can blend quantitative and qualitative approaches, bridging the artificial positivist-interpretive gap. Johnston and Kim (2011) dub this loose collection of methods "policy informatics," and one of the core methods of this new cluster of policy informatics is system dynamics modeling.

This article explains how system dynamics modeling can enhance the daily business of complex decision making, especially that of public administrators, by supporting double-loop learning.

> Double-loop learning is what allows organizations to be proactive or generative in their decision making.

Since its inception more than 50 years ago by Jay W. Forrester (1961), system dynamics modeling has been used in the public sector as a tool to apply systems theory (Easton 1965; Wiener 1948) in decision making (Andersen, Rich, and MacDonald 2009). Its application areas include welfare reform (Zagonel et al. 2004), health care (Hirsch et al. 2010; Lane, Monefeldt, and Rosenhead 2000), the policy-making process (Stave 2002), and defense and homeland security (Bakken and Gilljam 2003; Coyle 1996). Since the early 1990s, a number of researchers have been exploring ways to develop system dynamics models with public managers in face-to-face, facilitated problem-solving sessions. Schuman and Rohrbaugh (1991) and Reagan-Cirincione at al. (1991) first introduced ways to build system dynamics models with teams of public managers and a skilled facilitator. This early work gave rise to a subfield of system dynamics called *group model building* (Andersen and Richardson 1997; Luna-Reyes et al. 2006; Richardson and Andersen 1995; Vennix 1996). More recent work in group model building explores how the method interacts with other strategic planning and problem-solving methods (Ackermann et al. 2011; Eden et al. 2009), and how these approaches can be taught more effectively in professional degree programs (Andersen et al. 2006).

This article posits that systems theory and its associated simulation methods are well suited to support "double-loop learning" (Argyris and Schön 1996) for public managers, especially when the problems confronted are characterized by "dynamic complexity," as explained in the following section. Single-loop learning occurs when a management team modifies its strategy or action based on results from its previous actions. Double-loop learning goes further and involves modification of the management team's mental model or the theory that underlies the action. Double-loop learning is what allows organizations to be proactive or generative in their decision making.

Hyunjung Kim is assistant professor in the Department of Management at California State University, Chico. She received her doctorate in public administration from the University at Albany. She uses ethnographic methods and causal mapping methods to represent and understand the social process of aligning different mental models of decision makers. She also uses systems thinking and system dynamics modeling to enhance decision making in various management groups.
E-mail: hkim18@csuchico.edu

Roderick H. MacDonald is director of the Initiative for System Dynamics in the Public Sector at the Rockefeller College of Public Affairs and Policy, University at Albany. He earned his doctorate in public administration, focusing on the development of system dynamics simulation models as decision support tools. He has used system dynamics to address problems in various policy areas, including the delivery of mental health services, Social Security disability programs, DWI recidivism, and traffic safety.
E-mail: rod@isdps.org

David F. Andersen is Distinguished Service Professor of Public Administration, Public Policy, and Information Science at the University at Albany. He has served as a technical consultant to public and nonprofit agencies in the federal, state, and local sectors, as well as corporate clients in North America and Europe. He is coauthor of *Introduction to Computer Simulation: The System Dynamics Modeling Approach* and *Government Information Management.* He holds a doctorate in management from MIT's Sloan School.
E-mail: david.andersen@albany.edu

How does simulation modeling support double-loop learning in management teams? Sterman (1994) argues that experimenting with the virtual world enhances double-loop learning by reducing various learning impediments such as bounded rationality, system complexity, and information delays. Building on Sterman's seminal work, this article presents a framework to explain how the *process* of causal theory building and dynamic hypothesis testing in system dynamics aids double-loop learning in two seemingly contrasting epistemological paradigms. Building on studies on collective mental models (Cannon-Bowers, Salas, and Converse 1993; Kim 2009; Mohammed, Ferzandi, and Hamilton 2010), the framework assumes that collaborative decision making requires emergence, alignment, and modification of collective mental models. The goal of this article is to help public administrators understand the role of simulation modeling and manage their expectations as they engage in the modeling sessions. Furthermore, it seeks to encourage active participation of public managers in the modeling process, which is often flooded with technical jargon, computer software, and mathematical equations unfamiliar to nonmodelers, as this process can lead to double-loop learning in management teams.

The article is organized as follows: The first part uses the case of disability determination in New York State to explain the types of problems that may benefit from the simulation-based systems approach. This section discusses what "dynamic complexity" is and why it poses a special challenge for public managers. The second part will introduce the framework to discuss the role of system dynamics modeling and its contribution to double-loop learning in both positivistic and interpretive ways. The third section will describe the process of building and testing dynamic hypothesis in system dynamics, and the rest of the article will return to the case of disability determination to illustrate how the modeling process aided managerial decision making and double-loop learning.

Dynamic Complexity of Disability Determination in New York State

In 2003, the New York State Division of Disability Determination (DDD), within the Office of Temporary and Disability Assistance, faced a puzzling development that had been under way for several years. Since 1998, there had been a decline in initial disability recipients in New York, and this trend was running against national and regional trends. During that time, the DDD had changed from a demand environment to a planned environment in order to be more responsive to workload changes. As part of this process, the profiles of employees making disability determinations shifted to a more professional workforce with additional responsibilities and higher pay. The recent changes had led to high morale and improved productivity, but if the workload remained low, layoffs would be required and the lost positions would not be regained.

Statistical analysis, a traditional positivist method, was performed by outside experts to determine the precise magnitude of the downward trend and whether the trend would continue. The analysis indicated that the trend would continue for the next year, but when asked what the primary drivers of the trend were, the statistical

analyses could not provide an answer. Without an answer to the second question, the agency's ability to manage its workforce and work processes to deliver services in an efficient and effective manner was limited. The agency managers needed interpretive tools to move beyond a statistical description of what was happening, to understand why the decline was occurring, and to use this understanding to justify why a particular set of policies was to be implemented. The agency managers could come up with a number of possible explanations for the observed trend, but many of these explanations were outside the agency's control. What managers really wanted was a systematic way to elicit different explanations of the decline, even if those explanations were outside the perceived boundary of the agency, along with ways to test which policies could reverse or stabilize the declining trends and avoid layoffs. In other words, management was seeking a way to change how it was thinking about the problem in order to avoid what seemed like an almost inevitable layoff situation.

The decision environment surrounding the DDD was "dynamically complex." The complexity was created by multiple explanations—or hypotheses—for the observed decline in the initial disability recipients. The plausibility of these explanations competed with one another, and the complexity was intensified by possible interactions and feedback relationships between these explanations. The complexity was dynamic in the sense that each of the competing explanations was associated with different future growth patterns of initial disability recipients. While one hypothesis implied a lower equilibrium level of the number of initial disability recipients in New York State, another hypothesis suggested that the number could bounce back in the future. Yet another hypothesis implied that the number of initial disability recipients would oscillate over time. The difficulty of decision making was intensified as the expected future growth pattern could take place over different time spans.

