All models are wrong: reflections on becoming a systems scientist†

John D. Sterman*
Jay Wright Forrester Prize Lecture, 2002

John D. Sterman is the J. Spencer Standish Professor of Management and Director of the System Dynamics Group at the MIT Sloan School of Management.

Abstract

Thoughtful leaders increasingly recognize that we are not only failing to solve the persistent problems we face, but are in fact causing them. System dynamics is designed to help avoid such policy resistance and identify high-leverage policies for sustained improvement. What does it take to be an effective systems thinker, and to teach system dynamics fruitfully? Understanding complex systems requires mastery of concepts such as feedback, stocks and flows, time delays, and nonlinearity. Research shows that these concepts are highly counterintuitive and poorly understood. It also shows how they can be taught and learned. Doing so requires the use of formal models and simulations to test our mental models and develop our intuition about complex systems. Yet, though essential, these concepts and tools are not sufficient. Becoming an effective systems thinker also requires the rigorous and disciplined use of scientific inquiry skills so that we can uncover our hidden assumptions and biases. It requires respect and empathy for others and other viewpoints. Most important, and most difficult to learn, systems thinking requires understanding that all models are wrong and humility about the limitations of our knowledge. Such humility is essential in creating an environment in which we can learn about the complex systems in which we are embedded and work effectively to create the world we truly desire. The paper is based on the talk the author delivered at the 2002 International System Dynamics Conference upon presentation of the Jay W. Forrester Award.

Syst. Dyn. Rev. **18**, 501–531, (2002)

It's humbling to be honored by one's colleagues with the Jay W. Forrester Award. Thank you. I'm deeply appreciative.

When John Morecroft called to tell me about the award, he reminded me that one of my responsibilities was to give a talk here at the conference. Then he said, "But please, you have to make it shorter than the book."[1] This talk provides an opportunity to share some personal reflections on what I learned from the process of writing *Business Dynamics*, and from teaching system dynamics. Writing the book helped me become a better teacher, but I am keenly aware of how far short of my goals I fall in helping the students develop their systems thinking skills. And the students let me know it—one once told me I was a model professor. I thought this was high praise until I realized that a model is a small imitation of the real thing.

* John D. Sterman, MIT Sloan School of Management, Room E53-351, 30 Wadsworth Street, Cambridge MA 02142, U.S.A; E-mail: jsterman@mit.edu

† I thank Brian Dangerfield, Tom Fiddaman and my colleagues at MIT, including Jim Hines, Nelson Repenning, Anjali Sastry, and Peter Senge, for comments and assistance; all opinions and errors are mine.

System Dynamics Review Vol. 18, No. 4, (Winter 2002): 501–531
DOI: 10.1002/sdr.261

Received October 2002
Accepted October 2002

It takes a village

They say it takes a village to raise a child (a systems perspective if ever there was one). Well, it takes a worldwide community of systems thinkers to write a textbook. *Business Dynamics* benefited immensely from the advice, criticism, and encouragement of dozens of colleagues, students, and friends, many of whom are here today. It wouldn't have been possible to write the book at all without being able to draw on the great work and report the successful applications of system dynamics you all have done. You may find it hard to believe, but it was a struggle to keep the book as short as it is. It's a wonderful sign of the breadth and vitality of our field that one can write a thousand-page textbook and still only scratch the surface.

I am especially grateful to my colleagues at MIT and in the system dynamics community around the world who helped by providing data and examples, reviewing drafts, testing early versions in their courses, and in countless other ways. That list includes a great many of the people here today and many others who are not. So please excuse me if I can't name you all individually.

Still, a few words of thanks are needed. *Business Dynamics* would not be nearly as useful without the disc containing the more than 60 models developed in the book, and the software to run them. All the models are included in iThink, Powersim, and Vensim formats. This was possible only because High Performance Systems, Powersim Solutions, and Ventana Systems generously provided versions of their wonderful software for free. I'm most grateful to these companies and their hardworking people. My editor, Scott Isenberg, and the other terrific people at McGraw-Hill were enthusiastic and, above all, understanding, as the number of pages kept growing. I also want to recognize the vital help of my students, who constantly challenge me to make the discipline of system dynamics relevant, useful, and exciting. I can only hope they've learned as much from me as I've learned from them.

I owe an immeasurable debt of gratitude to my first system dynamics teachers, Dennis Meadows, the late Dana Meadows, and Jay Forrester. Besides the best system dynamics training I can imagine, they led by example, with high standards, integrity, and their passionate commitment. I feel blessed to have had the opportunity to learn from them. Finally, and most importantly, I thank my family. Though they couldn't be here today, *I* wouldn't be here without the love and support of my wife Cindy and my children David and Sarah.

Systems thinking *and* modeling for a complex world

Let me turn now to some reflections on the experience of writing *Business Dynamics*, of learning and teaching system dynamics for what is now 30 years. I'm sure I don't know the best way. I'm still learning and hope I'm still getting better at it. Of course, there isn't any one right or best way to teach system dynamics. There are many paths up the mountain. But not all paths are equal—many lead not to the summit but straight into the swamp. I've been down many of those.

Despite all these years of teaching and applying system dynamics, I still have trouble explaining it to people who ask me what I do. Is system dynamics science, engineering, or applied mathematics? Is it social science? Is it a philosophy? Is it a form of consulting, a theory of action? Is it hard or soft? The difficulty in answering the question ''what is system dynamics'' arises not because we don't know which of these things it is, but because it is all these things and more. The subtitle of *Business Dynamics* is *Systems Thinking and Modeling for a Complex World*. The word ''and'' here is important. System dynamics is grounded in control theory and the modern theory of nonlinear dynamics. There is an elegant and rigorous mathematical foundation for the theory and models we develop. System dynamics is also a practical tool policy makers can use to help solve important problems. And system dynamics is also a worldview, a paradigm in the sense of Thomas Kuhn. Like all powerful paradigms, it shapes every aspect of the way I experience the world.

Such breadth creates a tension. Many scientists and academics are deeply immersed in their specialties and skeptical of vague claims about ''complexity'' and ''systems'' studies that, they fear, lack rigor. Most managers have never studied science, nonlinear differential equations, or even calculus, or have forgotten it if they did. And most people, regardless of their background, are comfortable with their current philosophy, if they give such matters any thought at all. To be useful, system dynamics must be accessible to the widest range of scholars, students and policy makers, but without becoming a vague set of qualitative tools and unreliable generalizations. To be effective, it is often necessary to challenge some of our most deeply held beliefs, beliefs we often don't explicitly recognize. The resulting tension, the tension between qualitative systems thinking and formal modeling, between scientific rigor and the need to make decisions today, between gaining acceptance by clients and challenging dearly held beliefs, is compounded by the diversity of backgrounds within the community of managers, students and scholars interested in system dynamics, backgrounds ranging from people with no mathematics training to those with doctorates in physics.

The obvious strategy to deal with these tensions is to segment the market: Write a technical book on modeling for those with technical backgrounds, and a more popular book stressing the systems worldview for those with none; a treatise for the academic audience and a how-to book for practitioners. I rejected this strategy for several reasons. First, it has already been done very well. On the systems thinking side, Peter Senge's (1990) *Fifth Discipline* presents the concepts of system dynamics wonderfully, and places them in the context of learning and organizational change, all without any mathematics. On the technical side, there are many excellent treatises on the mathematics of complex nonlinear systems, though most focus on physical and biological examples (see, for example, Bar-Yam 1997, Mosekilde 1996, and Strogatz 1994).

But there is a more fundamental reason I rejected the segmentation strategy. The gulf between C.P. Snow's famous ''two cultures'' is unfortunately wider than when he first described it in 1959, while at the same time the need for basic numeracy and scientific literacy has never been greater (Snow 1959/1993; see also Paulos 2001). As systems thinkers, we must constantly strive to break down the false barriers that divide us, whether they rise up between the functional silos in a corporation, between scientific specialties, between the sciences and the humanities, or between the scholar's world of ideas and the policy maker's world of action. I believe a book focusing only on the technical side of system dynamics, or only on the qualitative, systems thinking side, only on theory or only on practice, only on business examples or only on public policy examples, would be inconsistent with the spirit and goals of system dynamics, would underestimate people's interests and capabilities, and would not provide readers with the resources they need to succeed in a world of growing dynamism and interconnection.

Policy resistance

While it's hard to define what system dynamics is, I don't have any trouble answering why it is valuable. As the world changes ever faster, thoughtful leaders increasingly recognize that we are not only failing to solve the persistent problems we face, but are in fact causing them. All too often, well-intentioned efforts to solve pressing problems create unanticipated ''side effects.'' Our decisions provoke reactions we did not foresee. Today's solutions become tomorrow's problems. The result is *policy resistance*, the tendency for interventions to be defeated by the response of the system to the intervention itself. From California's failed electricity reforms, to road building programs that create suburban sprawl and actually increase traffic congestion, to pathogens that evolve resistance to antibiotics, our best efforts to solve problems often make them worse.

At the root of this phenomenon lies the narrow, event-oriented, reductionist worldview most people live by. We have been trained to see the world as a series of events, to view our situation as the result of forces outside ourselves, forces largely unpredictable and uncontrollable. The concept of unanticipated events and ''side effects'' I just mentioned provides a good illustration. People frequently talk about unexpected surprises and side effects as if they were a feature of reality. A doctor may say, ''The patient was responding well to treatment, but died from unanticipated side effects.'' Our political leaders blame recession on unanticipated shocks such as corporate fraud or terrorism. Managers blame any difficulty on events outside their firms and (they want us to believe) outside their control, as for example when Cisco Systems blamed their record \$2.2 billion inventory writeoff and massive layoffs on ‘‘reduced capital spending and the global macroeconomic environment, which resulted in the reduction in our workforce and inventory charges we announced.’’ (Cisco Systems 2001 Annual Report). In fact, there is compelling evidence that, like other firms in the high-tech/telecommunications sector, Cisco’s own policies—from the design of its supply chain to pricing, production planning, and even the credit terms it offered customers—were central to the inflation and implosion of the great demand bubble (Goncalves 2002; Shi 2002).

