Proceedings of the 2018 Winter Simulation Conference
M. Rabe, A.A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, eds.

A PRIMER FOR SYSTEM DYNAMICS MODELING AND SIMULATION

Ignacio J. Martinez-Moyano

Argonne National Laboratory
Decision and Infrastructure Sciences Division
9700 South Cass Avenue
Argonne, IL 60439, USA

ABSTRACT

In real-world systems, change and complex interactions are the norm. As a result, stakeholders face challenges and problems when they try to complete tasks and accomplish important activities in these systems. Linear and traditional analytic approaches fall short in helping us to understand problematic behavior in complex systems and in changing such behavior; they provide faulty and misguided recommendations. The system dynamics approach is based on feedback and control theory, and is well suited for tackling dynamic phenomena. This paper discusses the use and applications of system dynamics modeling, general lessons related to how and when to use this approach, and relevant tools.

1 INTRODUCTION

Feedback is at the core of the system dynamics (SD) approach. Feedback is critical because every day you make decisions that change the world, which then creates new information that changes your next decision in an iterative, continuous set of feedback processes (for an exemplary treatise on the concept of feedback in social sciences and systems theory, see Richardson 1991). The SD approach was invented by Jay Forrester in the late 1950s at the Massachusetts Institute of Technology (MIT) in the United States. The SD approach emerged out of servomechanisms engineering and control theory concepts, and has been applied to manage a wide range of complex systems. The SD approach combines theory of information-feedback systems, knowledge of human decision-making processes, an experimental model approach to complex systems understanding, and computer simulation to create "a computer-aided approach to policy analysis and design" (Richardson 1996, p. 657). Jay Forrester published the very first applications of the approach under the title "Industrial Dynamics" (Forrester 1958; 1961). At the time, the approach was called industrial dynamics because the applicability of the ideas was thought to be narrower than what it was later discovered to be. In 1970, Forrester changed the name of the approach and that of the group at MIT in which the ideas were being developed, from Industrial Dynamics to System Dynamics (MIT System Dynamics Group 1970).

SD modelers are concerned with how and why systems change over time, in order to make better decisions within such systems to be able to solve complex problems. The SD approach focuses on how problems happen within the system because it favors an operational viewpoint. The SD approach emphasizes why things change because it is by answering *why* that modelers and stakeholders can gain deep understanding of the causal processes at play (cause and effect relationships) that can provide leverage for change to solve problems of interest. The SD approach favors a point of view that originates from within system behavior – an endogenous approach – to explain change. It uses a broad system boundary to provide endogenous structural explanations of system behavior, embodied in feedback mechanisms. In contrast to other modeling approaches, which focus on the use of data, SD aims to gain understanding and insights about the multiple interacting, causal feedback structures that give rise to system behavior, which creates the data that are observed and recorded. System dynamicists believe that it is more important to understand in detail the causal engine that creates the data (which at least in part is interpreted as a problem) than to focus only on understanding the data without regard to what mechanisms generate such data. However, system dynamicists also believe that the use of data is critical in understanding feedback structure and system behavior. In the SD approach, models are empirically grounded using all possible sources of data available, including written data, numerical data, and data contained within the minds of agents in the system – also known as mental models. Mental models represent the world from the point of view of the individual, filtered by his beliefs, assumptions, and preconceived notions about causes, effects, delays, interconnections, and the boundary of the problem (see Chapter 10 in Senge 1990).

In order to understand how systems change over time, system dynamicists, among other things, examine time series data for variables of interest that tell interesting stories about the problems they study in systems (for excellent stories about change over time, based on data, see Pinker 2011; 2018). Using data to support compelling system stories is a powerful way to connect with stakeholders and policymakers. Although "we live in an era that presumes Big Data to be solution to all our problems" (Pearl and Mackenzie 2018, p. 6), there are several problems with the use of data that should always be guarded against. Data must be used with caution, because there are many potential problems during collection, recording, measurement, curation, maintenance, visualization, and so forth. Data, just like models, are not perfect.

Figure 1 shows the monthly number of passengers at Boston Logan Airport for 2008 through 2017. The two panels in Figure 1 show the same data, but they use different temporal granularity (monthly averages for the year and monthly totals). In this case, and potentially in many others, the use of different time units (i.e., years, months) will yield different dynamic behavior patterns because of the characteristics of the underlying causal structure of the system. In the case of an airport passenger load, the same data – at the weekly, daily, or minute-by-minute level – would look very different in terms of the type of pattern it would depict. For example, in the case depicted in Figure 1, comparing passenger loads at the monthly level instead of comparing yearly averages alone, and comparing the trends year by year or creating longer-term representations of the behavior of the system (right panel in Figure 1) can be valid alternatives that help show comprehensive patterns of behavior. The left and right panels in Figure 1 show different aspects of the dynamics of the system that are potentially relevant to different stakeholders and to different policy interventions. In the right panel, the upward trend of behavior can be identified and very month-specific comparisons can be made. In addition, the upward shift of the almost-identical yearly trend of behavior is much more evident, which allows for possible extrapolation about what possible behavior to expect in future years. In the left panel, the oscillatory pattern is lost, making the dynamics of the problem (at the yearly level) different than at the monthly level. Depending on the purpose of the model and on the characteristics of the system, the data needed to understand the problem may need to be collected at a different level of aggregation.

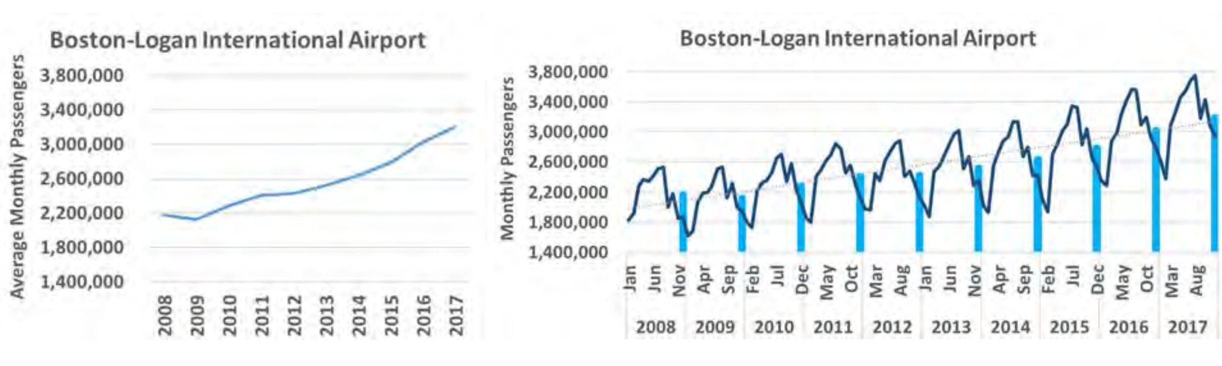

Figure 1: Passenger dynamics at Boston Logan Airport (Massport 2018).