The situation facing the management team at the DDD is an example of a very general set of problems often encountered by managers and policy makers (Lindblom 1959; Rittel and Webber 1973). When decision makers face dynamic complexity, as in the case of the DDD, system dynamics can be a powerful tool to support decision making. As will be illustrated in the following sections, system dynamics modeling can provide an opportunity for managers to move beyond understanding simply *what* is happening in their system to probe *why* changes are occurring over time. This change in mind-set enables double-loop learning, leading to more effective policies.

> System dynamics modeling can provide an opportunity for managers to move beyond understanding simply *what* is happening in their system to probe *why* changes are occurring over time.

Double-Loop Learning Framework for System Dynamics Modeling

A system is composed of physical and institutional structures and agents acting within the structures (Sterman 2000). The interaction of these structures and agents creates feedback loops, delays, accumulations, and nonlinearities that are responsible for various behaviors of the system. System dynamics modelers are interested in identifying the system structure that explains the observed system behaviors or creates the desired behaviors.

Public managers are decision-making agents in their policy system. The boundary of this system is defined by the policy or the problems that the managers are dealing with. As system agents, individual managers have their own understanding of the system structure that they manage. In the modeling community, this is referred to as a *mental model* of the system. The concept of mental model may be related to Tolman's (1948) much earlier concept of cognitive maps and certainly bears a relationship to Axelrod's (1976) and Eden's (1988) similarly named concepts. Johnson-Laird (1983), a cognitive psychologist, defines a mental model as a processor that translates external information into internal symbols and retranslates such symbols into external actions. Senge describes mental models as "deeply ingrained assumptions, generalizations, or even pictures or images that influences how we understand the world and how we take action" (1990, 8). Mental models guide what information we choose to observe, how we interpret the observed information, and how we make sense of the information to make a decision. In system dynamics, Richardson et al. (1994) and Doyle and Ford (1998) provide more in-depth discussion of a number of the distinct components of mental models as the term is used in the modeling practice.

Because managerial decisions are generated from mental models, mental models are viewed as the main leverage point for enhancing managerial decision making. When faced with dynamic complexity, a manager generates a policy action by closely monitoring the policy system. There may be day-to-day operational information, administrative reports on how the system is functioning, or anecdotal speculation about what could be going on. The observation is interpreted by the mental model, and the mental model identifies an appropriate policy action. Single-loop learning occurs when such observations of the policy system, especially the outcomes of previous policy actions, lead to changes in the policy action. When double-loop learning occurs, observations of the policy system have a more profound effect, leading to a modification of mental models or the "theory-in-use" (Argyris and Schön 1996). The top portion of figure 1 describes this typical double-loop learning process, in which feedback from direct action contributes to the modification of mental models.

Joining Mental Models at the Group Level
When decision making takes place at the group level, a team of managers can be thought of as having a collective mental model of the system. Such a collective mental model is a shared perception or interpretation of the policy system, and it may be quite different from the sum of individual mental models, for a number of reasons. First, individual models are based only on that individual's background and expertise, and thus they are likely to be incomplete and selective. Furthermore, members of a management team may hold individual mental models that are mutually inconsistent and even contradictory. As a result, social and political factors such as power dynamics in a decision group can influence how individual mental models are shared and integrated. The mental model concept at the group level has been explored by researchers in cognitive psychology, organizational behavior, and decision sciences (Klimoski
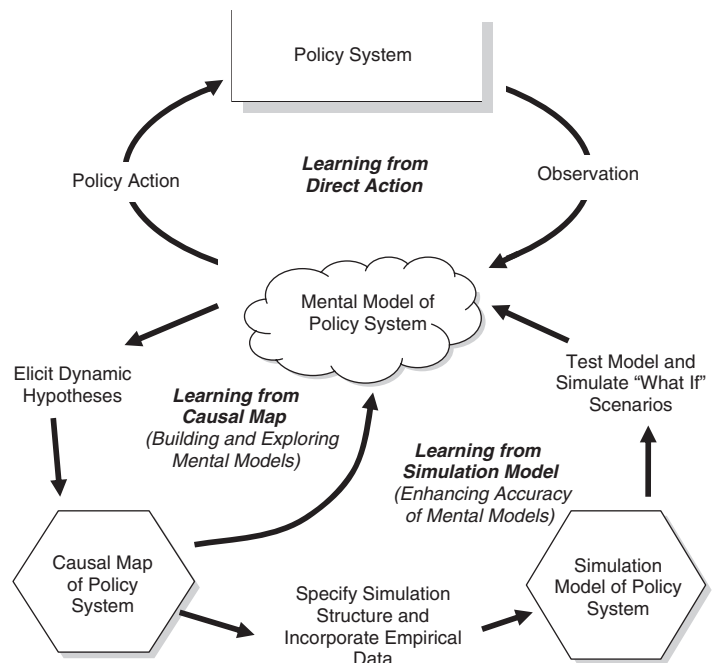
Figure 1 Double-Loop Learning Supported by Simulation-Based Dynamic Hypothesis Testing

and Mohammed 1994; Mohammed, Ferzandi, and Hamilton 2010; Schneider and Angelmar 1993; Walsh 1995). Kim (2009) finds that different names have been used to describe this mental model concept at the group level, and the subtle differences in their definitions and underlying assumptions emphasize different aspects of the group decision-making process. What is common, though, is that these theorists posit the existence of a mental model at the group level that empowers a group of managers and policy makers to take collective action in the face of complex situations.

For the current discussion, we focus on double-loop learning that can occur by joining individual mental models and creating and modifying collective mental models. The process is different from double-loop learning at the individual level, and Senge (1990) proposes that the ability to align the mental models of individual managers is one of the five core competencies of a learning organization. As shown in the lower portion of figure 1, our framework highlights that the *process* of building system dynamics simulation models (and, by extension, similar management science tools) is an effective way to promote double-loop learning in a decision-making group. This is achieved by (1) building and exploring collective mental models and (2) enhancing the accuracy of collective mental models. In the following section, we describe how these two approaches differ in terms of double-loop learning, especially in regard to their epistemological orientation.

> Our framework highlights that the *process* of building system dynamics simulation models (and, by extension, similar management science tools) is an effective way to promote double-loop learning in a decision-making group.