There are no side effects—only effects. Those we thought of in advance, the ones we like, we call the main, or intended, effects, and take credit for them. The ones we didn’t anticipate, the ones that came around and bit us in the rear—those are the ‘‘side effects’’. When we point to outside shocks and side effects to excuse the failure of our policies, we think we are describing a capricious and unpredictable reality. In fact, we are highlighting the limitations of our mental models. System dynamics helps us expand the boundaries of our mental models so that we become aware of and take responsibility for the feedbacks created by our decisions.

(Almost) nothing is exogenous

It is hard to underestimate the power of the feedback view. Indeed, almost nothing is exogenous. If you ask people to name processes that strongly affect human welfare but over which we have no control, many people name the weather, echoing Mark Twain’s famous quip that ‘‘Everybody talks about the weather, but nobody does anything about it.’’ But today even the weather is endogenous. We shape the weather around the globe, from global warming to urban heat islands, the Antarctic ozone hole to the ‘‘Asian brown cloud.’’[2] For those who feel that global warming, ozone holes, and the brown cloud are too distant to worry about, consider this: Human influence over the weather is now so great that it extends even to the chance of rain on the weekend. Cerveny and Balling (1998) showed that there is a seven-day cycle in the concentration of aerosol pollutants around the eastern seaboard of the United States. Pollution from autos and industry builds up throughout the workweek, and dissipates over the weekend. They further show that the probability of tropical cyclones around the eastern seaboard also varies with a seven-day cycle. Since there are no natural seven-day cycles, they suggest that the weekly forcing by pollutant aerosols affects cloud formation and hence the probability of rain. Their data show that the chance of rain is highest on the weekend, while on average the nicest day is Monday, when few are free to enjoy the out of doors. Few people understand that driving that SUV to work helps spoil their weekend plans.

In similar fashion, we are unaware of the majority of the feedback effects of our actions. Instead, we see most of our experience as a kind of weather: something that happens to us but over which we have no control. Failure to recognize the feedbacks in which we are embedded, the way in which we shape the situation in which we find ourselves, leads to policy resistance as we persistently react to the symptoms of difficulty, intervening at low leverage points and triggering delayed and distant, but powerful feedbacks. The problem intensifies, and we react by pulling those same policy levers with renewed vigor, at the least wasting our talents and energy, and all too often, triggering an unrecognized vicious cycle that carries us farther and farther from our goals. Policy resistance breeds a sense of futility about our ability to make a difference, a creeping cynicism about the possibility of changing our world for the better. One of the main challenges in teaching system dynamics is helping people to see themselves as part of a larger system, one in which their actions feed back to shape the world in ways large and small, desired and undesired. The greater challenge is to do so in a way that empowers people rather than reinforcing the belief that we are helpless, mere leaves tossed uncontrollably by storm systems of inscrutable complexity and scope.

Bathtub dynamics

As important as feedback is, it is only one of the basic building blocks of complex systems. Enhancing our capability to understand counterintuitive dynamics also requires understanding stocks and flows, time delays, and nonlinearities. Consider one of the most basic of these concepts: stocks and flows.

In *Business Dynamics* I devote two full chapters (6 and 7) to the concept of stocks and flows, providing extensive examples and challenges designed to help people learn how to identify stocks and flows, map them, and understand their dynamics. Several reviewers of the manuscript complained that readers did not need such a remedial treatment of elementary calculus. Many of my students at MIT similarly complain that the class time I devote to this material is review. After all, they've all had calculus. Most have backgrounds in engineering, the sciences, economics, or mathematics, and many have prior graduate degrees in these disciplines.

Why then include two chapters on stocks and flows, on graphical integration? Experimental studies show that most people do not have a good grasp of these concepts. Linda Booth Sweeney and I presented the students in my classes with simple stock–flow structures and asked them to infer the behavior of the stock from information on the flows (Booth Sweeney and Sterman 2000; Sterman and Booth Sweeney 2002). Figure 1 shows an example in which you are shown simple patterns for the inflow and outflow to a single stock and asked to sketch the path for the quantity of water in the bathtub. The task is among the simplest possible examples of stock-and-flow thinking. There are no feedback processes, no time delays, no nonlinearities. There is only one stock. The outflow is constant, and the inflow follows a simple pattern.

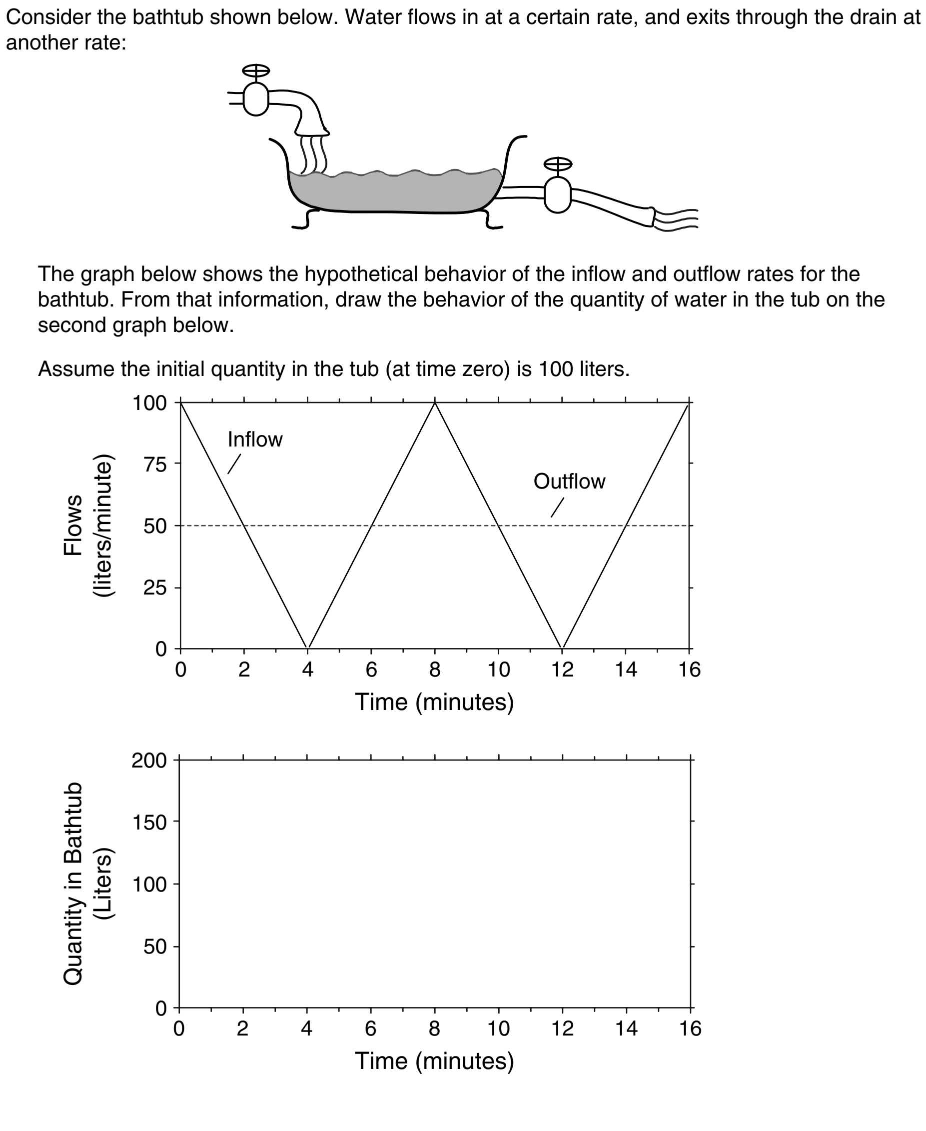
Unfortunately, only 36 percent of the MIT graduate students given this task answered correctly, even with generous coding criteria. Many appear to use a ''pattern matching'' heuristic, that is, assuming that the output of the system (the level of water in the tub) should follow the input (the sawtooth pattern of the net flow), as the erroneous responses in Figure 2 illustrate. Pattern matching often leads to wildly erroneous inferences about system behavior, causes people to dramatically underestimate the inertia of systems, and leads to incorrect policy conclusions. For example, a stock can rise even if the inflow is falling (obviously, when the inflow, though falling, remains above the outflow). A nation's debt grows even as it reduces its deficits. Of course, you may say. Yet many people find such behavior highly counterintuitive. When asked, for example, about global climate change, most people don't understand that atmospheric concentrations of greenhouse gases, already higher than at any time in the past 400,000 years, would continue to rise even if emissions fell to the rates called for in the Kyoto protocol—because current emission rates are roughly double the rate at which greenhouse gases are removed from the atmosphere by natural processes, while Kyoto calls for much smaller cuts. Most people believe that stabilizing emissions near current rates will stabilize the climate, when in fact stable emissions would guarantee continued increases in atmospheric greenhouse gas concentrations and a further increase in net radiative forcing, leading to still more warming. These errors are widespread even when people are explicitly told that current emissions are roughly double the natural uptake rate (Sterman and Booth Sweeney 2002).

These dismal results have now been replicated with diverse populations, from Austrian university students (Kainz and Ossimitz 2002; Ossimitz 2002) to MBAs at the University of Chicago (Sterman and Booth Sweeney 2002) to sophomores at the California Institute of Technology.[3] But are the results really a failure of systems thinking? Perhaps the reason people do poorly on these bathtub problems is not that they don't understand stocks and flows, but that they can't read graphs, or can't do the arithmetic, or aren't given enough time. So, inspired by a task developed by Günther Ossimitz (2002), I developed an even simpler challenge (Figure 3).