In general, visual inspection of the data is insufficient; a careful quantitative examination of the evidence used to empirically ground SD models is always required to ensure the proper use of data to adequately parametrize and calibrate models. The data presented in Figure 1 show what is going on at the aggregate level at Logan Airport. However, depending on the problem under investigation, disaggregating the data into relevant areas of the airport (e.g., terminals) could provide additional insight into the dynamics of the problem. The SD approach favors the highest level of aggregation possible which enables the dynamic behavior of interest to be explained thoroughly using an endogenous theory.

The SD approach distinguishes between dynamic complexity and detail complexity in systems (Senge 1990). Detail complexity has to do with the number of individual components in a system, while dynamic complexity is related to the number of interrelationships the components have among them. Some systems have few components (low detail complexity) with few interactions among them (low dynamic complexity), while others have many components (high detail complexity) with few interactions. Systems with low dynamic complexity are relatively easy to analyze and understand. Systems with high dynamic complexity, however, are more difficult to understand and control because their internal mechanisms are tightly interconnected and can produce counterintuitive behavior (for a detailed explanation of the concept of counterintuitive behavior of systems, see Forrester 1975). SD focuses particularly on systems with high dynamic complexity.

2 THE MODELING PROCESS

SD focuses on the analysis of feedback structures in order to learn about the drivers of problems in complex systems. The knowledge generated is designed to create actionable insights that allow stakeholders to design and implement policy and decisions to create the desired behavior in the systems under study. There is no single way of conducting SD studies; the main objective is to learn about the problem and the system. There are several different approaches in the literature (for a synthesis of four different approaches see Table 3 in Martinez-Moyano and Richardson 2013, p. 108). These approaches differ in the number of steps and in how they develop the modeling process. However, the main focus of the modeling effort is always the same: to increase the level of understanding of the problem being studied. In order to learn about the drivers of problems in complex systems, the SD approach follows an iterative model-building process designed to produce two important outcomes: *understanding of the problem and the system* and *understanding of the model* (for a graphical representation of the process, see Figure 2 in Martinez-Moyano and Richardson 2013, p. 108). The model-building process is iterative by design to allow for a new understanding about the problem to generate refinements of the model and of the problem itself.

The SD model building process, which is also developed with groups of stakeholders (Luna-Reyes et al. 2006), always "begins and ends with understandings of a system and its problems" (Richardson and Pugh 1981, p. 16). Thus, the process forms a closed loop and "never a linear sequence of steps" (Sterman 2000, p. 87). The existence of a problem drives the modeling effort. Without the focusing power of a problem, there is no way to engage in an effective modeling process. Therefore, the most important step in the modeling process is to identify and define the problem.

2.1 Problem Identification and Definition

The first step in modeling, derived in part from the initial understanding of the problem in the system, is to clearly identify, articulate, and define the problem of interest. In this step, the modeler clarifies the purpose of the modeling effort and makes sure that all stakeholders understand that the engagement is confined to a specific aspect of the system. Clarifying the purpose of the model is paramount, because "the art of model building is knowing what to cut out, and the purpose of the model acts as the logical knife" (Sterman 2000, p. 89) to accomplish this. Why should a model address a problem within the system and not try to replicate the entire business or social system? Because the usefulness of models derives from their capacity to simplify reality in meaningful ways that enhance our comprehension of reality. In order to adequately represent an entire business or social system, the resulting model would have to be as complex as the system it is trying to represent, which nullifies its usefulness. Representing every single aspect of reality in a model is not an adequate strategy to enhance the understanding of a complex system. A clear statement of the model's purpose gives the modeler a clear idea of "the imaginary line separating what is considered (for modeling purposes) to be inside the system and what is considered to be outside" (Richardson and Pugh 1981, p. 48) the system boundary. To be useful, the boundary chosen should enclose the system of interest. In addition, the boundary should include the minimum number of components whose interactions generate the problematic behavior of interest. Therefore, the problem of interest needs to be identified in dynamic terms before the boundary can be identified.

In order to identify and define a dynamic problem, system dynamicists use a triangulation research process (Neuman 2006) including data from the literature, written and numerical data related to the problem, and stakeholder input. This process helps them thoroughly understand the problem and clarify the purpose of the modeling effort. In addition, because the SD approach is interested in identifying the dynamic characteristics of the problems, it is common to elicit, and use, reference modes – graphs over time – of all the variables that the stakeholders see as central to the problem and to understand what drives those variables. The creation of reference modes is always connected to clear identification of a relevant and useful time unit and time horizon of analysis because the "time horizon should extend far enough back in history to show how the problem emerged and describe its symptoms" (Sterman 2000, p. 90). When a time horizon is too short for the dynamics of the problem, insufficient information will be available to characterize these dynamics. Potentially, many important variables could remain constant, creating the illusion of stability in the system. Alternatively, when a time horizon is too long for the dynamics of interest, the additional information included in the set could possibly distort or mask the relevant behavior or dynamics. The choice of the time horizon will have an important impact on the perception of the problem and on the assessment of the importance and usefulness of different policies associated with changing the problem's dynamic characteristics.

2.2 System Conceptualization

Once a first draft of the problem identification and definition is complete, the process continues by posing a theory of what creates the problematic behavior that is observed. This theory is called a dynamic hypothesis of the problem. A dynamic hypothesis, "an initial concise overview of the feedback structures believed to be responsible for the problem behavior" (Martinez-Moyano and Richardson 2013, p. 29), connected with the reference mode of behavior of interest, is a key in the development of a SD model (tools for mapping the feedback structure will be discussed in Section 3). It is a hypothesis, because it is the best guess of what is causing the problem, but it is always tentative and subject to change and testing. This hypothesis is used to guide the modeling development effort; it is always subject to being updated and revised as new evidence is obtained or as tests reject parts of its current form. A key aspect of this hypothesis is that it should capture the main drivers of the problematic behavior as seen from the different points of view of the stakeholders. Part of the purpose of the creation of the dynamic hypothesis, and of the dynamic model afterward, is to produce an endogenous explanation of the problematic behavior. Something is considered endogenous when it is caused by internal factors – inside the system under study. An endogenous explanation will be achievable with a broad model boundary. A broad model boundary "that captures important feedbacks is more important than a lot of detail in the specification of individual components" (Sterman 2000, p. 96), because the focus is on explaining the dynamics of interest, not on explaining the components of the system. In determining the level of aggregation of their models, because the components and mechanisms can be specified at various levels of aggregation, SD modelers favor parsimony. At the same time, they also try to maintain realistic operational descriptions of the relevant casual mechanisms they seek to represent. For some problems, a high level of aggregation may be the correct choice (aggregated stocks), while for other problems, a disaggregated conceptualization (individual agents) may be the best way to capture the underlying mechanisms responsible for the problematic behavior (for an example of a comparative analysis, see Rahmandad and Sterman 2008). Independent of the level of aggregation, as long as the choice adequately matches the problem characteristics and available data (if budget is not a constraint, additional data can be collected if necessary), the broad model boundary focusing on capturing the important feedback mechanisms is the key element that will allow the modeler and stakeholders to produce an endogenous explanation of the behavior of interest.