Positivist Tools for Enhancing Mental Model Accuracy
Managers often use decision support tools based on a positivist orientation to the world that assumes the existence of objective

social systems. Mental models are regarded as a subjective perception of the system that must be rejected when the perception deviates from the objective reality. By closely observing data collected from the policy system and by providing tools for systematically analyzing the data with an emphasis on minimizing subjectivity, this group of decision tools attempts to align mental models as closely as possible with the real system. Examples of such methods would include econometric models and optimization models that rely heavily on numerical and time series data. As mainstream double-loop learning instruments, these positivist decision tools have contributed extensively to public sector decision making. However, these tools are often not flexible enough to capture complex, dynamic patterns (McCaffrey et al. 1985), and they block out rich sources of qualitative information about our policy systems (Forrester 1994). Furthermore, these tools tend to use data from the past, whereas policy decisions are about the future.

Interpretive Tools for Building and Exploring Mental Models

Another set of decision tools comes from the interpretive view of policy systems. Mental models are regarded as constructed realities that have important attributes of their own—they are meaning systems that drive policy decisions and managerial actions. Policy systems have different meanings to different people, and interpretivists seek to explore different meanings assigned to the system (Berger and Luckmann 1966; Weick 1995). Double-loop learning from the interpretive perspective involves discovering and sharing these meaning systems, and its tools to support decision making provide ways to extract and describe mental models that are otherwise implicit and hard to observe. Interpretive tools are also useful for building and shaping collective mental models.

The need for such interpretive methods emerged largely in the field of strategic planning (Bryson and Roering 1988). Because different stakeholders and decision makers rarely agree on organizational missions, goals, values, or the best way to achieve the goals, exploring individual differences and aligning mental models became an important task for organizational success (Berry 2007; Miesing and Andersen 1991). Furthermore, public involvement (Thomas 1993) and cross-sector collaborations (Bryson, Crosby, and Stone 2006) in governmental decision-making processes called for an exploration of mental models as a way to enhance the experience and its outcome.

Senge (1990) postulates that learning organizations can improve managerial decisions by enhancing the mastery and alignment of mental models through the use of mapping tools. These tools are mainly used to build and explore mental models (Axelrod 1976; Bryson and Anderson 2000; Eden, Ackermann, and Cropper 1992; Huff 1990), but during the process, they frequently contribute to the development of new collective mental models.

Decision support tools from the interpretive paradigm also have limitations: interpretive data are robust and rich but can be biased and lack empirical accuracy. Relying solely on inferred mental models for decision making is a slow and ineffective way of organizational learning (Sterman 1994). In practice, effective managerial

> System dynamics modeling explicitly integrates both empirical and interpretive data in order to enhance the accuracy of mental models, as well as to build and align a management team's collective mental models.

decisions require empirical data collected from the policy system as well as interpretive data elicited from decision makers' mental models. Neither provides a complete picture of the policy system for generating the best policy action. As we present in the following sections with the case of the Division of Disability Determination in New York State, system dynamics simulation modeling is one way to incorporate both positivistic and interpretive approaches in decision making (Lane 2001; Yin 2009). System dynamics modeling explicitly integrates both empirical and interpretive data in order to enhance the accuracy of mental models, as well as to build and align a management team's collective mental models.

Simulation-Based Dynamic Hypothesis Testing

In the traditional parlance of the scientific method, a hypothesis presents a testable statement of how one or more measurable variables will behave under a clearly specified causal circumstance. It is a mental model that generates the causal statements to be tested against empirical data obtained from a policy system. If the data do not support the expected behavior generated by the causal assertion, then the hypothesis is rejected and the mental model is modified. As a result, the accuracy of the mental model is expected to improve. However, with simulation-based hypothesis testing, hypotheses are tested against traditionally collected data as well as the interpretive data generated from mental models.

In system dynamics, hypotheses are called *dynamic hypotheses.* Dynamic hypotheses make "statements of system structure that appear to have the potential to generate the problem behavior" (Richardson and Pugh 1989, 55). The lower portion of figure 1 summarizes the iterative process of dynamic hypothesis testing, emphasizing how it promotes double-loop learning. Next we describe each step of the dynamic hypothesis testing process.

Elicit dynamic hypotheses. The first step in dynamic hypothesis testing involves a management team exploring their views of how the policy system under study actually functions. They come up with structural explanations for various system behaviors, often in the form of "hunches" or guesses (Ackermann and Eden 2005). Such hunches emerge from the mental models of individual managers as they interact in social settings, and, with discussion, they evolve into more concrete dynamic hypotheses.

There are different techniques for eliciting dynamic hypothesis. In system dynamics, group model building (GMB) is frequently used to explore the management team's collective mental model. Interviews with key stakeholders or content analysis of the data collected from management team's decision-making processes are other examples. Luna-Reyes and Andersen (2003) review a broad array of qualitative research methods that can be used to elicit dynamic hypotheses.

Draw a causal map of the policy system. Dynamic hypotheses are then translated into a causal map to represent the structure of the policy system that is collectively perceived by the management team. In system dynamics, two types of causal maps can be used:

causal-loop diagrams and stock-and-flow diagrams. The former is used to capture feedback structures in the system, and the latter is used to capture accumulations and delays in the system. The process of creating a map helps the management team externalize their individually held mental models. When differences among the mental models become more explicit, the team members start to modify their individual mental models in order to create a collective mental model. As the management team builds and explores collective mental models, double-loop learning occurs.

Specify simulation structure and incorporate empirical data. The next step involves formalizing the causal map of the system so that the map can be integrated with empirical data. The loosely sketched causal structures are transformed into mathematical equations and nonlinear functions. Empirical data collected from the policy system is used to calibrate the equations and functions that describe the structure of the system. The techniques and methods for transforming qualitative maps into logically consistent simulation equations have been developed and documented in the system dynamics literature (Richardson and Pugh 1989; Sterman 2000).

Build a simulation model of the policy system. With help of computer simulation software, the modeler puts together the simulation structure developed in the previous step. Although building a simulation model requires technical knowledge and skills, the management team needs to communicate with the modeler on regular basis so that the team understands how the simulation model reflects the team's collective mental model.

Test the model and simulate "what if" scenarios. Once the simulation model is built, the model must be tested. There are various model validation techniques that can be used at this stage (Barlas 1996; Forrester and Senge 1980). Sterman (2000) lists 12 ways to assess simulation model structure, including the extreme condition test, sensitivity analysis, and time series replication test. When the policy system is appropriately represented with the computer simulation model, the management team can start testing various "what if" policy scenarios in the virtual world (Zagonel et al. 2004).

Update the mental model of the policy system. In the reiterative process of model validation, the computer simulation model is modified. If there is a discrepancy between the model-generated behavior and the system's actual behavior, the management team must check its assumptions as captured and embedded in the simulation model and explore the source of the discrepancy. The modification of the simulation model frequently requires updating the management team's mental model. This process continues until the mental model, the simulation model, and the policy system are all well aligned. As the management team enhances the accuracy of their mental models using the simulation model, double-loop learning continues to occur.