The task presents you with a graph showing, over 30 minutes, the rate at which people enter and leave a department store. You are asked four questions. The first two (when did the most people enter/leave the store?) test whether you can read the graph and know the difference between the number entering and the number leaving.

The figures in italics show the correct answer and the fraction of 172 subjects at the MIT Sloan School of Management responding correctly, along with the fraction who selected ''Can't be determined.'' Answers were considered correct if they were within ±1 of the correct response (e.g., 3, 4, or 5 for question 1). Half the subjects received the questions in the order shown; half received the two stock and flow questions (3 & 4) first. There were no significant differences

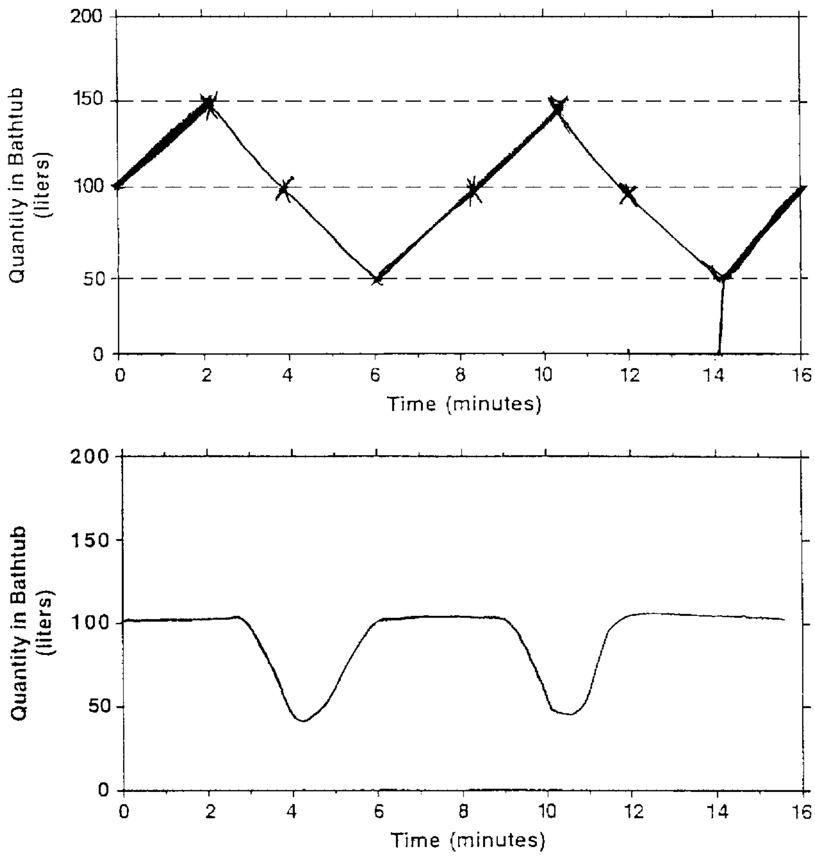
Fig. 1. A simple ‘bathtub’ task (Booth Sweeney and Sterman 2000)



in responses by question order; the results above aggregate the results of the two question order treatments.

Ninety-four percent of the MIT graduate students doing this task correctly answered these two questions.[4] The third and fourth questions (when were the most/fewest people in the store?) test your understanding of stocks and flows. To determine when the most people were in the store one need only recognize that the number in the store accumulates the flow of people entering less the flow of people leaving. Until minute 13 the number entering always exceeds

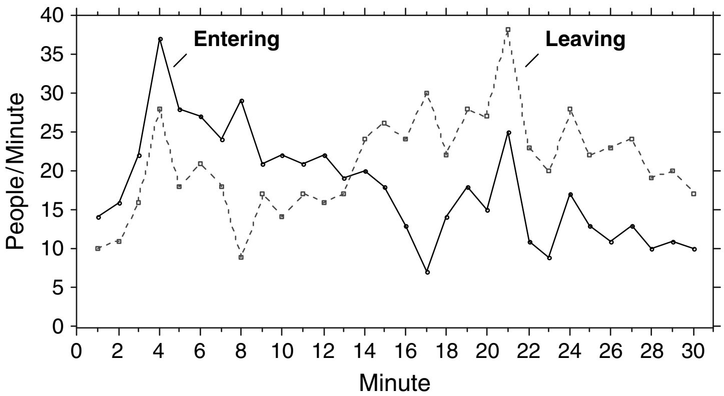
Fig. 2. Typical erroneous answers to the bathtub task (Booth Sweeney and Sterman 2000)



the number leaving, so the number in the store grows, while from minute 14 on, the number leaving exceeds the number entering, so the number of people in the store falls. The most people are in the store where the two curves cross. Only 42 percent correctly answered this question, and 17 percent indicated that the answer could not be determined from the information provided. Because the number in the store rises through minute 13 and falls thereafter, the fewest people are in the store either at the beginning or at the end. To determine which, you must judge whether more people enter up to minute 13 than leave afterwards, that is, whether the area between the rate of entering and rate of leaving up to minute 13 is greater or smaller than the area between the two curves from minute 14 on. Inspection of the graph readily reveals that the area between the curves from minute 14 on is larger than the area between the curves through minute 13 (in fact it is twice as large). More people left after minute 13 than were added up to that point. The fewest people are therefore in the store at minute 30. Only 30 percent correctly answered this question.

Fig. 3. The department store task

The graph below shows the number of people **entering** and **leaving** a department store over a 30 minute period.



Please answer the following questions.

Check the box if the answer cannot be determined from the information provided.

1. During which minute did the most people enter the store?

   Minute ___4___ *(94%)* ❑ Can't be determined *(0%))*

2. During which minute did the most people leave the store?

   Minute ___21___ *(94%)* ❑ Can't be determined *(0%)*

3. During which minute were the most people in the store?

   Minute ___13___ *(42%)* ❑ Can't be determined *(17%)*

4. During which minute were the fewest people in the store?

   Minute ___30___ *(30%)* ❑ Can't be determined *(28%)*

Fully 28 percent indicated that the question could not be answered, including one subject who wrote:

☑ Can't be determined$_\wedge$ *by me.*

Note that determining when the most people are in the store does not require any calculation—one need only understand that a stock rises when its inflow exceeds its outflow and falls when outflow exceeds inflow, then note where the two curves cross. Determining when the fewest are in the store does require a qualitative judgment of whether the area between the curves is largest before or after minute 13, but people have no trouble determining which area is larger when asked. The problem, at least among these highly educated subjects, is not the inability to read graphs, but difficulty with the concept of accumulation, with stocks and flows.

I now ask students to try these simple bathtub tasks before we introduce the concept of stocks and flows. The purpose is not to embarrass, but rather to motivate them to build their intuitive understanding of these critical concepts. Most see that, far from being a waste of time, they can gain significant insight into dynamics by developing their ability to identify, map, and understand the behavior of stocks and flows, even if they have a strong technical background.

Yet people should learn how stocks and flows are related before they reach graduate school. The concepts can and ought to be taught early in grade school. Calculus is not necessary. Students usually wonder why they were never taught these skills, and we have had some wonderful conversations about the fractionation of knowledge in the K-12 grades (ages 5–18), the drive to learn formulae and do well on standardized tests rather than build intuition, the lack of real-world relevance in the curriculum, and other pressures that may contribute to this deficit in our reasoning skills. We also talk about the exciting work being done by the growing stock of teachers who successfully incorporate system dynamics concepts into their teaching.[5] It's gratifying to see so many business students engage in such a spirited way with questions of social policy.

These basics (feedback, stocks and flows, time delays, nonlinearities) are essential foundations for effective systems thinking and modeling. It is clear that people have poor understanding of these concepts. At the same time, we know that people *can* learn to think in feedback terms, to recognize and understand stocks and flows, time delays, and nonlinearities. It takes training and practice, and must be supported by tools such as management flight simulators. But it can be done. More difficult for people to learn, and perhaps even more important, are other core concepts of systems thinking and system dynamics. Failure to appreciate and live by these concepts hurts us more than failing to understand feedback and time delays.

Model boundary: Invisible fences in the mind

The first system dynamics article I ever read was Jay Forrester's (1971a) *Counterintuitive Behavior of Social Systems*. Jay argues that most people believe cause and effect are closely related in time and space, while in complex dynamic systems cause and effect are often distant in time and space. One of the goals of system dynamics is to expand the boundaries of our mental models, to lengthen the time horizon we consider so we can see the patterns of behavior created by the underlying feedback structure, not only the most recent events. I found and still find his argument compelling.

But expanding the boundaries of our mental models is much more than just recognizing the delayed and distant effects of our decisions. It requires crossing disciplinary boundaries, boundaries between departments and functions in a company, between specialties in the academy. It requires breaching barriers erected by culture and class, by race and religion.

In affluent suburbs of the United States many dog owners now use invisible fences. The invisible fence folks bury a cable around the perimeter of your yard. They put a special collar on your dog. Whenever the dog gets too close to edge of the yard, the collar detects a radio signal from the buried cable and gives the dog a shock. Dogs quickly learn where the boundary is and refuse to cross it. After a short training period, you can turn off the collar. The dog will still not cross the invisible fence.

