Learning from Evidence in a Complex World

| John D. Sterman, PhD |

The United States spends more on health care than any other nation (15.3% of gross domestic product [GDP] in 2003, up from 5.1% in 1960). Yet the return on this huge investment is discouraging: the United States ranks 33rd in life expectancy and 35th in investment is discouraging; the United States ranks 33rd in life expectancy and 35th in infant mortality. More than 40 million have no health insurance. Minorities and the poor have significantly lower life expectancy than others. Nearly two thirds of US adults are overweight, and almost one third are obese. Diabetes and cardiovascular disease are rampant. The number of unhealthy days Americans experience is growing.

Policies to promote public health and welfare often fail or worsen the problems they are intended to solve. Evidence-based learning should prevent such policy resistance, but learning in complex systems is often weak and slow. Complexity hinders our ability to discover the delayed and distal impacts of interventions, generating unintended “side effects.” Yet learning often fails even when strong evidence is available: common mental models lead to erroneous but self-confirming inferences, allowing harmful beliefs and behaviors to persist and undermining implementation of beneficial policies.

Here I show how systems thinking and simulation modeling can help expand the boundaries of our mental models, enhance our ability to generate and learn from evidence, and catalyze effective change in public health and beyond. (Am J Public Health. 2006;96:505–514. doi:10.2105/AJPH.2005.066043)

The diligent adherence to scientific method, continued application of the scientific method, and the growing scholarly and practitioner literature on evidence in complex settings, showing how policies to thwart bioterrorists. Replication is difficult or impossible (we have only one climate and cannot compare a high–greenhouse gas [GHG] future to a low one). Decisions taken in one part of the system ripple out across geographic and disciplinary boundaries. Long time delays mean we never experience the full consequences of our actions.

Generating reliable evidence through scientific method requires the ability to conduct controlled experiments, discriminate among rival hypotheses, and replicate results. But the more complex the phenomenon, the more difficult are these tasks. Medical interventions and health policies are embedded in intricate networks of physical, biological, ecological, technical, economic, social, political, and other relationships. Experiments in complex human systems are often unethical or simply infeasible (we cannot release smallpox to test policies to thwart bioterrorists). Replication is difficult or impossible (we have only one climate and cannot compare a high–greenhouse gas [GHG] future to a low one). Decisions taken in one part of the system ripple out across geographic and disciplinary boundaries. Long time delays mean we never experience the full consequences of our actions.
irrelevant. Complexity hinders the *generation* of evidence.

Learning often fails even when reliable evidence is available. More than 2 and one-half centuries passed from the first demonstration that citrus fruits prevent scurvy until citrus use was mandated in the British merchant marine, despite the importance of the problem and unambiguous evidence supplied by controlled experiments. Some argue that today we are smarter and learn faster. Yet adoption of medical treatments varies widely across regions, socioeconomic strata, and nations, indicating either overuse by some or underuse by others—despite access to the same evidence on risks and benefits.

Although economic theory suggests market forces, publications, benchmarking, training, and imitation should cause performance to converge to optimal levels, many studies document large, persistent differences in performance across organizations. Consider cystic fibrosis. Conditions for learning are excellent: the stakes are literally life and death. The Cystic Fibrosis Foundation, NIH, and medical schools conduct research, collect clinical evidence, train specialists, and disseminate best practices. Yet although life expectancy for cystic fibrosis patients has risen significantly over the past decades, large performance differences persist across treatment centers. At the same time that many beneficial innovations diffuse slowly and unevenly, widely held superstitions remain immune to evidence (e.g., copper bracelets for arthritis treatment, “feed a cold, starve a fever,” astrology). Many of the heuristics we use to interpret evidence lead to systematically erroneous but strongly self-confirming inferences. Complexity hinders *learning* from evidence.

Many scientists respond to the complexity and learning problems by arguing that policy should be left to the experts. But this “Manhattan Project” approach (where experts secretly provide advice to inform decisions made without consulting the public or their elected representatives) fails when success requires behavior change throughout society. Effective interventions for problems from HIV/AIDS to global warming require changes in the beliefs and behaviors of a large majority of the population, supported by complementary changes in education, incentives, and institutions. Decisions once taken by experts are now seen to affect multiple stakeholders, the public at large, and future generations. People are often suspicious of experts and their evidence, believing—often with just cause—that those with power and authority routinely manipulate the policy process for ideological, political, or pecuniary purposes.

Unable to assess the reliability of evidence about complex issues on their own, and frequently excluded from the policy process, citizen noncompliance and active resistance grow, from motorcycle helmet laws to measles–mumps–rubella immunization. Complexity hinders the *implementation* of policies on the basis of evidence.