> As the management team enhances the accuracy of their mental models using the simulation model, double-loop learning continues to occur.

Simulation Modeling in the Division of Disability Determination

The following section describes how the steps of dynamic hypothesis testing took place in the New York State Division of Disability Determination (DDD). First, the group model-building techniques

(Andersen and Richardson 1997), including the graph-over-time exercise, group interviews, and facilitation, were used to elicit from the management team five dynamic hypotheses associated with the observed decline in initial disability recipients between 1998 and 2004. They can be summarized as follows:

1. Decline in the proactive outreach of the Social Security Administration (SSA): To keep up with the national productivity goal introduced in the late 1990s, the SSA's field offices in New York City shifted their resources away from proactive outreach efforts to activities more closely related to the agency's productivity goals.
2. Welfare reform: The Personal Responsibility and Work Opportunity Reconciliation Act of 1996 removed the state's incentive to move people from welfare to disability benefits.
3. Decline in manufacturing jobs: The number of manufacturing jobs in New York State has been declining for decades. Approximately 400,000 manufacturing jobs have been lost since 1990. People in manufacturing jobs apply for disability at a higher rate than workers in other industries.
4. Influence of other applicants: Potential applicants are likely to apply for benefits if they know other applicants who have obtained benefits successfully. This word-of-mouth phenomenon can work in the opposite direction if the number of disability benefit recipients is declining.
5. Market saturation: The potential population eligible for disability benefit is limited, and, over time, the pool of people remaining who have not filed claims becomes smaller. The smaller the pool of remaining potential recipients, the harder it will be to get these people to apply.

Around these hypotheses, the management team's perception of the system was explored and represented in causal maps. In this phase of the work, many detailed sketches were drawn, discussed, and refined by the DDD's management team. Figure 2A presents a high-level summary of multiple maps generated in this phase. This causal mapping process promoted double-loop learning because it required the management team to articulate and represent divergent mental models held by individual members and then to align and reconcile the individual maps in order to create a collective mental model.

This high-level systems view in figure 2A indicates that the elicited hypotheses are not independent of one another: they are interrelated parts of one system. Furthermore, these interconnections create balancing loops (B1 ~ B4), where each balancing loop represents a self-correcting mechanism. A change in any of the variables in the loop would be restored to an equilibrium state by the feedback structure.

The *DDD Staffing Loop* (B1) represents the main problem of the case. Staffing levels within the DDD are adjusted to meet the need to process initial applications for initial disability claims. In order to plan and manage staffing, the DDD needed to determine whether the observed decline in the number of initial claims was permanent or temporary. Not shown in loop B1, but detailed in other parts of the model, are pieces of the system structure that would allow the DDD staff to take action to adjust either their
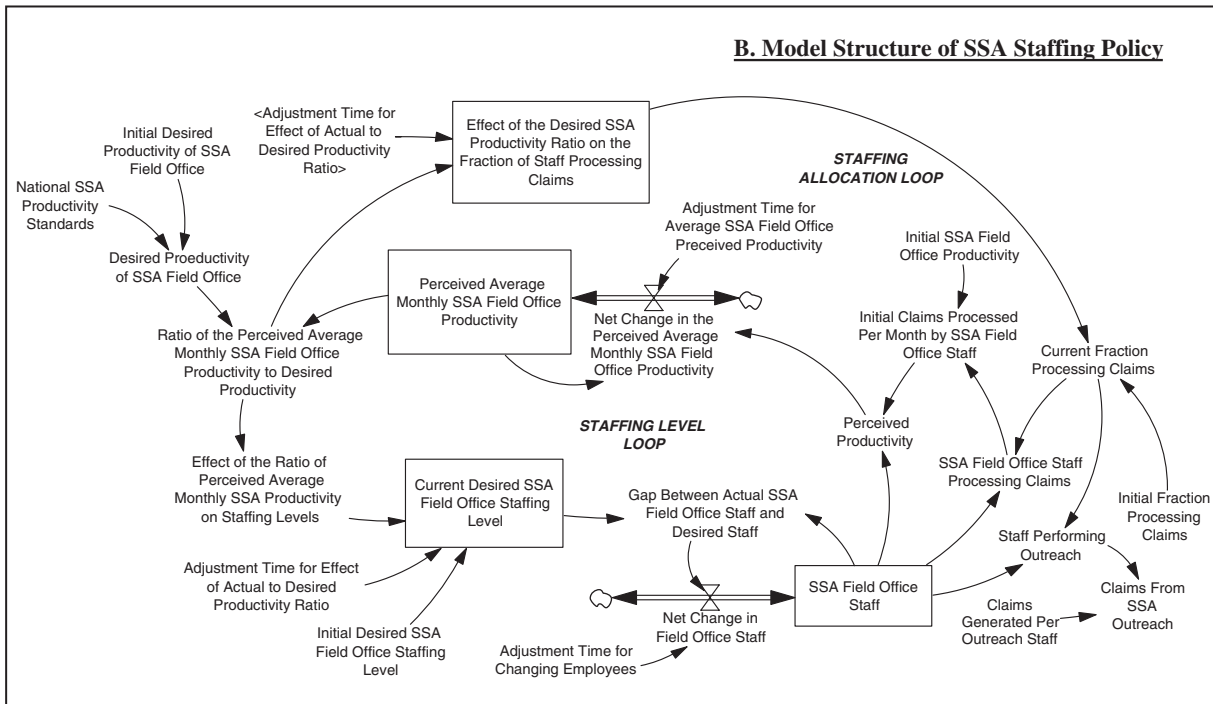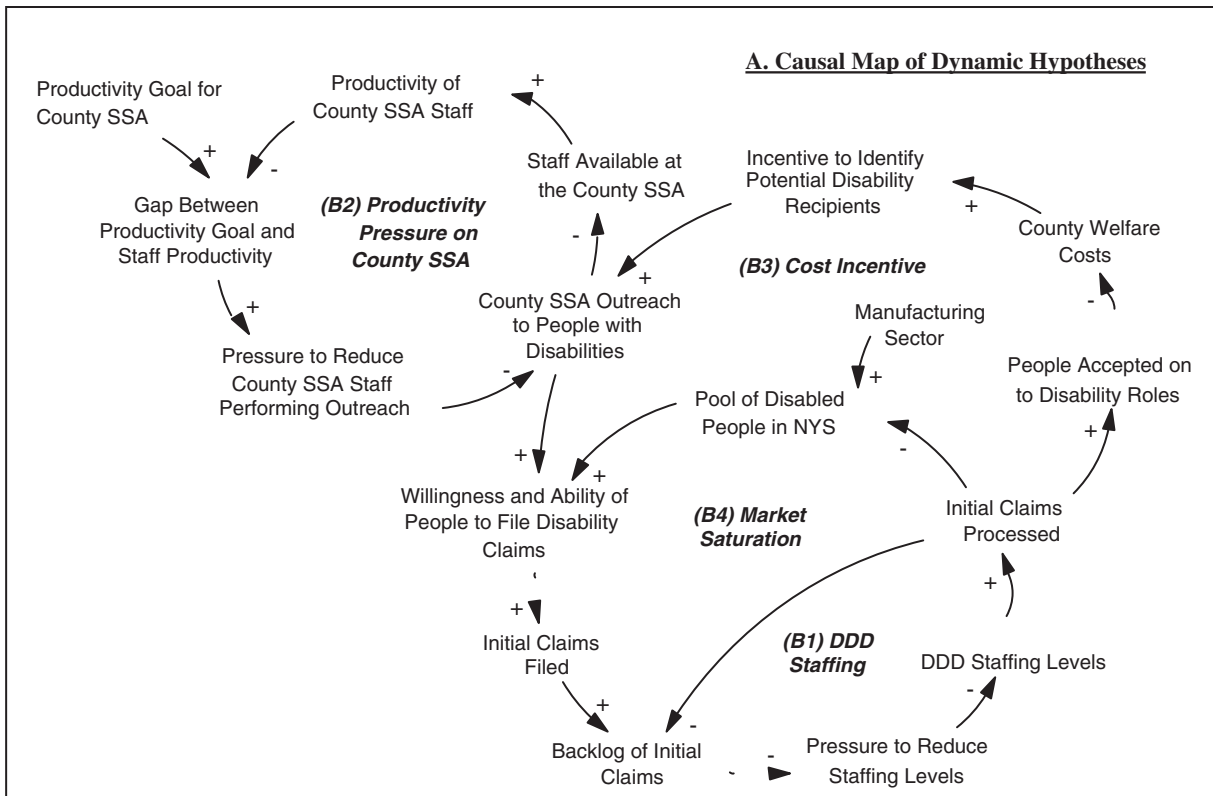
Figure 2 A Partial View of the Causal Map and the Model Structure Linking Dynamics Hypotheses
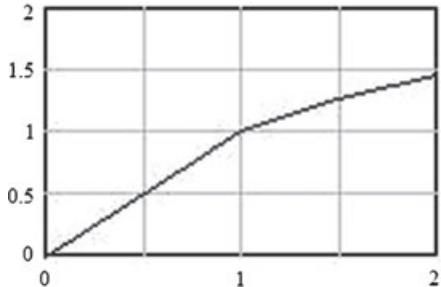
staffing levels or the number of clients they serve in order to close any gap between staff available and clients needing to be processed. The question of which to adjust—staffing levels or clients served—can only be answered by taking a holistic approach.