Explanations are endogenous when they focus on internal factors of the system, not on external influences that are outside of the stakeholders' control. For example, the two teams playing in the FIFA World Cup Final (FIFA stands for Fédération Internationale de Football Association, which translates to International Federation of Association Football in English) face exactly the same external influences (rules of the game, meteorological conditions, etc.), but one team will eventually win and one will lose. That is the design of the game. Each team can choose to explain their defeat or win in endogenous or exogenous terms. For example, the winners could explain their victory by saying that the referees were more sympathetic to them than to the opposing team (exogenous explanation) or by saying that the opposing team gave up early in the game (exogenous explanation). Alternatively, the winners could say that their offensive strategy worked against the defense stance of the opposing team (endogenous explanation) or that their training paid off by giving them long-lasting endurance and an unbeatable attitude (endogenous explanation). In general, explanations about success tend to be endogenous (success is generally linked to internal action and effort), while explanations about failure tend to be exogenous (failure is generally linked to external causes or to things the stakeholders have no control over). As in the case of the FIFA World Cup Final, factors in real-world scenarios are never purely endogenous or exogenous; real scenarios are always affected by a mix of causes in which a mode is predominant, or more useful, in trying to understand a state of affairs. Often, the mode of analysis used to try to understand the state of the world does not match the predominant state of the world, which creates results that are not as desirable as they could have been. The SD approach is more advantageous when applied to feedback-intensive problems.

2.3 Model Formulation

After the problem has been adequately conceptualized, the next step is to formulate a model. A formal model of the problematic behavior of interest should always be based on the dynamic hypothesis obtained during system conceptualization. The model should start small and simple, as close as possible to the original dynamic hypothesis, and build toward complexity and completeness by adding variables and loops only as evidence is found and understanding develops. Dimensional consistency should be used from the beginning to facilitate formulation and to enhance the modelers' understanding and capacity to explain both the physics of the system and the behavioral and social aspects related to decision processes. It is best to always have a running model that is both understandable and clear. All equations written during formulation should make sense and should follow real-world processes. Only information that is available to decision makers in the real world can be available to decision makers in the model. Decision makers in the model cannot have access to instantaneous information that is not instantaneous in the real world (such as reports or states of the world). SD models are operational in the sense that they are representations of how the real world operates; therefore, stakeholders should be able to recognize that every parameter and variable has a real-life meaning. In SD models, there is no room for factors that amalgamate several components into unrecognizable items that magically modify the world. This is why all parameters and variables should have explicitly stated units of measurement, should be consistent, and should make sense. One should be clear about the fact that one is adding apples to apples (as opposed to adding apples to oranges). The operational nature of the SD approach pushes modelers to ensure that all relevant elements of the models are represented so that the models are both simple and understandable. For example, a good SD model of milk production should explicitly represent cows (Olaya 2015). All of these considerations, which are simple in isolation, become incredibly complex when put together in models with tens or hundreds of equations. This is why it is important to always start with a small model, to manage model complexity to fit the problem and the audience, to always strive for clarity and simplicity, to simulate early and often, and to always assure dimensional consistency from the very beginning – never adding variables or equations that do not have real-world counterparts or that cannot be easily explained to the real system's stakeholders.

2.4 Model Testing and Evaluation

Although "validation and verification of models is impossible" (Sterman 2000, p. 846), model testing should start with the first equation. It is never too early to thoroughly test a model and it is never too late to abandon a model if it does not help in understanding the problem of interest. Models should be tested to find flaws and to identify ways to improve them. In this way, testing models becomes a way to falsify the theories behind the models. Following Popper's (1959) falsification theory of scientific discovery, all tests should aim to falsify the theories we pose, not to confirm such theories. If, by conducting a test, the only thing we learn is that we know what we know, then the test is useless. For example, if we want to test a theory that all swans are white by looking for white swans, what we will learn when we collect evidence is that there are a number of white swans. Nothing else. It does not matter how many white swans we manage to collect; we will not learn anything about our theory because the test was designed to confirm the theory, not to falsify it. A more useful test would be to try to find a black (or any non-white) swan to test our theory because if we manage to find just one of these, our theory would be falsified and we would be able to learn something about swans that we did not know before conducting the test. We do not need more than one instance of a non-white swan to have enough evidence to reject the original theory and expand it to accommodate the new empirical evidence. Testing models, therefore, becomes a way to increase confidence in their usefulness and results. The more tests we conduct, the more likely it is that others will be confident that the model is robust and that the modeler did all the necessary work to adequately represent the problematic behavior. In this way, rigorous, quantitative, objective model testing is a social mechanism that generates intersubjective agreement about the model's usability and usefulness. There is little hope for testing and evaluation that is not transparent and reproducible.

Over decades, modelers interested in SD have developed specific tests to uncover flaws and improve many aspects of dynamic models (Forrester and Senge 1980; Barlas 1996; Sterman 2000). According to Sterman (2000, p. 858–861), some general areas of assessment include tests related to boundary adequacy, structure assessment, dimensional consistency, parameter assessment, extreme conditions, integration error, behavior reproduction, behavior anomaly, family member (class of system), surprise behavior, sensitivity analysis, and system improvement. Among these, behavior reproduction – a model's ability to produce simulated behavior consistent with historical or reference behavior modes – is one test of confidence that, although weak, many stakeholders (and practitioners of other simulation approaches) think of as powerful proof of model value. In general, tests explore the behavior and structure of the models we build. The dimensional consistency test is the most basic test of model value. If a model is not dimensionally consistent, this is evidence that the modeler does not understand the problem's basic mechanisms of operation. Modelers should ensure that all variables have real-world meaning and that all variables, consequently, have real-world units of measurement that are dimensionally consistent when used in mathematical formulae. For example, in a cash-flow model, when adding two streams of income, the two streams should have the same periodicity. The choice of periodicity does not really matter as long as it is the same for both streams so that the units of measurement are compatible. If one stream of cash is measured in dollars/year, the other stream should be measured in the same units. Alternatively, if one stream of cash is measured in dollars/month, then the other stream should also be measured in dollars/month. Similarly, if one stream of cash is comprised of dollars, then the other stream should also be comprised of dollars (to be precise both should be, for example, U.S. dollars). One cannot add together dollars and pesos because, although both are units of cash, they are inconsistent and do not represent the same thing. In order to consistently add together cash from two streams that are measured in different currencies, one would need to convert one of the two into the other unit of measurement using an adequate exchange factor with the correct units of measurement (e.g., units of currency A per units of currency B). Once one is certain that all variables have clearly identifiable, real-world counterparts (and units of measurement), one should test for the effects of extreme values on the behavior of the model because "the model should behave in a realistic fashion no matter how extreme the inputs or policies imposed on it may be" (Sterman 2000, p. 869). In a milk production model, no milk should be produced when there are no cows available. In an epidemic model, there should be no additional infections when no susceptible population is available to be infected. Demand for goods and services should decrease to zero when the price is high enough (it does not need to be infinite), and the price of goods and services should never rise to infinity.