We are just the same. We live in a society that trains us to stay within artificial and damaging boundaries far more effectively than any invisible fence trains a dog. Much of our education consists of getting punished for crossing boundaries. School teaches us that every subject is different, and knowledge is fragmented (math is completely separate from social studies, which is different from literature). You learn that there are jocks and nerds, our team and their team, good guys and bad guys; that you are either ''with us or against us.'' These invisible lines in the mind are the boundaries of our mental models (Meadows 1991: 281–283). Like dogs, we waste a lot of time barking uselessly at people who get too close to our territory. Academics too often look down on those outside their own specialties, which are defined ever more narrowly. Or consider discussions of the economy. We hear pundits pontificate about how economic events will affect workers, consumers, taxpayers, and investors, as if these were separate species competing for survival in a zero-sum world, when each of us is all of the above: we work, we consume, we pay taxes, we benefit from government services, and we invest our savings for retirement. We are told logging old-growth forests is another case of jobs versus the environment, as if the economy could exist without a healthy environment, or the environment remain healthy if people have no jobs. Or perhaps you pursue a business career. What do you hear? That's a marketing problem. That's an operations problem. That's a human resources problem. And whatever you do, don't bring your personal problems to work.[6]

But we do not face marketing problems, operations problems, financial problems, and people problems; we don't have workplace issues, personal problems, and family problems. We don't have problems as workers, consumers, taxpayers, or investors. We just have problems. We create these boundaries and impose these categories on the world to simplify its overwhelming complexity. Some boundaries are necessary and inevitable. But all too often, the invisible fences in our minds cut critical feedbacks, deny us the insights of people with different experience and perspectives, and breed arrogance about our ability to control nature and other people—and then our problems grow worse.

In system dynamics we've developed tools and processes to help expand the boundaries of our mental and formal models. We build model boundary charts, listing the variables that are endogenous, exogenous, and, as best we can, excluded. We have sophisticated protocols for group modeling. We are trained to be suspicious of exogenous variables. Perhaps, we say to an innocent-looking coefficient in an equation, you are not really constant, but part of the feedback structure of the system. As the late Barry Richmond urged, we ''challenge the clouds'' in our stock and flow networks. Sources and sinks are modeling abstractions—is it really acceptable to assume there is an unlimited source and infinite absorption capacity for the material flowing through the stocks in your model? Is there an infinite pool of customers to buy your product? Is there an infinite sink to absorb the wastes we spew into the environment?

Yet we and other modelers use these simple disciplines too little. Narrow model boundaries are all too common, from the mental models of the person on the street to the formal models published in the most highly respected scientific journals. By model boundary I mean not only substantive assumptions such as whether the interest rate is endogenous or exogenous or whether the production function assumes constant returns to scale, but also the more subtle boundaries imposed by all modeling methodologies, such as the assumption that data are numbers, or that human beings make decisions to maximize expected utility. Most of the critical assumptions in any model, mental or formal, are the implicit ones, the ones buried so deep that the modelers themselves are unaware of them (Meadows 1980; Meadows and Robinson 1985; 2002). The most important assumptions of a model are not in the equations, but what's not in them; not in the documentation, but unstated; not in the variables on the computer screen, but in the blank spaces around them.

Let me illustrate with two examples, both drawn from resource economics. First, consider the debate over the future supply of energy and mineral resources. Here's what Morris Adelman, a leading energy economist, had to say in 1993:

> Minerals are inexhaustible and will never be depleted. A stream of investment creates additions to proved *reserves*, a very large in-ground inventory, constantly renewed as it is extracted. . .. How much was in the ground at the start and how much will be left at the end are unknown and irrelevant. (p. xi)
> The fixed stock does not exist. (p. xiii)
> What exists, and can be observed and measured, is not a stock but a flow. (p. xiv)

Figure 4 shows the stock and flow structure corresponding to Adelman's statements. The only stock is the stock of proven reserves, increased by a flow of investment, and drained by extraction. Adelman's assertion that ''The much larger amount in the ground is unknowable and irrelevant, a nonbinding constraint'' (p. xiii) means additions to proven reserves, in his view, are best modeled as flowing from an infinite source.

Adelman's statements violate conservation of matter. Every ton of titanium and every barrel of oil added to the stock of proven reserves reduces the stock of titanium and oil remaining to be found in the future. Every ton and barrel extracted reduces the quantity remaining in the ground. As exploration adds to the stock of proven reserves, the stock of undiscovered resource falls. *Ceteris paribus*, the smaller the stock of resources remaining to be discovered, the lower the productivity of exploration activity must be (on

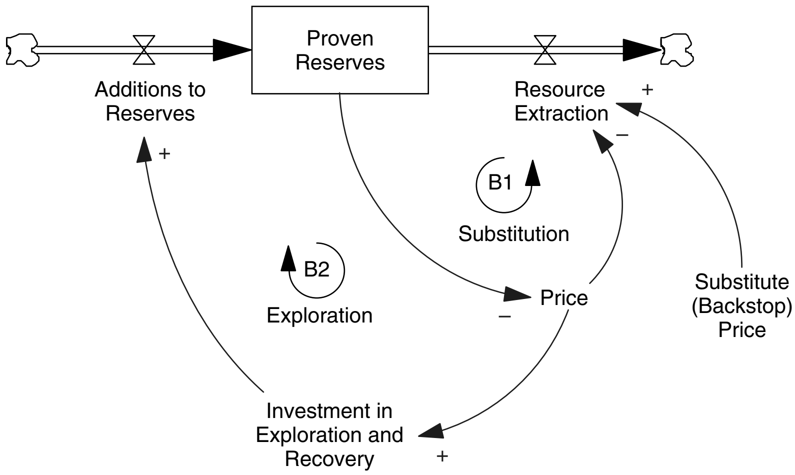


Fig. 4. Simplistic economic model of mineral resources. Investment includes improvements in technology

average), and the smaller the rate of addition to proven reserves will be for any investment rate. In the limit, if the stock of undiscovered resource fell to zero, the rate of additions to proven reserves would necessarily fall to zero.

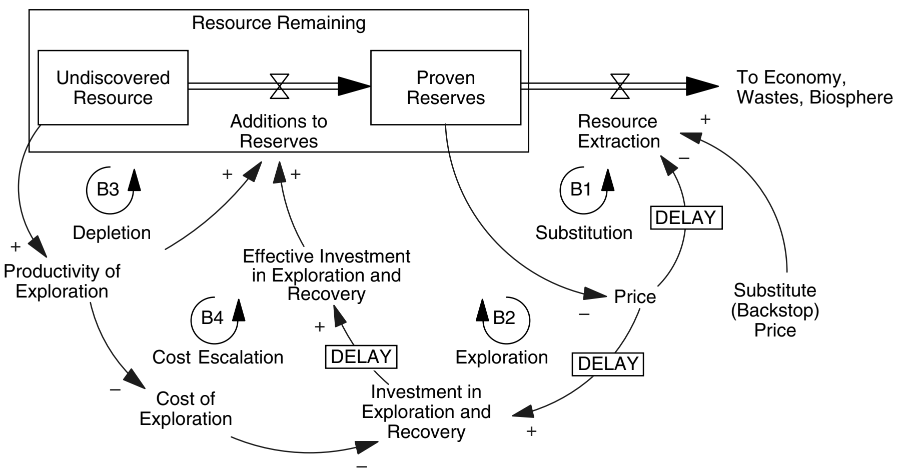
Economists argue that a drop in proven reserves will raise prices, leading through the familiar feedbacks of the free market to substitution of other resources (the Substitution loop B1 in Figure 4) and inducing additional exploration activity and improvements in technology that can increase exploration and recovery (the Exploration loop B2). And they are right. But additional exploration only drains the stock of undiscovered resource faster. Depletion must continue—the stock of resources in the ground must fall—as long as there is any extraction. Only if there is a "backstop" technology that can fully substitute for all uses of the nonrenewable resource at a finite price, in finite time, will demand fall to zero and halt depletion. How large the resource base is, what the costs of backstop technologies are, and whether a backstop technology can be developed before depletion constrains extraction and reduces economic welfare are empirical questions, not matters of faith. The very possibility that depletion might matter cannot be assumed away, to be made untestable with models in which resources are assumed infinite, the price system always functions perfectly, delays are short and technology provides backstops at low cost.

Turnabout is fair play. Narrow model boundaries are not restricted to economic models. Models that focus only on geological factors and calculate the lifetime of the resource remaining by assuming exogenous technology and extraction rates make the equally serious error of omitting the possibility that prices can alter incentives for innovation, exploration, and substitution. I am not arguing against economic models, or against geological models, but against narrow model boundaries. Models that consider *only* the price system or *only* geological factors omit important feedbacks, provide a poor guide to the future, and offer harmful policy advice.

A better model is shown in Figure 5. Here the total stock of resource remaining is the sum of proven reserves and the stock of undiscovered resources. For clarity the diagram omits many relevant feedbacks and aggregates resources into the two stocks shown. (See Davidsen, Sterman, and Richardson 1990 for a model in which the resource base is disaggregated into the standard categories used by the USGS, and with a full range of economic, technological, and geological feedbacks.) The total stock of resource in place falls with extraction.[7] Falling reserves raise prices, leading to substitution and boosting exploration, as in the simplistic economic model (though with delays). However, as exploration activity identifies more of the resource, the productivity of current exploration activity falls (the Depletion loop B3). The lower the productivity of exploration activity, the lower the expected return to exploration will be at any given price, so future investment in exploration drops (the Cost Escalation loop B4).