The *Productivity Pressure on SSA Staff Loop* (B2) describes the structure underlying the first dynamic hypothesis of declining SSA

proactive outreach. The loop says that if there is a gap between the productivity goal and staff productivity, there is pressure to close the gap by shifting staff away from proactive outreach activities to activities directly related to productivity measures. The *Cost Incentive Loop* (B3) presents a high-level view of the second dynamic hypothesis of welfare reform, illustrating how managers in county welfare systems react to cost incentives to identify welfare clients as

Table 1 Model Formulation and Data Integration

Examples of Model Formulation:

Variable	Equation	Unit
Ratio of the Perceived Average Monthly SSA Field Office Productivity to Desired Productivity	= Perceived Average Monthly SSA Field Office Productivity/Desired Productivity of SSA Field Office (x-axis) Ratio of the Perceived Average Monthly SSA Field Office Productivity to Desired Productivity	Dimensionless
(y-axis) Effect of the Ratio of Perceived Average Monthly SSA Productivity on Staffing Levels		Dimensionless
Current Desired SSA Field Office Staffing Level	= SMOOTH (Effect of the Ratio of Perceived Average Monthly SSA Productivity on Staffing Levels* Initial Desired SSA Field Office Staffing Level, Adjustment Time for Effect of Actual to Desired Productivity Ratio) where, $A = smooth\ (B, \Delta t)$ means $A = A_0 + \int \frac{B-A}{\Delta t} .dt$ and A_0 is equal to B_0	Staff

Four Types of Data Sources and Examples of Parameter Estimated:

Data Source	N	Example	Comment Pertaining to Example
Direct administrative or historical data	34	Initial Desired SSA Field Office Staffing Level = 2,450 (Staff)	The New York State field staff was reduced from 2,450 to 2,232 between 1994 and 2004. The model replicates this from 1997 to 2004.
Imported from other empirical studies	7	Initial Fraction of Claims for Which an Appeal is Filed =.058 (dimensionless)	Information from David Stapleton's paper "The Eligibility Definition Used in the Social Security Programs for People with Disabilities Needs to be Changed in a fundamental Way," draft, March 22, 2004.
Interview data	24	Adjustment Time for Average SSA Field Office Perceived Productivity = 3 (months)	SSA field offices are continuously monitored for productivity purposes, but reports are generated on a quarterly basis.
Model logic	5	Adjustment Time for Changes in the Pool of Temporary Assistance = 3 (months)	Modeler estimated. Assumption is that data are compiled on a quarterly basis and changes would not be noticed immediately.
Total	**70**		

DDD clients. Similarly, the *Market Saturation Loop* (B4) describes, at a relatively high level, the story told by the market saturation hypotheses. Notice that the dynamic hypothesis dealing with decline in manufacturing jobs is handled as an exogenous effect that is not captured within any of the feedback loops of the system under study.

When the mapping process was completed, the abstract map of the policy system was transformed into mathematical equations. This formalization process allowed the map to be populated with empirical data and simulated for testing. This process incorporated data and feedback mechanisms in the policy system that are not perceived by the decision makers (Eden et al. 2009), and it enhanced the accuracy of the mental model by aligning the mental model with the policy system.

Figure 2B presents key portions of a formal simulation model of the SSA sector processing claims, as described in loop B2 in figure 2A. Figure 2B presents a small fraction of the total system structure within the simulation model—it is only one of the 15 sectors that contain details for all five competing dynamic hypotheses. Compared to the qualitative map, the formal simulation model explicitly identifies the structure that adjusts staffing level and that allocates staff. Table 1 illustrates the equations used to define the relationship between selected variables specified in figure 2B. The three equations illustrating model formulation in table 1 are meant to be read as a

whole, illustrating how the simulation model transforms two model variables, *Perceived Average Monthly SSA Field Office Productivity* and *Desired Productivity of SSA Field Office,* into a third variable, *Current Desired SSA Field Office Staffing Level.* Full documentation of the model structure containing 321 equations and its technical details can be found in MacDonald and Kim (2008).