On a more mechanical note, SD models are normally, although not always, formulated in continuous time and solved using numerical integration, which creates the possibility for the introduction of integration error. Testing for integration error is easy, and this error should be corrected by cutting down the integration interval to the point where the integration error is negligible. The results of the model should never be a function of the choice of integration method or integration interval (also known as time step).

2.5 Model Use, Implementation, and Dissemination

Extensive model testing and evaluation is a great start. However, modeling should not "stop at comprehensive analysis and reporting" (Roberts 2007, p. 134). A running model that has been thoroughly tested should be used to explore the problem of interest with the stakeholders and other interested parties, keeping in mind implementation and system change. The whole modeling process should revolve around the stakeholders' problems of concern, as well as increasing the understanding of what to do about these problems. When using the model, it is important to tell clear, coherent stories about the system of interest based on model results, using a language that the stakeholders can relate to and pictures (representations) of model structure that they can recognize in the real system. For example, in order for the model to be useful, the results should help stakeholders see why and how the problem of interest emerged and what can be done about it operationally. It is not enough to magically point to a solution that the stakeholders cannot identify with or where they cannot see how their actions in the system will create the necessary change to get from the current state to the desired state (how implementation will occur). The path to the desired state – transient dynamics – is as important as the end state of the system itself (the new equilibrium of the system). It is critical to involve the stakeholders throughout the entire modeling process to develop ownership and commitment to the results and insights. In order to focus on implementation of possible policies and solutions, it will be critical to recognize that the way from here to there might be counterintuitive. Stakeholders participating in the modeling effort need to understand the likely dynamics that the system will experience in the process of reaching a new steady state and how the implemented policies may have to be explained to other interested parties in the system if the end result is desired. For example, imagine that a system is in equilibrium, where the original behavior is represented by a dashed blue line at time 0, and that an intervention is implemented at time "a" (Figure 2). We will explore four possible transient dynamics resulting from the policy intervention introduced at time "a". First, after the policy intervention, the system may experience what has been termed "better-before-worse" behavior (alpha trajectory in the left panel of Figure 2). In the better-before-worse behavior, in the short run and peaking at time "b", system behavior improves as a result of the intervention; then it declines well below the original state to a final "alternative state A" at time "n".

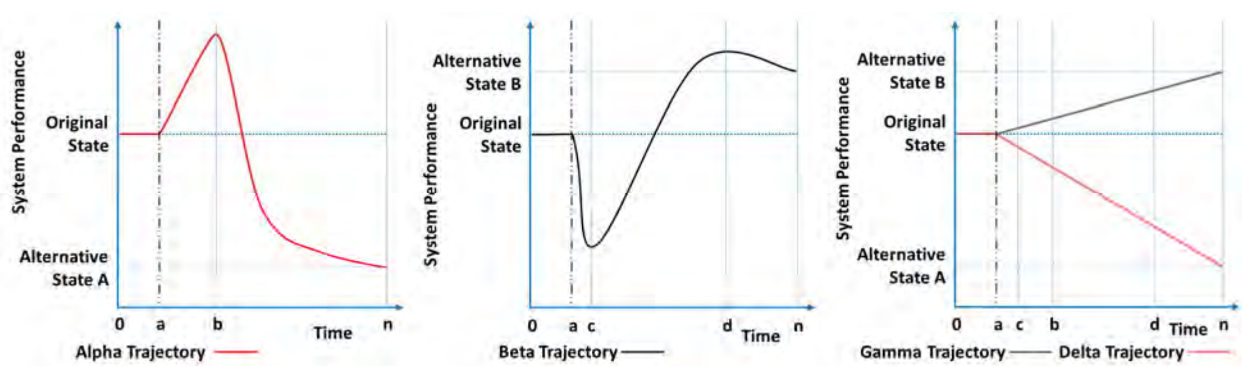

Figure 2: Trajectories of possible behavior over time.

During the time between the intervention and time "b", the system and its stakeholders will experience an improved state of affairs (the "better" part of the behavior mode), only to experience a subsequent long period of decline (the "worse" part of the behavior mode), which leads to a new state of affairs well below the original equilibrium state. This type of behavior is not uncommon in real systems. If unaware of the long-term dynamics, the stakeholders involved in implementing a policy that results in this type of behavior may think that they are creating value and that they have made "the right choice". In some cases, the decision makers may not have enough time to experience the long-term consequences of the policies they implemented and might not even realize that their choices created a long-term decline in performance for the system. If the designer of the policy intervention that generates the alpha trajectory is promoted at some point before time "b", she will not experience any portion of the downturn and will believe that the policy was a complete success. She will carry that thinking to her next assignment, perpetuating the spurious learning about what works and what does not work (for a deep exploration of learning in complex systems, see Sterman 1994).

Alternatively, after a policy intervention, a system may experience "worse-before-better" behavior (beta trajectory in the central panel of Figure 2). In this case, the system first experiences worse behavior than the original, while the system adjusts to interacting forces. Then, the system adjusts back to the original state and surpasses it. In some cases, the overall trajectory includes a peak performance point (time "d") followed by a second period of slight decline leading to a final equilibrium state (at a higher level than the original system state). To successfully implement a policy conducive to this type of behavior, the stakeholders would most likely face major resistance and doubt, especially between times "a" and "c". In many systems, unfortunately, at time "c" most policy designers or decision makers will have been substituted due to the large decrease in performance over a short period of time as a result of their policy choice. If these decision makers are not given enough time to see the consequences of their policy implementation (eventually positive in this case), they also will learn incorrect lessons about what influences the system and will carry these lessons with them to their next assignments. Even worse (for the system as a whole), the new decision makers brought in to "fix" the problem will also learn the wrong lessons; most likely, they will only "ride the wave" of the "better" part of the behavior that was going to happen as a natural consequence of the policy implemented before they arrived. In some cases, the newcomers will be praised for the increase in performance (erroneously attributed to their effort) and will be promoted without having made a contribution themselves. In other cases, the newcomers may change the natural course of the system with a new intervention that could further decrease system performance.