System dynamics models that integrate these geological, economic, and technological feedbacks date at least to the early 1970s (see Behrens 1973 and other models in Meadows and Meadows 1973); Sterman and Richardson (1985), Sterman, Richardson, and Davidsen (1988), and Davidsen, Sterman, and Richardson (1990) develop models for US and world petroleum resources integrating depletion and market forces with explicit, endogenous technology for exploration and recovery. These models show that extraction often grows rapidly and real prices often fall in the first part of the resource lifecycle as new discoveries and improving technology build proven reserves and lower exploration and extraction costs. As production grows, however, the stock of

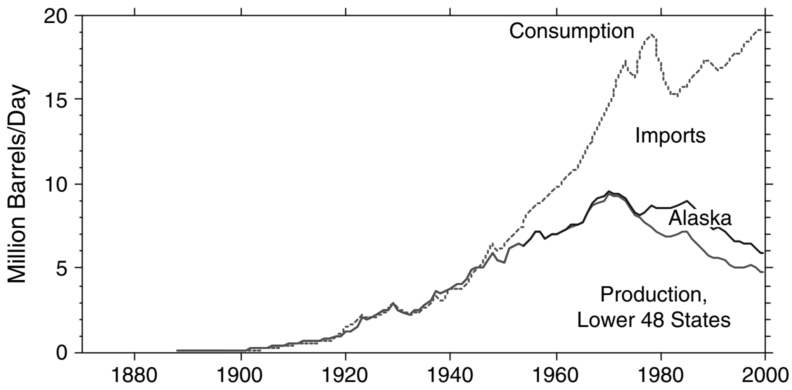
Fig. 5. Improved model of mineral resources, integrating economic and geological feedbacks



resource remaining falls. Unless and until a backstop technology completely substitutes for the resource, the quantity of resource remaining continues to drop, triggering the inevitable shift in feedback loop dominance from the production-enhancing feedbacks of technology and price to the production-limiting depletion and cost-escalation loops. To illustrate, Figure 6 shows production and consumption of petroleum in the Unites States from roughly 1859, when Col. Drake drilled the first modern well in Titusville, Pennsylvania. For about a century, exploration and innovation in exploration and recovery technology allowed production to keep pace with the extraordinary exponential growth of the rapidly industrializing economy. Estimates of the ultimate recoverable resource base rose dramatically, from less than 20 billion barrels for the lower 48 states in 1910 to as high as 600 billion barrels by 1960 (Sterman and Richardson 1985). Exploration and extraction were so successful, however, that by the 1950s new discoveries slowed and the yield to exploration effort fell. Production peaked in 1970 and has fallen ever since. Production from the lower 48 states and adjacent offshore areas is now less than half its peak value and continues to sink, while imports have grown to more than half of total consumption. Estimates of the ultimate recoverable resource fell by more than half. All this despite dramatic improvements in exploration and recovery technology and extended periods in which real oil prices and drilling activity reached all-time highs.

Yet the narrow boundaries in resource models persist. In the early 1990s William Nordhaus developed the DICE (Dynamic Integrated Climate Economy) model. One of the first and most influential of the so-called ‘‘Integrated Climate-Economy Models,’’ DICE has many features system dynamics modelers should view with approval. It links the climate and global warming with the dynamics of the economy. Until the integrated models were developed, research programs in climate change were fragmented. On the one hand were models of the biogeochemical processes governing the climate, such as the detailed GCMs

Fig. 6. US petroleum production and consumption



(general circulation models) that simulate global climate by tracking insolation, heat transport in the atmosphere and oceans, etc. in a spatially disaggregated framework. The concentration of greenhouse gases (GHGs) in these models is exogenous. Likewise carbon cycle models, which generate GHG concentrations, take emissions as exogenous. On the other hand were traditional economic growth models in which climate had no role whatever. Integrated models like DICE close an important feedback: the economy generates GHGs, which alter the climate, which feeds back to reduce economic growth and emissions.

Despite its virtues, the DICE are loaded. Consider its carbon cycle (Eq. 8 in Nordhaus 1992):

$$M(t) = \beta E(t) + (1-\delta_M)M(t-1) \tag{1}$$

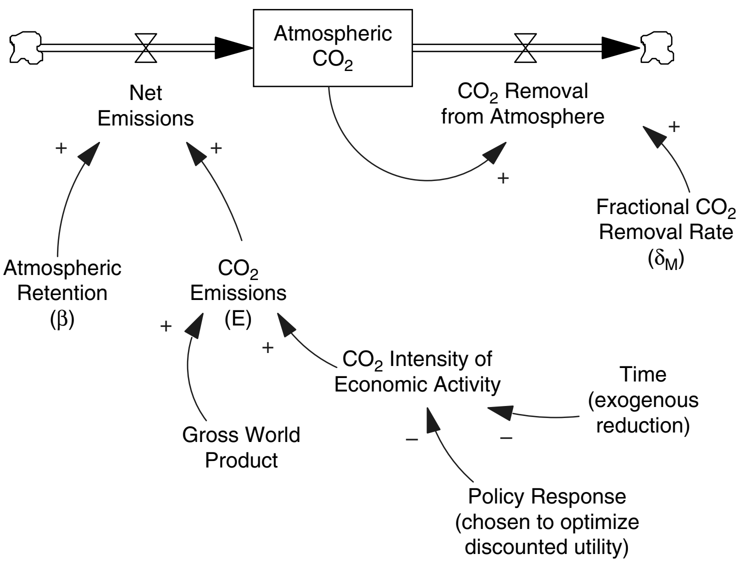
''where $M(t)$ is CO_2 concentrations relative to preindustrial times, β is the marginal atmospheric retention ratio, and δ_M is the rate of transfer from the rapidly mixing reservoirs to the deep ocean'' (p. 1316). Figure 7 shows the stock and flow structure for the model's carbon cycle. First note that the carbon sinks that remove CO_2 from the atmosphere (as it is taken up by biomass and dissolves in the ocean) are assumed to have infinite absorption capacity. Carbon, once removed from the atmosphere, disappears forever. In fact, these carbon sinks are finite. The carbon taken up by the land and oceans reduces net transfer (for example as the concentration of carbon in the ocean increases) and eventually makes its way back into the atmosphere. Further, the transfer rate is assumed to be a linear, first-order process (the removal time constant $1/\delta_M$ is constant and set to 120 years). However, there are important nonlinear constraints on carbon uptake by biomass as primary production is increasingly constrained by other nutrients; similarly ocean uptake is sharply constrained by the rise in the partial pressure of CO_2 as oceanic carbon concentrations grow (Oeschger *et al.* 1975). Carbon cycle models show that these feedbacks cause the fractional removal rate δ_M to decline as atmospheric CO_2 concentrations rise, as terrestrial and oceanic carbon sinks saturate, and as global mean temperatures increase (e.g., Sarmiento *et al.* 1995).

On the emissions front, Nordhaus estimated the ''atmospheric retention ratio'' β by regression, finding it to be 0.64. More than one-third of all CO_2 generated by the economy never enters the atmosphere. Where does it go? The value of β was estimated with annual data from the following equation:

$$M(t) - 0.9917\ M(t-1) = \beta E(t) \tag{2}$$

where $0.9917 = 1-\delta_M = (1-1/120)$. The left side is the net change in atmospheric CO_2, *assuming* removal is governed by a linear, first-order removal process with a 120-year time constant. A charitable interpretation is that 36 percent of total emissions is quickly absorbed out of the atmosphere (within a year), with the rest requiring an average of 120 years to be removed. However, the emissions that leave the atmosphere quickly are presumably absorbed by biomass or by the mixed, surface layer of the ocean. As these stocks

Fig. 7. Carbon cycle in the DICE model (Nordhaus 1992). Atmospheric CO_2 is defined as the excess above the preindustrial level



fill, additional removal from the atmosphere is constrained. And the carbon absorbed into these stocks either cycles back to the atmosphere, reducing net transfer, or is transferred to longer-lived carbon sinks such as the deep ocean, contributing to their saturation and reducing net transfer out of the atmosphere. Since none of these carbon reservoirs are represented, however, Nordhaus has not only assumed that much of the carbon quickly leaves the atmosphere, but that 36 percent of total emissions disappear forever, without a trace.

The DICE model violates conservation of matter, and the violation matters. Even a basic relaxation of these limiting assumptions, to account for sink capacities and conserve carbon, increases the warming generated by a given rate of CO_2 emissions, thus working against Nordhaus' conclusion that optimal carbon taxes are low (Fiddaman 2002).

Professor Nordhaus later developed the DICE model into RICE, the Regional Integrated Climate Economy model (Nordhaus and Boyer 2000, Nordhaus 2001[8]). RICE is similar to DICE, but disaggregated to represent eight regions so that carbon permit trading and other policies such as the Kyoto–Bonn accord can be represented. The climate sector now conserves carbon, and includes three compartments for carbon (atmosphere, ocean surface/biosphere, and deep ocean), but is still linear and remains overoptimistic. The other core assumptions of the original DICE model have been retained. Nordhaus (2001) describes the model, then comments on its limitations:

> Economic models, whether of the economics of global warming or of other phenomena such as business cycles, have great difficulty incorporating the many ''frictions'' that arise in real-world markets. In the present case, frictions are likely to plague the

> emissions market and to prevent equalization of carbon prices (that is, the prices of permits to emit carbon dioxide) in all participating countries and industries. Important frictions include impediments to trade, . . .; the inability of countries to get full credit for ''forestry'' options if regulations are tightly written; limits on the sale of permits by countries to ensure that ''overbooking'' of allowances does not occur; and a host of features such as transactions costs, regulatory and tax differences, risk and uncertainty, and unfamiliarity. Such frictions will force carbon prices to diverge in different regions or industries and thereby lead to higher costs of attaining the accord's emissions reductions targets.

These are indeed important issues that may affect the design and impact of policies to mitigate warming. But the next sentence is:

> Notwithstanding their importance, frictions are omitted from the present simulations.

There is something fundamentally wrong with a modeling process and peer review system that encourages modelers to build and allows the publication of models in which many of the factors the modelers themselves view as important are omitted. Now, we should commend Professor Nordhaus for listing some of the heroic boundary assumptions of his model. Many modelers are not so forthcoming, and the audience and client are left to discover the limitations of the models on their own, something most are ill-equipped to do, even in the too-rare circumstance that the model is available and properly documented.