It is also important to document how the key model parameters and relationships are estimated. In the full DDD model, 70 parameters and relationships were estimated. The lower portion of table 1 summarizes how the data sources for parameter estimation were documented for the DDD project in order to form the empirical basis of the dynamic hypothesis testing.

The final step involved testing the soundness of the model and simulating various policy scenarios. One of the frequently used model tests for building confidence in the simulation model is time-series replication tests. It compares the model-generated behaviors with empirical data—data that are different from those used to calibrate the model. Figure 3 compares the historical data on the number of average yearly New York State initial disability claims received by the DDD in a 70-month period between 1998 and 2004 with the output generated by the simulation model. The simulation result is an example of hundreds of simulation runs generated by the simulation model. We refer to this particular run as the *base run* because it contains the specific set of parameters that are most compatible with
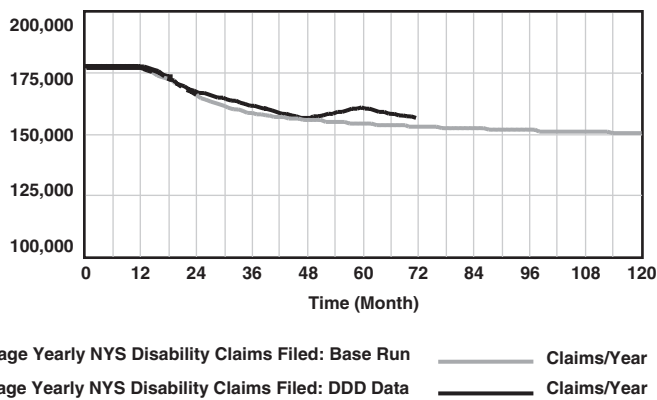
Figure 3 Simulated versus Actual Average New York State Disability Claims Filed

the time series and other data used to define the DDD's original problem. "What if" policy scenario runs are typically compared to this base run. It is also important to note that the simulation run in figure 3 reflects active feedback dynamics from all 15 sectors of the model, not just the pieces of system structure shown in figure 2B. The ability of the simulation model to endogenously generate the historical data would help build the client's confidence in the simulation model.

The DDD data in figure 3 end when this study was completed. The simulated time series captures both the replication of the historical data and the projection of a future possible scenario. In this case, the simulation (i.e., the base run in figure 3) forecasts a decline and stabilization in the number of disability claims if no other policy actions are implemented.

Once the running simulation model has been constructed and the management team has a reasonable level of confidence in it, the model can be used to explore policy scenarios. Although there were five possible causes originally proposed as to why the number of initial disability claims in New York State had declined, the model output indicated that most of the decline could be attributed to shifting productivity goals and staffing policy changes in the Social Security Administration. These initial hunches were verified by repeatedly running the simulation model and by performing sensitivity analysis on the data used to construct the model.

After the iterative process of running different policy scenarios in the model, the DDD management concluded that the fall in initial disability claims could be controlled without layoffs, if they expanded their view of the system.

Results of the Intervention
In the process of building and running a simulation model, the management team at the New York State Division of Disability Determination merged an interpretivist approach to their mental models with a positivist approach to the empirical data to support their management decision. The simulation model was based on causal structure drawn from the managers' mental models (summarized in figure 2A), as well as empirical data collected from 70 different sources (listed in table 1) used to calibrate the formal simulation structure (as one partial structure shown in figure 2B.)

An important feature of this modeling exercise was its ability to clarify the boundary of the policy system under study. The simulation model made clear distinctions between endogenous (within the boundary) and exogenous (outside the boundary) effects in the overall policy system. In the simulation model, of the five initial dynamic hypotheses, four were generated within the boundary of the model—only *Decline in Manufacturing Jobs* was conceptualized as an exogenous dynamic effect. However, some effects modeled as endogenous to the model were actually viewed by the DDD team as exogenous to the DDD system. For example, the DDD management perceived productivity pressures on the Social Security Administration as an exogenous factor. It was not under the direct control of DDD, and therefore it was perceived as outside the boundary of DDD system. On the other hand, the feedback effect of productivity pressure on the SSA was captured endogenously in the model in order to build a simulation structure that could endogenously generate the DDD's response to the simulated changes in the outside pressures.

These distinctions between endogenous and exogenous effects proved to have important practical implications for the DDD. When the management defined the boundary of their policy system as what is under their direct control, the simulation result of New York State disability claims continued to decline at a rate greater than the result shown in figure 3. This meant a need for layoffs. In fact, at the time that the modeling started, the DDD managers saw layoffs as a most likely but unwanted action in their future. However, the simulation model incorporating the workforce dynamics of SSA and county welfare offices as endogenous effects not only suggested that these agencies were having huge impact on the declining trend in initial disability claims, but also presented different future paths and policy options for the DDD.

After the iterative process of running different policy scenarios in the model, the DDD management concluded that the fall in initial disability claims could be controlled without layoffs, if they expanded their view of the system. In the short term, the management team decided to partner with agencies in other states to bring in out-of-state cases to maintain a stable workflow. For a longer-term measure, the DDD decided to start its own proactive outreach program to compensate for the decline in proactive outreach among the county SSA.

The DDD management saw the need for its own proactive outreach program because, as they expanded their mental model boundary, they came to realize that the SSA and county welfare offices had the types of resources constraints similar to what they were experiencing and that these agencies must manage their scarce human resources to meet their organizational goals. This indicated that the outside agencies' ability to generate additional initial disability claims for the DDD was determined by each agency's organizational goals and management decisions. Rather than blaming these agencies for the decline, the DDD decided to focus on what it could do in terms of outreach to generate appropriate initial disability claims. MacDonald and Kim (2008) give a more complete account of how

the simulation model supported "what if" policy analysis with an expanded system boundary.

Implications and Discussions

This article presented a framework that explains the role of simulation modeling as a way to enhance double-loop learning in management teams based on a case from the New York State Division of Disability Determination. Through the process of building, running, and modifying the computer simulation model, the accuracy and alignment of the management team's collective mental model was enhanced and the boundary of the mental model was expanded to include new policy options.

When we returned after several years for a qualitative and retrospective analysis of this case, managers reported the following:

> [W]e were pleased with the results of this study: it gave us the confidence to go ahead and make the changes that we needed. The simulation model made sense. We understood or had a hypothesis about why the initial disability claims had fallen and this allowed us to make decisions about future staffing levels with confidence. We ended up taking in work from other states so that we could keep the staff that we had. Based on this modeling, we started new outreach programs and our staff could stop doing the work from other states and focus on New York claims once the outreach started to work. In the longer run this strategy worked well for us.