Finally, the gamma and delta trajectories (right panel in Figure 2) represent linear paths that end up exactly where the alpha and beta trajectories do. In these cases, the paths are uniform, predictable, and easy to extrapolate and plan for. These paths represent typical paths that system operators or stakeholders would use for planning. However, these paths are not at all connected to the nonlinear behavior of more realistic system behavior paths. Therefore, they should be avoided as guidelines for implementation.

2.6 Design of Learning Strategy and Infrastructure

After thoughtfully considering the model and how to implement lessons derived from it, model-based system stories – illustrated using simplified causal diagrams – can be used with the stakeholders as a powerful mechanism to create a long-term learning process for change. In a model-based study, small, simplified models can be developed and used to explain counterintuitive model behavior that gives rise to real system insights and interesting patterns of behavior. With such models, learning devices can be designed to help stakeholders who make decisions within the system learn how the complexities of the system interact and play out over time, and how policies can change the outcomes they care about. In this way, the model becomes part of a larger learning strategy; it becomes a boundary object, which enables knowledge to be shared and transferred between and among stakeholders (Carlile 2004). Although the model is a powerful embodiment – at a specific point in time – of knowledge about the problem and about the system, the modeling process itself is also a source of learning, knowledge, and insight for stakeholders and modelers (Forrester 1985). The modeling process is iterative; it builds on knowledge and insights developed in previous iterations, which allows for enhanced understanding, additional knowledge, and more powerful insights to develop every time the process is completed. As the level of understanding increases and insights about what to do are generated, goals may change, new assumptions will be made, and the environment in which the problem exists (if the problem persists) will change, starting the process anew.

3 TOOLS FOR SYSTEMS THINKING AND SYSTEM DYNAMICS

The two most commonly used tools to map the feedback structure of systems are causal-loop diagrams and stock-and-flow diagrams. There is no general agreement about which of the two should be used first. Although some scholars in SD advocate for the use of causal-loop diagrams first, because this type of diagram is considered simpler (see Table 11 in Martinez-Moyano and Richardson 2013), the use of causal-loop diagrams is not necessarily problem-free (Richardson 1997) or completely straightforward (Coyle 2000; Homer and Oliva 2001). Using stock-and-flow diagrams as the preferred structure mapping tool has its origin in the belief that it is easier to identify the (relatively) few stocks that are relevant to the problem than to focus the modelers' attention on several of possibly thousands of feedback loops that could influence the problem.

3.1 Causal-loop Diagrams: Causal Links and Causal Loops

Causal-loop diagrams represent closed causal connections, forming a circular relationship. The shortest possible loop consists of two variables (called vertices or nodes in network and graph theory) connected with two directional causal links (called edges) represented by directional arrows. The length of a loop is a function of the number of variables in it. Shorter loops tend to operate faster than longer loops do. Arrows indicate the direction of causality between variables. Each causal link is represented with a polarity sign to characterize the type of influence the variable at the beginning of the arrow (cause) has on the variable at the end of the arrow (effect). There are two possible link polarities: positive and negative. A positive causal polarity ("+" sign) means that, all else equal, the two variables will follow the same trajectory over time; increases (or decreases) in the variable at the beginning of the arrow will result in increases (or decreases) in the variable at the end of the arrow. Similarly, a negative causal polarity ("–" sign) means that, all else equal, the two variables will follow opposite trajectories over time; increases (or decreases) in the variable at the beginning of the arrow will result in decreases (or increases) in the variable at the end of the arrow. For example, elaborating on an example in Sterman (2000, p. 13), as shown in the left panel of Figure 3, "Eggs" and "Chickens" are connected with a positive causal link. Therefore, eggs and chickens follow the same trajectory of behavior over time, when eggs increase (or decrease), chickens increase (or decrease).

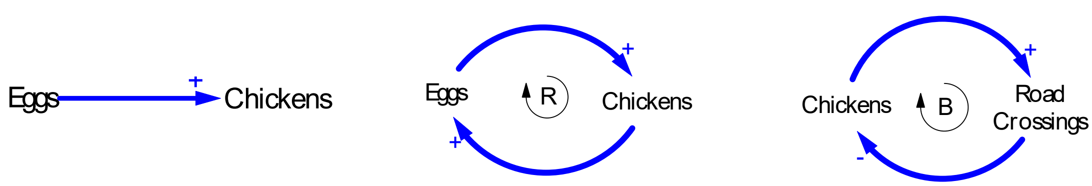

Figure 3: Causal links and causal loops.

Causal-loop diagrams are always analyzed in causal pairs, assuming that everything else remains constant (*ceteris paribus*), in order to isolate the effect of each cause when multiple causes and multiple causal loops exist at the same time. For example, as portrayed in the central panel of Figure 3, eggs and chickens are now part of a closed set of two causal links, one that describes the causal relationship between eggs and chickens and another that describes the causal relationship between chickens and eggs. Both causal connections are positive and together they create a positive, or reinforcing (identified with an "R"), causal (or feedback) loop. In this diagram, the modeler is saying that as eggs increase (or decrease), chickens increase (or decrease) and that as chickens increase (or decrease), eggs increase (or decrease) even more. The description of what happens in the system starts with a type of behavior that, after the loop is completed, is reinforced, creating more of the original behavior. Reinforcing cycles are engines of growth (or demise) that, when left unchecked, can generate explosive dynamics. What type of world would exist if the reinforcing cycle in the central panel of Figure 3 were the only force of nature in the world? How many eggs would you expect to see? How many chickens? If, in the real world, only reinforcing cycles existed, all processes of nature would seek explosive behavior that would resemble either dramatic growth or dramatic decline. In addition to positive causal links, there are negative causal links that represent causal relationships where balance is achieved as part of the process.

For example, continuing with the example of chickens (Figure 3), we can imagine that as chickens increase, road crossings increase due to a multitude of factors, and as road crossings increase, chickens decrease, closing a negative, or balancing, causal loop (identified with a "B" in the right panel of Figure 3). In order to create a balancing loop, an odd number of negative causal links must exist in the causal loop. In this case, there are two causal links, one positive (from chickens to road crossings) and one negative (from road crossings to chickens). An odd number of negative causal links is needed to create a negative (or balancing) loop because, for each negative causal link, the trajectory of the behavioral path reverses. Therefore, if the number of negative causal links is even, the original behavioral trajectory is maintained (in addition, the behavior follows its mathematical characterization; when you multiply two negative signs, the result is a positive sign). For example, as chickens increase, road crossings increase and as road crossing increase, chickens decrease (because the causal connection between road crossings and chickens is negative). Therefore, after one full cycle, an original increase in chickens has the final effect of a decrease in chickens due to its intermediate effect of increasing road crossings, which in turn decreases chickens. Why do chickens decrease when road crossings increase? That is a question that may have many answers. The most relevant answer points to the fact that when we draw a causal connection between two variables, we make a number of assumptions about the world that remain, for the most part, implicit in our models. In order to avoid confusion about assumptions used to develop causal-loop diagrams, and to increase model transparency, it is important to document as thoroughly as possible all aspects and sources of the data used to create such diagrams. Transparency "is an important attribute of useful models because it enables users to identify and understand the assumptions, relationships, and data used" (Martinez-Moyano 2012, p. 199).