Yet several serious problems remain. The omissions cited above constitute only a subset of the important boundary and methodological assumptions in the updated DICE and RICE models. As in many of the integrated climate–economy models, Professor Nordhaus makes many other assumptions, assumptions that work against his conclusions, assumptions that are not questioned or tested. These include:

- Consumers and producers make decisions that are consistent with global, intertemporal optimization under full information. (*We never make mistakes in economic decisions; the distant and delayed effects of our decisions, even those occurring over centuries, are fully internalized.*)
- Instant or rapid equilibration of factor inputs to prices. (*The economy and energy demand respond to prices very quickly; there are no significant lags in the turnover of carbon energy consuming capital stocks, the development of new technologies, changes in settlement patterns or transportation infrastructure, and so on.*)
- Energy efficiency improves and the carbon intensity of the economy falls exogenously. (*Technology improves automatically and without costs, delays, or ''side effects''.*)
- All non-energy resources are excluded. (*Interactions between climate change and other issues are unimportant.*)

- World population stabilizes early in the next century. (*All regions of the world quickly move through a demographic transition similar to but faster than the transition in the industrialized world.*)[9]
- Future potential economic output per person declines exogenously, limiting economic growth, carbon emissions, and the demand on the world's climate and other resources. (*The end of greed: the people of the world stop seeking economic growth and higher incomes.*)
- Nature and other species only matter insofar as they contribute to economic output and to the extent that their contribution might be compromised by warming. (*The potential extinction of the Orangutan or polar bear is irrelevant unless their loss reduces gross world product.*)
- Utility is determined by economic output per capita and is discounted over time. (*Your children are less important than you.*)[10]

As Tom Fiddaman (1997; 2002) shows, these assumptions bias the results of integrated models towards the conclusion that significant reductions in emissions are economically suboptimal. I submit that these assumptions range from the debatable (world population growth will stabilize rapidly, at about 11.3 billion) to the counterfactual (consumers and producers are intertemporal optimizers; there are no market failures; the economy equilibrates quickly; technology improves automatically and without cost or delays) to the immoral (the objective of policy is to maximize discounted utility with a utility function in which our children are much less important than we are and in which nature's only role is to promote production—indeed, if the extinction of the polar bear increased the availability of fish for humans it would be counted as a benefit of warming).

The assumption that growth in economic output per capita (or, equivalently, in the growth of factor productivity) declines autonomously so that gross world product eventually stabilizes, even absent any climate pressures, is particularly ironic, coming as it does from a critic of studies such as *World Dynamics* (Forrester 1971b) and *The Limits to Growth* (Meadows *et al.* 1972, 1974). The assumed exogenous reduction in growth can result only from Malthusian pressures other than climate change—which Nordhaus and Boyer (1999: 3–15) explicitly rule out—or from the utopian assumption that the people of the world are spontaneously becoming content with their material standard of living, even though large income gaps between rich and poor regions remain in the RICE projections. Assuming instead that people will continue to strive for higher incomes leads to such high emissions and such large climate changes that optimal policy would call for significant carbon taxes and deep emissions cuts today. The conclusions of the DICE and RICE models are not robust to correction of their errors and alternative plausible assumptions.

Model testing

The importance and difficulty of uncovering hidden assumptions requires a far greater role for model testing than is common in the social sciences. System dynamics has long had a sophisticated, flexible approach to testing. We stress multiple tests, from dimensional consistency to extreme conditions tests to tests of sensitivity to structural assumptions and aggregation. We emphasize the use of all types of data, not only statistical tests on numerical data. Because all models are wrong, we reject the notion that models can be validated in the dictionary definition sense of 'establishing truthfulness', instead focusing on creating models that are useful, on the process of testing, on the ongoing comparison of the model against all data of all types, and on the continual iteration between experiments with the virtual world of the model and experiments in the real world. We argue that focusing on the process of modeling rather than on the results of any particular model speeds learning and leads to better models, better policies, and a greater chance of implementation and system improvement (Forrester 1971/1985).

When Jay Forrester first articulated these views he was a lone voice in the modeling and social science community, which was dominated by the logical positivism imported to economics by Milton Friedman (1953). Today, more and more social scientists recognize the impossibility of validation, the provisional character of all models, and the need for a more eclectic and diverse set of tests (see Oreskes *et al.* 1994, Sterman 1994, and Chapter 21 in *Business Dynamics)*; see also Herbert Simon's (1963, 1979) and Paul Sameulson's (1963) critiques of Friedman's positivism). But we have a long way to go. Many important tests are simply never done. Many modelers focus excessively on replication of historical data without regard to the appropriateness of underlying assumptions, robustness, and the sensitivity of results to assumptions about model boundary and feedback structure. Modelers often fail to document their work, preventing others from replicating and extending it (see *Business Dynamics*, Chapter 21). Modelers and clients often suffer from confirmation bias, selectively presenting data favorable to their preconceptions. Such behavior only succeeds in generating mistrust of the model and suspicion about the intentions of the modelers, counter to the modeler's goals (a fine example of policy resistance).

If modeling is to fulfill its promise, a different approach is needed. Models rarely fail because we used the wrong regression technique or because the model didn't fit the historical data well enough. Models fail because more basic questions about the suitability of the model to the purpose weren't asked, because a narrow boundary cut critical feedbacks, because we kept the assumptions hidden from the clients, or because we failed to include important stakeholders in the process (Meadows and Robinson 1985, 2002).

To avoid such problems, whether as modeler or client, we must recognize that no one test is adequate. We must strive to use all types of data, both numerical and qualitative. We must do a better job of testing the robustness of our conclusions to uncertainty in our assumptions. Such testing requires far more than merely assessing parametric sensitivity, though we should encourage greater use of Monte-Carlo and other multiparameter tests in our work. Model results are usually far more sensitive to assumptions about the model boundary, level of aggregation, and representation of decision-making than to variations in parameters, yet sensitivity to these issues is only rarely assessed.[11]

We must insist on the highest standards of documentation. Models must be fully replicable and available for critical review. Build into the budget and time line sufficient resources to assess the impact of the work and document it fully so others can help you improve it. We must open the modeling process to the widest range of people we can, including our critics. We must design assessment into our work from the start so we can discover errors more quickly, measure the extent to which we meet our goals, and learn how to work more effectively in the future.

Unfortunately, all too often testing is inadequate, documentation is incomplete, important critics and stakeholders are excluded, and assessment is never undertaken. Worse, model testing is often designed to ‘‘prove’’ the model is ‘‘right’’ and model tests are presented as evidence designed to promote client acceptance. We are continually pressured by our clients, our students, our colleagues, and our own egos to slip out of the role of questioner and learner into the role of expert and teacher. Doing so often fails, by generating defensiveness and resistance. The phrase ‘‘getting client buy-in’’ should be banned from our lexicon. Taking the perspective that we are selling a ‘‘product’’ to the client is antithetical to a genuine inquiry process. Such an approach is designed to deflect criticism and promote the status and authority of the modeler. Instead, it makes learning difficult and ultimately erodes the impact of the model and the credibility of the modeler—and of all modelers.

Not surprisingly, the highest leverage point to enhance the impact of our modeling work is counterintuitive. Implementation success requires changing the clients' mental models. To do so the clients must become partners with us in the modeling process. Ultimately, our chances of success are greatest when we work with our clients to find the flaws in our models, mental and formal, then work together to improve them. In this fashion we all—modelers and clients—gradually develop a deeper understanding of the system and the confidence to use that understanding to take action. Paradoxically, a testing process designed to highlight the shortcomings of our models increases the chances of implementation and sustained success.

A hard look at soft variables

Another source of puzzlement for students of system dynamics relates to so-called soft variables and the role of numerical data. Jay Forrester argued early and, I believe correctly, that data are not only numerical data, that 'soft' (unmeasured) variables should be included in our models if they are important to the purpose. Despite the critical importance of qualitative information some modelers restrict the constructs and variables in their models to those for which numerical data are available, and include only those parameters that can be estimated statistically. These modelers defend the rejection of soft variables as being more scientific than ''making up'' the values of parameters and relationships. How, they ask, can the accuracy of estimates for soft variables be tested? How can statistical tests be performed without numerical data?

Omitting structures or variables known to be important because numerical data are unavailable is actually less scientific and less accurate than using your best judgment to estimate their values. ''To omit such variables is equivalent to saying they have zero effect—probably the only value that is known to be wrong!'' Forrester (1961, p. 57).

We should never compromise this principle. Omitting concepts because we have no numerical data is a sure route to narrow model boundaries, biased results, and policy resistance. Of course, we must evaluate the sensitivity of our results to uncertainty in assumptions—whether we estimated the parameters judgmentally or by statistical means. Modelers who follow these principles for modeling and testing developed owe no apology to those who would judge model ''validity'' by historical fit and statistical tests alone.

That said, it is important to use proper statistical methods to estimate parameters and assess the ability of the model to replicate historical data when numerical data are available. Unfortunately, some advocates of systems thinking go to the extreme of discounting the role of statistical parameter estimation and numerical data in general. They argue that qualitative insights are more important than numerical precision and that model behavior is insensitive to variations in most parameter values. They say that building a model for insight means they don't have to assess the behavior of the model against the historical data. These are serious errors, even when the purpose of a model is insight. Rigorously defining constructs, attempting to measure them, and using the most appropriate methods to estimate their magnitudes are important antidotes to casual empiricism, muddled formulations, and the erroneous conclusions we often draw from our mental models (Homer 1996; 1997). Ignoring numerical data or failing to use statistical tools when appropriate is sloppy and lazy. In my experience, many who avoid the proper use of numerical data do so not because they believe it is the best way to help people learn or solve important problems but because they don't want to take the time or don't have the skills to do it. No excuse. Failing to use these tools increases the chance that the insights you derive from your model will be wrong or harmful to the client.