The statement illustrates well how double-loop learning as defined in this article occurred in the case. Evaluations such as the foregoing are not unusual in simulation-based strategy and problem-solving studies. In addition to case studies, there are various formal methods for evaluating the effectiveness of modeling projects, including retrospective surveys (McCartt and Rohrbaugh 1989), experimental measurements (Huz et al. 1997), and more micro-level analyses of cognitive outcomes by people involved in the problem-solving sessions (Rouwette 2011).

This study raises a number of points regarding public managers' decision making in dynamically complex environments. First, effective team decisions are, by definition, forward looking (Andersen 1980). To make a decision about the future, managers must be guided by data about the past and their mental models. With limited data availability, the leverage point for enhanced decision making exists in mental model interventions. Second, effective tools for supporting decision making should combine an ability to elicit, represent, and create mental models with an ability to enhance the accuracy of mental models by incorporating a wide array of historical and administrative data, as well as expert judgment. Third, effective decision-making tools should promote double-loop learning. The modeling process initiates two types of double-loop learning. The first occurs when full system maps are created to build and explore mental models. The second occurs when simulation models are created to enhance the accuracy of mental models. Simulation-based hypothesis testing is especially valuable when there are multiple causal explanations for a problem of interest. When these multiple causal structures are active all at once, mental models lack the ability to consistently project the expected behavior of the system (Sterman 1994). Computer simulation models can aid our mental model simulation.

To date, large private sector firms such as General Motors and Boeing Aircraft have opened special internal consulting operations to make this simulation-based systems approach available to their managers. Specialized consulting firms such as McKinsey & Company and PA Consulting support other firms that do not have in-house capability. The federal government and national laboratories now use large-scale simulation models for defense and homeland security. But the question remains, how can this simulation-based approach penetrate the vast array of application possibilities in state and local governments? The real payoff for simulation-based dynamic hypothesis testing rests on the ability to promote double-loop learning in the daily decision-making practices of public sector managers. The framework and the case illustrated in this article should provide useful guidelines for future application of simulation-based systems approaches in wider variety of contexts to aid public sector decision making.

Acknowledgments

References

Ackermann, Fran, David F. Andersen, Colin Eden, and George P. Richardson. 2011. ScriptsMap: A Tool for Designing Multi-Method Policy-Making Workshops. *Omega* 39(4): 427–34.

Ackermann, Fran, and Colin Eden. 2005. Using Causal Mapping with Group Support Systems to Elicit an Understanding of Failure in Complex Projects: Some Implications for Organizational Research. *Group Decision and Negotiation* 14(5): 355–76.

Andersen, David F. 1980. How Differences in Analytic Paradigms Can Lead to Differences in Policy Conclusions. In *Elements of the System Dynamics Method,* edited by Jørgen Randers, 61–74. Cambridge, MA: Productivity Press.

Andersen, David F., John M. Bryson, George P. Richardson, Fran Ackermann, Colin Eden, and Charles B. Finn. 2006. Integrating Modes of Systems Thinking into Strategic Planning Education and Practice: The Thinking Person's Institute Approach. *Journal of Public Affairs Education* 12(3): 265–93.

Andersen, David F., Eliot Rich, and Roderick H. MacDonald. 2009. System Dynamics Applications to Public Policy. In *Encyclopedia of Complexity and Systems Science,* edited by Robert A. Meyers, 7051–61. New York: Springer.

Andersen, David F., and George P. Richardson. 1997. Scripts for Group Model Building. *System Dynamics Review* 13(2): 107–29.

Argyris, Chris, and Donald Schön. 1996. *Organizational Learning: Theory, Method, and Practice.* 2nd ed. Reading, MA: Addison-Wesley.

Axelrod, Robert, ed. 1976. *The Structure of Decision: The Cognitive Maps of Political Elites.* Princeton, NJ: Princeton University Press.

Bakken, Bjørn Tallak, and Martin Gilljam. 2003. Dynamic Intuition in Military Command and Control: Why It Is Important, and How It Should Be Developed. *Cognition, Technology and Work* 5(3): 197–205.

Barlas, Yaman. 1996. Formal Aspects of Model Validity and Validation in System Dynamics. *System Dynamics Review* 12(3): 183–210.

Berger, Peter L., and Thomas Luckmann. 1966. *The Social Construction of Reality: A Treatise in the Sociology of Knowledge.* Garden City, NY: Doubleday.

Berry, Frances S. 2007. Strategic Planning as a Tool for Managing Organizational Change. *International Journal of Public Administration* 30(3): 331–46.

Bryson, John M., and Sharon R. Anderson. 2000. Applying Large-Group Interaction Methods in the Planning and Implementation of Major Change Efforts. *Public Administration Review* 60(2): 143–62.

Bryson, John M., Barbara C. Crosby, and Melissa Middleton Stone. 2006. The Design and Implementation of Cross-Sector Collaborations: Propositions from the Literature. Special issue, *Public Administration Review* 66: 44–55.

Bryson, John M., and William D. Roering. 1988. Initiation of Strategic Planning by Governments. *Public Administration Review* 48(6): 995–1004.

Cannon-Bowers, Janis A., Eduardo Salas, and Sharolyn Converse. 1993. Shared Mental Models in Expert Team Decision Making. In *Individaul and Group Decision Making,* edited by N. John Castellan, 221–46. Hillsdale, NJ: Lawrence Erlbaum.

Coyle, R. Geoffrey. 1996. System Dynamics Applied to Defense Analysis: A Literature Survey. *Defense Analysis* 12(2): 141–60.

Doyle, James K., and David N. Ford. 1998. Mental Models Concepts for System Dynamics Research. *System Dynamics Review* 14(1): 3–29.

Easton, David. 1965. *A Systems Analysis of Political Life.* New York: Wiley.

Eden, Colin. 1988. Cognitive Mapping. *European Journal of Operational Research* 36(1): 1–13.

Eden, Colin, Fran Ackermann, John M. Bryson, George P. Richardson, David F. Andersen, and Charles B. Finn. 2009. Integrating Modes of Policy Analysis and Strategic Management Practice: Requisite Elements and Dilemmas. *Journal of Operational Research Society* 60(1): 2–13.

Eden, Colin, Fran Ackermann, and Steve Cropper. 1992. The Analysis of Cause Maps. *Journal of Management Studies* 29(3): 309–24.

Forrester, Jay W. 1961. *Industrial Dynamics.* Cambridge MA: Productivity Press.

———. 1994. Policies, Decisions, and Information Sources for Modeling. In *Modeling for Learning Organizations,* edited by John D. W. Morecroft and John D. Sterman, 42–63. Portland, OR: Productivity Press.