In real-world applications of causal-loop diagrams, it is common to have multi-loop diagrams, because the real world is inherently a multi-loop environment. In order to create multi-loop diagrams, all that is necessary is to connect loops that have shared variables. As shown in Figure 4, expanding our representation of the world in which eggs, chickens, and road crossings exist, the causal-loop diagram has become a multi-loop diagram with two loops (one reinforcing and one balancing), four causal links (three positive and one negative), and three variables. Reinforcing loops accelerate change and are destabilizing processes in the system, while balancing loops counteract change and are stabilizing processes in the system.

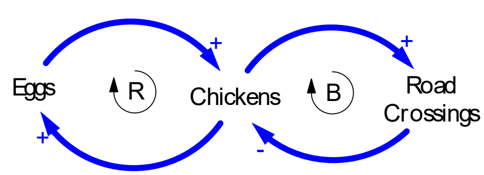

Figure 4: Multi-loop diagram.

In the system in Figure 4, the behavior of the two interacting causal loops depends on several factors that are not explicit in the diagram and that will determine the strength of each loop's gain. The gain of a feedback loop is "the product of the gains of its constituent links" (Oliva 2016, p. 30) or "the strength of the signal returned by the loop" (Sterman 2000, p. 145). The gain of the link between two variables is essentially how much the "effect" variable changes given a change in the "cause" variable (ratio of the output to the input). If, in the case of Figure 4, the reinforcing cycle has higher gain than the balancing cycle, the type of behavior expected will be different than if the two loops have identical gain. For example, a loop gain of positive two means that a change in a variable is doubled after a full cycle is completed, while a loop gain of negative 0.5 means that, after a full cycle, the effect on the variable is opposite to the original change with half its original strength (Sterman 2000). What kind of world would you expect to experience under a stronger balancing loop regime? What if an alternating loop strength process was achievable?

3.2 Stocks and Flows

Stocks, also called "levels," are accumulations in the system that "characterize the state of the system and generate the information upon which decisions and actions are based" (Sterman 2000, p. 192). When conceptualizing stocks in a system, the metaphor of a bathtub is useful. A bathtub can be seen as a stock because it accumulates water when the inflow of water from the faucet exceeds the outflow of water through the drain. In the same way, in systems, stocks accumulate material and nonmaterial things over time through changes in their inflows and outflows. Mathematically, the stock is the integral of the difference between the inflow and the outflow, plus the initial condition of the stock. For example, a stock of "population" grows with inflow from "birth rate" and declines from outflow from "death rate." If the birth rate is higher than the death rate, all other things being equal, the population level will increase. Alternatively, if the birth rate is lower than the death rate, the population level will necessarily decrease. The same type of behavior could be described for money in a bank account, food in a refrigerator, people in a convenience store, knowledge in your brain, students in a university, Nobel Prize winners, alcohol in the bloodstream, CO_2 concentration in the atmosphere, and so forth. All of us experience stocks, and the flows that control them, on a regular basis in our daily lives. Unfortunately, although not necessarily, we often do not identify their value in terms of the inertia, memory, delays, and disequilibrium potential they provide. For the most part, and for most of us, the role that stocks play in complex behavior is not clear; it is almost invisible. In order to identify stocks in real systems, imagine stopping time for an instant; as in a photograph, the stocks will remain visible while the flows disappear. You would know how many students are at your university at that specific time, but not how many were admitted and how many graduated that year; you could see how much food you have in your refrigerator, but not how much food you bought and consumed during the week; you could see how much money you have in your bank account, but not how much you earned and spent during that month. Stocks persist, but flows do not. In addition, in order to distinguish between stocks and flows, units of measurement can be used. Stocks are usually measured in basic units (people, dollars, gallons), while flows are necessarily measured in combined units because flows are always measured in the same units as the stocks, per unit of time. If the stock is measured in "dollars" (and the unit of time used is "month"), its corresponding flows, necessarily, will be measured in "dollars/month." Therefore, the amount of cash in your bank account (the stock) at any given point in time is measured in dollars, while the inflows to it, and outflows from it, are measured in dollars/month.

In SD, stocks are represented with rectangles and the flows that change them are represented with double arrows (representing pipes) and faucet-like icons (representing valves that control the flows) going into and out of the stocks (Figure 5). Stocks can only be modified by flows. Cloud-like icons represent both sources and sinks for the flows. These sources or sinks represent infinite stocks; these feed the flows and are outside the boundaries of the model, and their associated causal mechanisms are not of interest to the modeler and can be ignored – or collapsed into the source or sink.

Continuing with the eggs and chickens example, the egg stock is modified by egg laying (inflow) and egg hatching (outflow). In order for the egg stock to grow, either the egg-laying rate must increase (*ceteris paribus*) or the egg-hatching rate must decrease (*ceteris paribus*). Correspondingly, in order for the chicken stock to increase, either the birth rate of chickens (inflow) must increase or the death rate of chickens (outflow) must decrease. Each side in Figure 5 shows one stock with two flows. There is no limit to how many flows (inflow or outflow) a stock can have. If a stock has zero flows, it is basically a constant in the model because it will not change. If a stock has only inflows without outflows, that means it will only be able to grow over time (this type of state is called an absorbing state). For example, in some traditions, the stock of deceased people is a stock without an outflow (once deceased, there is no way to come back).

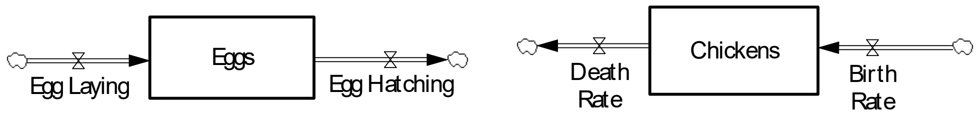

Figure 5: Egg and chicken stocks.

4 CLOSING THE LOOP: CONNECTING STRUCTURE AND BEHAVIOR

The most powerful use of SD is to deduce dynamic behavior from a feedback structure via computer simulation. In this way, as described in Section 2, policy insights and policy implementation can be identified and tested in combination with stakeholders and other system actors.