Most important, we should not accept the availability of data as given, as outside the boundaries of our project or research. We must ask why concepts our modeling suggests are important have not been measured. Frequently, it is because no one thought these concepts were important. That perception, in turn, most often stems from the narrow boundaries of our understanding. There is a strong feedback: we measure what we care about, and those measurements alter what we believe is important. Because we tend to have short-term, event-oriented mental models, with narrow boundaries, with few feedbacks, and with weak understanding of the systems in which we are embedded, we tend to think what's important is what's salient, tangible, and familiar. As we measure these things they become even more real, while the delayed and distant effects of our decisions, the unfamiliar, and the intangible fade like wraiths. Thus we confuse the net income of the firm with the health of the enterprise, the amount we spend on training with the skills and knowledge of our employees, GDP per capita with happiness, and the size of our houses with the quality of our home life.

Human creativity is great: once we recognize the importance of a concept, we can almost always find ways to measure it. Within living memory there were no national income accounts, no survey methodologies to assess political sentiments, no psychological inventories for depression or subjective well-being, no protocols for semi-structured interviews or coding criteria for ethnographic data. Today, many apparently soft variables such as customer perceptions of quality, employee morale, investor optimism, and political values are routinely quantified with tools such as content analysis, surveys, and conjoint analysis. Of course, all measurements are imperfect. Metrics for so-called soft variables continue to be refined, just as metrics for so-called hard variables are. Quantification often yields important insights into the structure and dynamics of a problem. Often the greatest benefit of a modeling project is to help the client see the importance of and begin to measure and account for soft variables and concepts previously ignored.[12]

Why simulation is essential

Some advocates of systems thinking go even further, arguing that it is not necessary to build a formal, working simulation at all—that causal maps or other purely conceptual models are sufficient. They are mistaken. Simulation is essential for effective systems thinking, even when the purpose is insight, even when we are faced with a "mess" rather than a well-structured problem.

I am not opposed to all qualitative modeling. I do it myself. Building a formal model takes time. The data you need to build and test your model are rarely available without significant cost and effort. We must constantly make judgments about whether the time and cost of additional modeling and data collection are justified. A good qualitative mapping process will surface the mental models of the client. Often these have narrow boundaries and are dynamically impoverished. There is no doubt that many students, senior executives, and policy makers derive enormous value from expanding their mental models to include previously unrecognized feedbacks. But we must recognize that such qualitative modeling exposes us to one of the most fundamental bounds on human cognition: our inability to simulate mentally the dynamics of complex nonlinear systems. Indeed, our experimental studies show that people are unable to accurately infer the behavior of even the simplest systems, systems far simpler than those emerging from qualitative modeling work. Formal models, grounded in data and subjected to a wide range of tests, lead to more reliable inferences about dynamics and uncover errors in our mental simulations.[13]

Most importantly, computer simulations help build our intuition and improve our mental simulation capability. It is no accident that the most effective practitioners of qualitative modeling have extensive backgrounds in formal modeling. Their ability to identify the important feedbacks in a messy situation and draw useful and compelling inferences from them developed from their years of experience with formal modeling and simulation.

There is an even more fundamental reason why simulation is essential. There is no learning without feedback, without knowledge of the results of our actions. Traditionally, scientists generated that feedback through experimentation. But experiments are impossible in many of the most important systems. When experimentation is too slow, too costly, unethical, or just plain impossible, when the consequences of our decisions take months, years, or centuries to manifest, that is, for most of the important issues we face, simulation becomes the main—perhaps the only—way we can discover *for ourselves* how complex systems work, where the high leverage points may lie. The alternative is rote learning based on the authority of a consultant, teacher, or textbook, a method that dulls creativity, stunts the very systems thinking and scientific reasoning skills we hope to develop, and thwarts implementation.

All decisions are based on models . . . and all models are wrong

The concepts of system dynamics people find most difficult to grasp are these: All decisions are based on models, and all models are wrong. These statements are deeply counterintuitive. Few people actually believe them. Yet accepting them is central to effective systems thinking.

Most people are what philosophers call ‘‘naïve realists’’: they believe what they see *is*, that some things are just plain True—and that they know what they are. Instead, we stress that human perception and knowledge are limited, that we operate from the basis of mental models, that we can never place our mental models on a solid foundation of Truth because a model is a simplification, an abstraction, a selection, because our models are inevitably incomplete, incorrect—wrong. Many systems thinkers illustrate this with the famous story of the ancient astronomer who taught that the world is supported on the shoulders of a giant. ''But where does the giant stand?'' asks a student. ''On an immense turtle,'' the master replies. ''But on what does the turtle stand'' persists the student. ''Another turtle.'' This goes on a while, until the exasperated master shouts out ''It's turtles all the way down!''

Recognizing the limitations of our knowledge, the ''inevitable a priori'' assumptions at the root of everything we think we know, is deeply threatening (Meadows 1980). It's one thing to point out that someone *else's* opinions are 'just a model'—it's quite something else to recognize the limitations of our *own* beliefs. And how are we to make decisions if all models are wrong? The concept that it's turtles all the way down, that there is no ultimate, absolute foundation for our beliefs, is so deeply counterintuitive, so threatening, that most people reject it as ''obviously false'' or become so dizzy with doubt that they run screaming as fast as they can to someone who claims to offer the Truth. Much of the misery people inflict on others arises from the arrogant belief that only we know the True Path, and the resulting intolerance and fear of any who profess beliefs different than ours. Fundamentalism, whether religious or secular, whether the unquestioning belief in an all-powerful deity, the all-powerful state or the all-powerful free market, breeds persecution, hatred and war.

To help people open up to a new perspective, a new model, and change deeply entrenched behaviors, we must often first help them see the limitations of their current beliefs. Doing so is difficult. But even when we succeed, it is only part of the challenge. Yes, we might solve an important problem if we can help people see through a new lens, improve their mental models, and thus make better decisions. But in a deeper sense, we fail our clients and students when all we do is facilitate the old organizational change recipe of ''unfreeze, change, refreeze.'' We may only succeed in replacing one dogma with another, while strengthening people's belief that the scales have now fallen away from their eyes, that *now* they have the Truth. We must strive for more: helping people develop the critical thinking skills and confidence to continually challenge their own models, to uncover their own biases.

Yet we must recognize the inherent tension between being humble about the limitations of our knowledge on the one hand, and being able to argue for our views, respond to criticism, and make decisions on the other. Developing the capacity to see the world through multiple lenses and to respect differences cannot become an excuse for indecision, for a retreat to impotent scholasticism. We have to act. We must make the best decisions we can despite the inevitable limitations of our knowledge and models, then take personal responsibility for them. Mastering this tension is an exceptionally difficult discipline, but one essential for effective systems thinking and learning.

Too many "Why" questions?

Every semester I solicit midterm feedback from the students to help me improve my teaching. A student once commented that *Business Dynamics* and the homework assignments in my course were too difficult, writing "Too many 'why' questions." Reading this I knew I had failed that student. It's by asking those "why" questions that we come to understand that we are all embedded in systems, some natural, like the global climate, and some of our own making, like our schools, businesses, communities, and economies. It's by asking those 'why' questions that we gain insight into how we are both shaped by and shape the world, where we can act most effectively, where we can make a difference—and what we are striving for.

When human beings evolved, the challenge was survival in a world dominated by systems we could barely influence but that determined how we lived and died. Today the challenges we face are the result of systems we have created. The hurricane or earthquake do not pose the greatest danger. It is the unanticipated "side effects" of our own actions, side effects created by our inability to understand and act in consonance with our long-term goals and deepest aspirations.

What prevents us from overcoming policy resistance is not a lack of resources, technical knowledge, or a genuine commitment to change. What thwarts us is our lack of a meaningful systems thinking capability. That capability requires, but is much more than, the ability to understand complexity, to understand stocks and flows, feedback, and time delays. It requires, but is much more than, the use of formal models and simulations. It requires an unswerving commitment to the highest standards, the rigorous application of the scientific method, and the inquiry skills we need to expose our hidden assumptions and biases. It requires that we listen with respect and empathy to others. It requires the curiosity to keep asking those "why" questions. It requires the humility we need to learn and the courage we need to lead, though all our maps are wrong. That is the real purpose of system dynamics: To create the future we truly desire—not just in the here and now, but globally and for the long term. Not just for us, but for our children. Not just for our children, but for all the children.

It's demanding work. But it's also a joy. As Gauss said, "It is not knowledge, but the act of learning, not possession but the act of getting there, which grants the greatest enjoyment."[14]

Notes

1. Actually, *Business Dynamics* is even longer than it appears. Besides the text itself, there is a 500-page instructor's manual, available on-line

and on paper to anyone who teaches system dynamics. The manual contains suggested solutions to every one of the more than seventy *Challenges*—modeling exercises—in the book, from the frequency response of delays to policy responses to the HIV/AIDS pandemic.