Forrester, Jay W., and Peter M. Senge. 1980. Tests for Building Confidence in System Dynamics Models. In *System Dynamics,* edited by Augusto A. Legasto, Jr., Jay W. Forrester, and James M. Lyneis, 209–28. New York: North-Holland.

Hirsch, Gary, Jack Homer, Elizabeth Evans, and Ann Zielinski. 2010. A System Dynamics Model for Planning Cardiovascular Disease Interventions. *American Journal of Public Health* 100(4): 616–22.

Huff, Anne Sigismund. 1990. Mapping Strategic Thought. In *Mapping Strategic Thought,* edited by Anne Sigismund Huff, 11–49. West Sussex, UK: Wiley.

Huz, Steven, David F. Andersen, George P. Richardson, and Roger Boothroyd. 1997. A Framework for Evaluating Systems Thinking Interventions: An Experimental Approach to Mental Health System Change. *System Dynamics Review* 13(2): 149–69.

Johnson-Laird, P. N. 1983. *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness.* Cambridge, MA: Harvard University Press.

Johnston, Erik, and Yushim Kim. 2011. Introduction to the Special Issue on Policy Informatics. *The Innovation Journal: The Public Sector Innovation Journal* 16(1): 1–4.

Kim, Hyunjung. 2009. In Search of a Mental Model-Like Concept for Group-Level Modeling. *System Dynamics Review* 25(3): 207–23.

Klimoski, Richard, and Susan Mohammed. 1994. Team Mental Model: Construct or Metaphor? *Journal of Management* 20(2): 403–37.

Lane, David C. 2001. Rerum Cognoscere Causas: Part 1—How Do the Ideas of System Dynamics Relate to Traditional Social Theories and the Voluntarism/Determinism Debate? *System Dynamics Review* 17(2): 97–118.

Lane, David C., C. Monefeldt, and J. V. Rosenhead. 2000. Looking in the Wrong Place for Healthcare Improvements: A System Dynamics Study of an Accident and Emergency Department. *Journal of the Operational Research Society* 51(5): 518–31.

Lindblom, Charles E. 1959. The Science of "Muddling Through." *Public Administration Review* 19(2): 79–88.

Luna-Reyes, Luis F., and Deborah L. Andersen. 2003. Collecting and Analyzing Qualitative Data for System Dynamics: Methods and Models. *System Dynamics Review* 19(4): 271–96.

Luna-Reyes, Luis F., Ignacio J. Martinez-Moyano, Theresa A. Pardo, Anthony M. Cresswell, David F. Andersen, and George P. Richardson. 2006. Anatomy of a Group Model-Building Intervention: Building Dynamic Theory from Case Study Research. *System Dynamics Review* 22(4): 291–320.

MacDonald, Roderick H., and Hyunjung Kim. 2008. Understanding the Dynamics of Disability Determination in New York State. Paper presented at the International System Dynamics Conference, July 20–24, Athens, Greece.

McCaffrey, David P., David F. Andersen, Paul McCold, and Doa Hoon Kim. 1985. Modeling Complexity: Using Dynamic Simulation to Link Regression and Case Studies. *Journal of Policy Analysis and Management* 4(2): 196–216.

McCartt, Anne T., and John Rohrbaugh. 1989. Evaluating Group Decision Support System Effectiveness: A Performance Study of Decision Conferencing. *Decision Support Systems* 5(2): 243–53.

Miesing, Paul, and David F. Andersen. 1991. The Size and Scope of Strategic Planning in State Agencies: The New York Experience. *American Review of Public Administration* 21(2): 119–37.

Mohammed, Susan, Lori Ferzandi, and Katherine Hamilton. 2010. Metaphor No More: A 15-Year Review of the Team Mental Model Construct. *Journal of Management* 36(4): 876–910.

Reagan-Cirincione, Patricia, Sandor Schuman, George P. Richardson, and Stanley A. Dorf. 1991. Decision Modeling: Tools for Strategic Thinking. *Interfaces* 21(6): 52–65.

Richardson, George P., and David F. Andersen. 1995. Teamwork in Group Model Building. *System Dynamics Review* 11(2): 113–37.

Richardson, George P., David F. Andersen, Terrence A. Maxwell, and Thomas R. Stewart. 1994. Foundations of Mental Model Research. Paper presented at the 12th International System Dynamics Conference, July 11–15, Stirling, Scotland.

Richardson, George P., and Alexander L. Pugh III. 1989. *Introduction to System Dynamics Modeling.* Waltham, MA: Pegasus Communications.

Rittel, Horst W., and Melvin M. Webber. 1973. Dilemmas in a General Theory of Planning. *Policy Sciences* 4: 155–69.

Rouwette, Etienne A. J. A. 2011. Facilitated Modelling in Strategy Development: Measuring the Impact on Communication, Consensus and Commitment. *Journal of the Operational Research Society* 62(5): 879–87.

Schneider, Susan C., and Reinhard Angelmar. 1993. Cognition in Organizational Analysis: Who's Minding the Store? *Organization Studies* 14(3): 347–74.

Schuman, Sandor P., and John Rohrbaugh. 1991. Decision Conferencing for Systems Planning. *Information and Management* 21(3): 149–59.

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

Stave, Krystyna A. 2002. Using System Dynamics to Improve Public Participation in Environmental Decisions. *System Dynamics Review* 18(2): 139–67.

Sterman, John D. 1994. Learning in and about Complex Systems. *System Dynamics Review* 10(2–3): 291–330.

———. 2000. *Business Dynamics: Systems Thinking and Modeling for a Complex World.* Boston: Irwin/McGraw-Hill.

Thomas, John Clayton. 1993. Public Involvement and Governmental Effectiveness: A Decision-Making Model for Public Managers. *Administration & Society* 24(4): 444–69.

Tolman, Edward C. 1948. Cognitive Maps in Rats and Men. *Psychological Review* 55(4): 189–208.

Vennix, Jac A. M. 1996. *Group Model Building: Facilitating Team Learning Using System Dynamics.* Chichester, UK: Wiley.

Walsh, James P. 1995. Managerial and Organizational Cognition: Notes from a Trip Down Memory Lane. *Organization Science* 6(3): 280–321.

Weick, Karl E. 1995. *Sensemaking in Organizations.* Thousand Oaks, CA: Sage Publications.

Wiener, Norbert. 1948. *Cybernetics or Control and Communication in the Animal and the Machine.* Cambridge, MA: MIT Press.

Yin, Robert K. 2009. *Case Study Research.* 4th ed. Thousand Oaks, CA: Sage Publications.

Zagonel, Aldo A., John Rohrbaugh, George P. Richardson, and David F. Andersen. 2004. Using Simulation Models to Address "What If" Questions about Welfare Reform. *Journal of Policy Analysis and Management* 23(4): 890–901.