4.1 Modeling Disease Progression: The SIR Model

Models of disease progression exemplify the power the SD approach has to explain nonlinear behavior of complex systems. In particular, the SIR (susceptible, infectious, recovered) model has been used in the SD literature to show how the feedback structure is responsible for system behavior (Sterman 2000, p. 303–304). Figure 6 shows the structure of the SIR model.

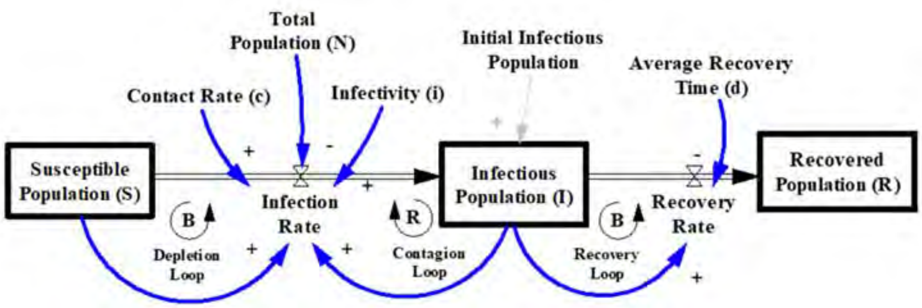

Figure 6: SIR model structure (from Martinez-Moyano and Macal 2013).

The main accumulations in the system are connected via the infection and the recovery rates. Three feedback loops change the rates that create the behavior shown in Figure 7. In the resulting s-shaped behavior, there is a “shift in loop dominance” (Kampmann and Oliva 2009; Oliva 2016).

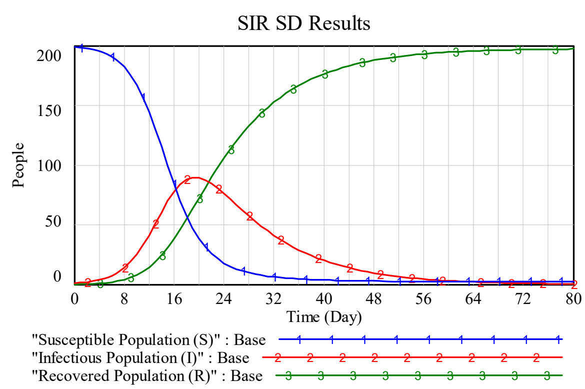

Figure 7: SIR model behavior (from Martinez-Moyano and Macal 2013).

4.2 Modeling the Criminal Justice System: Endogenous Causes of Policy Failure

In the state of Illinois, the Sentencing Policy Advisory Council (SPAC) and the Sentencing Commission are members of the Illinois State Commission on Criminal Justice and Sentencing Reform. In 2015, the Commission recognized that the overpopulation of the Illinois prison system had reached critical levels. This led to Executive Order 15-14, which created the Illinois State Commission on Criminal Justice and Sentencing Reform. The Commission's goal is to study the criminal justice system and to make recommendations for amending state laws, polices, and procedures to reduce the State's prison population by 25% by 2025. In 2014, together with the Illinois SPAC, the Office of Research of the Cook County (CC) Sheriff's Office, which oversees the largest jail in the United States, initiated an effort to use SD modeling to understand the dynamics of the jail population. Figure 8 depicts two important stocks identified in the system (inmates and bonded individuals) and three main "support" capacities. The two main stocks are interconnected and are not independent. At the core of the dynamics in this system, as it is the case in the swamping insight model (Ghaffarzadegan et al. 2011), the recidivism process – the process by which inmates leave the jail and then come back – was identified to be crucial.

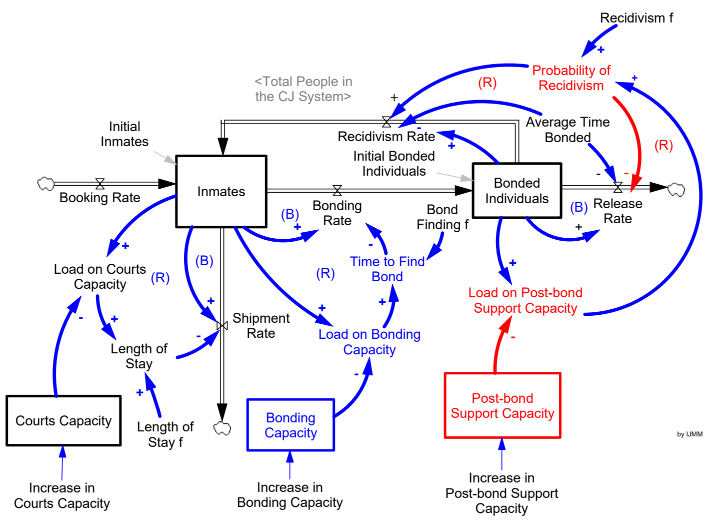

Figure 8: Criminal justice system conceptual model.

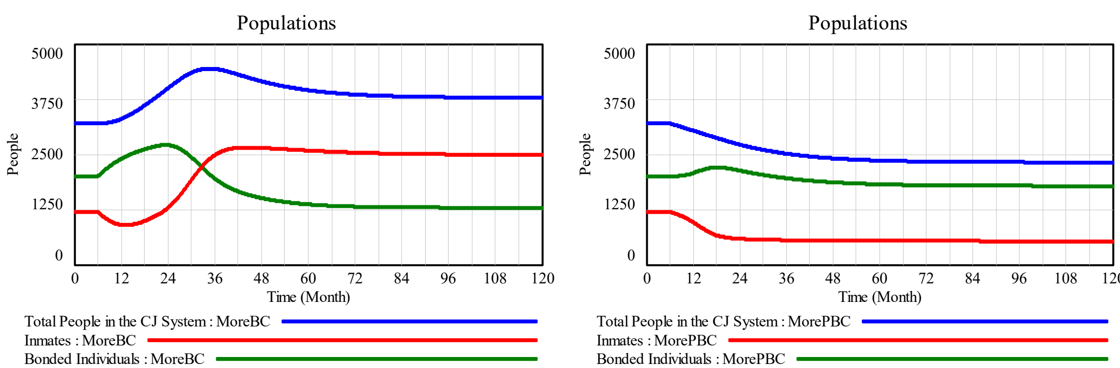

Figure 9: Criminal justice system conceptual model behavior (left more bonding capacity, right more post-bond support capacity).

Figure 9 shows simulation results, based on notional data. The model used to produce these results was initialized in equilibrium. Increasing the bonding capacity by 20% (MoreBC run, left panel in Figure 9) results in an increase in the total people in the criminal justice system (from 3,200 to 3,783 people; up 18%), while increasing the post-bond support capacity by 20% (MorePBC run, right panel in Figure 9) has the opposite net effect (from 3,200 to 2,318 people; down 28%).