2. The Asian brown cloud is a thick haze of pollution, three-quarters anthropogenic in origin, that causes drought in parts of Asia while flooding others, lowers agricultural productivity by reducing ground-level sunlight, and causes respiratory illness and death (see http://www.eapap.unep.org/issues/air/impactstudy/index.cfm).
3. Colin Camerer (personal communication).
4. In coding the responses I considered a response correct if it was within ± 1 minute of the correct answer, so subjects were not penalized for incorrectly reading the x-axis values.
5. The Creative Learning Exchange serves as a clearinghouse and source for system dynamics materials and tools for K-12 grades. See http://www.clexchange.org.
6. Or, in Tom Lehrer's (1965) immortal lyric, "'Once the rockets are up who cares where they come down. That's not my department' says Wernher von Braun." This gives an entirely new meaning to the concept of "silo thinking."
7. A good modeler should consider the possibility that new high-grade resources may be formed by biological or geological processes. For resources such as metals and fossil fuels the scientific consensus is that formation of new resources takes place over geological time scales and is negligible within the time horizon of concern in policy models. Good modeling practice also requires considering the fate of the resource after extraction, possibly including the stocks of resources in use, their disposal or consumption, waste generation, and, if possible, the potential for recycling. Note further that while the total stock of, say, petroleum or Ti in the earth may be enormous (and resources such as Ti, as elements, are conserved), there is a distribution of grades and extraction costs, with most of the total stock consisting of low-concentration, deep, or remote deposits with high extraction costs. Rational extractors develop and deplete the lowest-cost resources first, while lower-grade deposits in more costly locales remain uneconomic and unexploited until prices rise or technology improves. The depletion feedback in Figure 5 captures the long-run supply curve relating the yield of discovery and extraction effort to the distribution of the quantity remaining and the level of technology. That is, the productivity of exploration (measured in, say, tons of Ti or barrels of oil per \$ of exploration effort) $= f(R, T)$ where R is the stock of resource remaining and T is the state of exploration technology, with $\partial f(R, T)/\partial R \geq 0$ and $\partial f(R, T)/\partial T \geq 0$. The function depends on the distribution of resources in the earth, but must satisfy $f(0, T) = 0$.

8. The DICE and RICE models are available at http://www.econ.yale.edu/~nordhaus/homepage/dicemodels.htm.
9. World population in RICE stabilizes at about 11.3 billion and reaches almost 95 percent of this final value by 2100. These estimates assume rapid demographic transition in the developing world and continued low fertility in developed nations (see file <Rice99 web version.xls> at the site in note 8).
10. The social discount rate in the RICE model is an exogenous variable assumed to decline modestly over time. Nevertheless, the welfare of those alive in 2025 is weighted only 42 percent as much as the welfare of those in 1995. The weight falls to just 15 percent by 2065, when the grandchildren of today's undergraduates may themselves be in college. After 2155 the weight given our descendants is less than 2 percent of the 1995 value (see file <Rice99 web version.xls> at the site in note 8, row 37 on the World worksheet).
11. For a good example of testing robustness against deep methodological assumptions, see Repenning (2000; 2002), in which the dynamics of process improvement programs are modeled using both a rational actor game-theoretic framework and a behavioral, disequilibrium framework. Repenning shows that the impact of job security on employee participation in improvement programs is robust to assumptions about the degree of rationality of workers and managers.
12. For one of many examples, see *Business Dynamics*, Section 2.2, where system dynamics modeling of the auto leasing market led General Motors to create new market research instruments to assess how people trade off new versus high-quality off-lease used vehicles, an area the firm had previously ignored because under prevailing mental models the used and new vehicle markets were separate.
13. Sastry (1997) provides one of many examples.
14. As cited in http://www-gap.dcs.st-and.ac.uk/~history/Quotations/Gauss.html.

References

Adelman MA. 1993. *The Economics of Petroleum Supply*. MIT Press: Cambridge, MA.

Bar-Yam Y. 1997. *Dynamics of Complex Systems*. Addison-Wesley: Reading, MA.

Behrens W. 1973. The dynamics of natural resource utilization. In *Toward Global Equilibrium*, Meadows DL, Meadows DH (eds). Pegasus Communications: Waltham, MA, 141–164.

Booth Sweeney L, Sterman JD. 2000. Bathtub dynamics: initial results of a systems thinking inventory. *System Dynamics Review* **16**(4): 249–294.

Cerveny R, Balling R. 1998. Weekly cycles of air pollutants, precipitation, and tropical cyclones in the coastal NW Atlantic region. *Nature* **394**(6 August): 561–563.

Davidsen P, Sterman J, Richardson G. 1990. A petroleum life cycle model for the United States with endogenous technology, exploration, recovery, and demand. *System Dynamics Review* **6**(1): 66–93.
Fiddaman T. 1997. Feedback Complexity in Integrated Climate–Economy Models. Ph.D. thesis, MIT Sloan School of Management, Cambridge, MA 02142; available at http://home.earthlink.net/~tomfid/ (accessed December 2002).
______. (2002). Exploring policy options with a behavioral climate–economy model. *System Dynamics Review* **18**(2): 243–267.
Friedman M. 1953. *Essays in Positive Economics*. University of Chicago Press: Chicago.
Forrester JW. 1961. *Industrial Dynamics*. Pegasus Communications: Waltham, MA.
______. 1971a. Counterintuitive behavior of social systems. *Technology Review* **73**(3): 52–68.
______. 1971b. *World Dynamics*. Pegasus Communications: Waltham, MA.
______. 1971/1985. The model versus a modeling process. *System Dynamics Review* **1**(1): 133–134.
Goncalves P. 2002. When do minor shortages inflate to great bubbles? *Proceedings of the 2002 International System Dynamics Conference*. System Dynamics Society: Albany, NY.
Homer J. 1996. Why we iterate: scientific modeling in theory and practice. *System Dynamics Review* **12**(1): 1–19.
______. 1997. Structure, data and compelling conclusions: notes from the field (1997 Jay Wright Forrester Award Lecture). *System Dynamics Review* **13**(4): 293–309.
Kainz D, Ossimitz G. 2002. Can students learn stock–flow thinking? an empirical investigation. *Proceedings of the 2002 International System Dynamics Conference*. System Dynamics Society: Albany, NY.
Lehrer T. 1965. Wernher von Braun, *That was the year that was*. Reprise Records.
Meadows DH. 1991. *The Global Citizen*. Island Press: Washington, DC.
______. 1980. The unavoidable a priori. In *Elements of the System Dynamics Method*, Randers J (ed.). Pegasus Communications: Waltham, MA.
Meadows DH, Robinson J. 1985. *The Electronic Oracle: Computer Models and Social Decisions*. Wiley: Chichester.
______, ______. 2002. *The Electronic Oracle: Computer Models and Social Decisions* (excerpt). *System Dynamics Review*. **18**(2): 271–308.
Meadows DH, Meadows DL, Randers J, Behrens W. 1972. *The Limits to Growth*. Universe Books: New York.
Meadows DL, Meadows DH (eds). 1973. *Toward Global Equilibrium: Collected Papers*. Waltham, MA: Pegasus Communications.
Meadows DL, *et al*. 1974. *Dynamics of Growth in a Finite World*. Pegasus Communications: Waltham, MA.
Mosekilde E. 1996. *Topics in Nonlinear Dynamics: Applications to Physics, Biology, and Economic Systems*. World Scientific: River Edge, NJ.
Nordhaus W. 1992. An optimal transition path for controlling greenhouse gases. *Science* **258**(20 November): 1315–1319.
______. 2001. Global Warming Economics. *Science* **299**(9 November): 1283–1284.
Nordhaus W, Boyer J. 2000. *Warming the World: Economic Models of Global Warming*. MIT Press: Cambridge, MA. Web version dated 1999 available at http://www.econ.yale.edu/~nordhaus/homepage/dicemodels.htm (accessed December 2002).

Oeschger H, Siegenthaler U, Schotterer U, Gugelmann A. 1975. A box diffusion model to study the carbon dioxide exchange in nature. *Tellus* **27**: 168–192.

Oreskes N, Shrader-Frechette K, Belitz K. 1994. Verification, validation, and confirmation of numerical models in the earth sciences. *Science* **263**(4 February): 641–646.

Ossimitz G. 2002. Stock–Flow Thinking and Reading stock–flow-related graphs: An empirical investigation in dynamic thinking abilities. *Proceedings of the 2002 International System Dynamics Conference*. System Dynamics Society: Albany, NY.

Paulos JA. 2001. *Innumeracy: Mathematical Illiteracy and Its Consequences*. Hill & Wang: New York.

Repenning N. 2000. Drive out fear (unless you can drive it in): the role of agency and job security in process improvement efforts. *Management Science* **46**(11): 1385–1396.

———. 2002. A simulation-based approach to understanding the dynamics of innovation implementation. *Organization Science* **13**(2): 109–127.

Samuelson P. 1963. Discussion: problems of methodology. *American Economic Review, Proceedings* **53**(May): 231–236.

Sarmiento JL, Le Quéré C, Pacala SW. 1995. Limiting future atmospheric carbon dioxide. *Global Biogeochemical Cycles* **9**(1): 121–137.

Sastry M. 1997. Problems and paradoxes in a model of punctuated organizational change. *Administrative Science Quarterly* **42**(2): 237–275.

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

Shi S. 2002. Phantom Orders Roil Cisco's Supply Chain–A System Dynamics Study of Networking Equipment Industry's Supply Chain. Unpublished master's thesis, MIT, Cambridge, MA.

Simon H. 1963. Discussion: problems of methodology. *American Economic Review, Proceedings* **53**(May): 229–231.

———. 1979. Rational decision-making in business organizations. *American Economic Review* **69**: 493–513.

Snow CP. 1959/1993. *The Two Cultures*. Cambridge University Press: New York.

Sterman JD. 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*. Irwin/McGraw-Hill: New York.

Sterman JD, Booth Sweeney L. 2002. Cloudy skies: assessing public understanding of global warming. *System Dynamics Review* **18**(4): 207–240.

Sterman JD, Richardson GP. 1985. An experiment to evaluate methods for estimating fossil fuel resources. *Journal of Forecasting* **4**(2): 197–226.

Sterman J, Richardson G, Davidsen P. 1988. Modeling the estimation of petroleum resources in the United States. *Technological Forecasting and Social Change* **33**(3): 219–249.

Strogatz S. 1994. *Nonlinear Dynamics and Chaos, With Applications to Physics, Biology, Chemistry, and Engineering*. Addison-Wesley: Reading, MA.