5 CLOSING REMARKS

At the core of the SD approach are the concepts of feedback and endogeneity. The SD approach, consequently, uses a closed-loop approach to go beyond open-loop thinking and reveal in a holistic way *why* things change over time. In order to produce a reliable answer, SD develops empirically based, broad-boundary models that provide explanations, based on their feedback structures, of behavioral evolution in the problems of interest. In this approach, the feedback structure of the systems is responsible for the observed problematic behavior of systems. In order to develop models that follow this approach, "models go through constant iteration, continual questioning, testing and refinement" (Sterman 2000, p. 87), creating the possibility of extremely rapid deduction-induction cycles in the process. The SD approach follows an iterative model-building process designed to produce two important outcomes: understanding of the problem and the system and understanding of the model produced. The SD approach uses the focusing power of problems to clarify the important aspects to include in the study and in the model, and uses all relevant available sources of data including written, numerical, and mental models.

ACKNOWLEDGMENTS

This work is supported by the U.S. Department of Energy under Contract No. DE-AC02-06CH11357.

REFERENCES

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

Carlile, P. R. 2004. "Transferring, Translating, and Transforming: An Integrative Framework for Managing Knowledge across Boundaries". *Organization Science* 15(5):555–568.

Coyle, G. 2000. "Qualitative and Quantitative Modelling in System Dynamics: Some Research Questions". *System Dynamics Review* 16(3):225–244.

Forrester, J. W. 1958. "Industrial Dynamics: A Major Breakthrough for Decision Makers". *Harvard Business Review* 36(4):37–66.

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

Forrester, J. W. 1975. "Counterintuitive Behavior of Social Systems". In *Collected Papers of Jay W. Forrester*, 211–244. Cambridge, MA: Productivity Press.

Forrester, J. W. 1985. "'The' Model Versus a Modeling 'Process'". *System Dynamics Review* 1(1):133–134.

Forrester, J. W., and P. M. Senge. 1980. "Tests for Building Confidence in System Dynamics Models". *TIMS Studies in Management Sciences* 14:209–228.

Ghaffarzadegan, N., J. Lyneis, and G. P. Richardson. 2011. "How Small System Dynamics Models Can Help the Public Policy Process". *System Dynamics Review* 27(1):22-44.

Homer, J., and R. Oliva. 2001. "Maps and Models in System Dynamics: A Response to Coyle". *System Dynamics Review* 17(4):347–355.

Kampmann, C. E., and R. Oliva. 2009. "System Dynamics, Analytical Methods for Structural Dominance Analysis In". In *Encyclopedia of Complexity and Systems Science*, edited by R. A. Meyers, 8948–8967. New York: Springer.

Luna-Reyes, L. F., I. J. Martinez-Moyano, T. A. Pardo, A. M. Cresswell, D. F. Andersen, and G. P. Richardson. 2006. "Anatomy of a Group Model-building Intervention: Building Dynamic Theory from Case Study Research". *System Dynamics Review* 22(4):291–320.

Martinez-Moyano, I. J. 2012. "Documentation for Model Transparency". *System Dynamics Review* 28(2):199–208.

Martinez-Moyano, I. J., and C. M. Macal. 2013. "Exploring Feedback and Endogeneity in Agent-based Models". In *Proceedings of the 2013 Winter Simulation Conference*, edited by R. Pasupathy, et al., 1637–1648. Piscataway, New Jersey: IEEE.

Martinez-Moyano, I. J., and G. P. Richardson. 2013. "Best Practices in System Dynamics Modeling". *System Dynamics Review* 29(2):102–123.

Massport. 2018. Airport Statistics. https://www.massport.com/logan-airport/about-logan/airport-statistics, accessed August 21, 2018.

MIT System Dynamics Group. 1970. "D-1399 System Dynamics Staff Meeting Minutes". Boston, MA: MIT.

Neuman, L. W. 2006. *Social Research Methods: Qualitative and Quantitative Approaches*. Boston, MA: Pearson/AandB.

Olaya, C. 2015. "Cows, Agency, and the Significance of Operational Thinking". *System Dynamics Review* 31(4):183–219.

Oliva, R. 2016. "Structural Dominance Analysis of Large and Stochastic Models". *System Dynamics Review* 32(1):26–51.

Pearl, J., and D. Mackenzie. 2018. *The Book of Why: The New Science of Cause and Effect*. New York, NY: Basic Books.

Pinker, S. 2011. *The Better Angels of Our Nature: Why Violence Has Declined*. New York, NY: Viking Books.

Pinker, S. 2018. *Enlightenment Now: The Case for Reason, Science, Humanism, and Progress*. New York, NY: Viking Books.

Popper, K. 1959. *The Logic of Scientific Discovery*. New York, NY: Routledge Classics, Taylor & Francis Group.

Rahmandad, H., and J. Sterman. 2008. "Heterogeneity and Network Structure in the Dynamics of Diffusion: Comparing Agent-based and Differential Equation Models". *Management Science* 54(5):998–1014.

Richardson, G. P. 1991. *Feedback Thought in Social Science and Systems Theory*. Waltham, MA: Pegasus Communications.

Richardson, G. P. 1996. "System Dynamics". In *Encyclopedia of Operations Research and Management Science*, edited by S. I. Gass, and C. M. Harris, 656–660. Boston, MA: Kluwer Academic Publishers.

Richardson, G. P. 1997. "Problems in Causal Loop Diagrams Revisited". *System Dynamics Review* 13(3):247–252.

Richardson, G. P., and A. L. Pugh, III. 1981. *Introduction to System Dynamics Modeling with Dynamo*. Cambridge, MA: Productivity Press.

Roberts, E. B. 2007. "Making System Dynamics Useful: A Personal Memoir". *System Dynamics Review* 23(2/3):119–136.

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

Sterman, J. D. 1994. "Learning in and About Complex Systems". *System Dynamics Review* 10(2/3):291–330.

Sterman, J. D. 2000. *Business Dynamics: Systems Thinking and Modeling for a Complex World*. Boston, MA: Irwin McGraw-Hill.

AUTHOR BIOGRAPHY

IGNACIO MARTINEZ-MOYANO is a Computational Social Scientist in the Decision and Infrastructure Sciences Division at Argonne National Laboratory and Senior Scientist at Large of the Consortium for Advanced Science and Engineering (CASE) at the University of Chicago. Ignacio is also a Lecturer at the Graham School of the University of Chicago. Ignacio is President of the System Dynamics Society (2018) and Managing Editor of the *System Dynamics Review*. He has published in academic journals such as *Organization Science*, *Journal of Public Administration Research and Theory*, *ACM Transactions on Modeling and Computer Simulation* (TOMACS), *Computers & Security*, *System Dynamics Review*, and *Government Information Quarterly*. His email address is imartinez@anl.gov.