**DYNAMIC COMPLEXITY**

Most people define complexity in terms of the number of components or possible states in a system. In pharmaceutical development, for example, optimally screening new compounds for therapeutic activity is highly complex, but the complexity lies in finding the best solution out of an astronomical number of possibilities. Such needle-in-a-haystack problems have high levels of *combinatorial* complexity. However, most cases of policy resistance arise from *dynamic complexity*—the often counterintuitive behavior of complex systems that arises from the interactions of the agents over time. The text box on page 507 describes some of the characteristics of complex systems. Where the world is dynamic, evolving, and interconnected, we tend to make decisions using mental models that are static, narrow, and reductionist. Among the elements of dynamic complexity people find most problematic are feedback, time delays, and stocks and flows.
Feedback

Like organisms, social systems contain intricate networks of feedback processes, both self-reinforcing (positive) and self-correcting (negative) loops. However, studies show that people recognize few feedbacks; rather, people usually think in short, causal chains, tend to assume each effect has a single cause, and often cease their search for explanations when the first sufficient cause is found.26,27 Failure to focus on feedback in policy design has critical consequences. Suppose the hospital you run faces a deficit, caught between rising costs and increasing numbers of uninsured patients. In response, you might initiate quality improvement programs to boost productivity, announce a round of layoffs, and accelerate plans to offer new high-margin elective surgical services. Your advisors and spreadsheets suggest that these decisions will cut costs and boost income. Problem solved—or so it seems.

Contrary to the open-loop model behind these decisions, the world reacts to our interventions (Figure 1). There is feedback: our actions alter the environment and, therefore, the decisions we take tomorrow. Our actions may trigger so-called side effects that we did not anticipate. Other agents, seeking to achieve their goals, act to restore the balance that we have upset; their actions also generate intended and unintended consequences. Goals are also endogenous, evolving in response to changing circumstances. For example, we strive to earn more in a quest for greater happiness, but habituation and social comparison rapidly erode any increase in subjective well-being.26

Policy resistance arises because we do not understand the full range of feedbacks surrounding—and created by—our decisions. The improvement initiatives you mandated never get off the ground because layoffs destroyed morale and increased the workload for the remaining employees. New services were rushed to market before all the kinks were worked out; unfavorable word of mouth causes the number of lucrative elective procedures to fall as patients flock to competitors. More chronically ill patients show up in your ER with complications after staff cuts slashed resources for patient education and follow-up; the additional workload forces still greater cuts in prevention. Stressed by long hours and continual crisis, your most experienced nurses and doctors leave for jobs with competitors, further raising the workload and undercutting quality of care. Hospital-acquired infections and preventable errors increase. Malpractice claims multiply. Yesterday’s solutions become today’s problems.

Ignoring the feedbacks in which we are embedded leads to policy resistance as we persistently react to the symptoms of difficulty, intervening at low leverage points and triggering delayed and distant effects. The problem...

Causes of Policy Resistance

Policy Resistance Arises Because Systems Are

- **Constantly changing.** Heraclitus said, “All is change.” What appears to be unchanging is, over a longer time horizon, seen to vary. Change occurs at many time scales, and these different scales sometimes interact. A star evolves over billions of years as it burns its hydrogen fuel, but can explode as a supernova in seconds. Bull markets can rise for years, then crash in a matter of hours.
- **Tightly coupled.** The actors in a system interact strongly with one another and with the natural world. Everything is connected to everything else. “You can’t just one thing.”
- ** Governed by feedback.** Because of the tight coupleings among actors, our actions feed back on themselves. Our decisions alter the state of the world, causing changes in nature and triggering others to act, thus giving rise to a new situation, which then influences our next decisions.
- **Nonlinear.** Effect is rarely proportional to cause, and what happens locally in a system (near the current operating point) often does not apply in distant regions (other states of the system). Nonlinearity often arises from basic physics: insufficient inventory may cause you to boost production, but production can never fail below zero no matter how much excess inventory you have. Nonlinearity also arises as multiple factors interact in decisionmaking: Pressure from the boss for greater achievement increases your motivation and effort—up to the point where you perceive the goal to be impossible. Frustration then dominates motivation—and you give up or get a new boss.
- **History-dependent.** Many actions are irreversible: you can’t unscramble an egg (the second law of thermodynamics). Stocks and flows (accumulations) and long time delays often mean doing and undoing have fundamentally different time constants: during the 50 years of the Cold War arms race, the nuclear nations created more than 250 tons of weapon-grade plutonium (239Pu). The half-life of 239Pu is about 24,000 years.
- **Self-organizing.** The dynamics of systems arise spontaneously from their internal structure. Often, small, random perturbations are amplified and molded by the feedback structure, generating patterns in space and time. The stripes on a zebra, the rhythmic contraction of your heart, and persistent cycles in measles and the real estate market all emerge spontaneously from the feedbacks among the agents and elements of the system.
- **Adaptive and evolving.** The capabilities and behaviors of the agents in complex systems change over time. Evolution leads to selection and proliferation of some agents while others become extinct. People adapt in response to experience, learning new ways to achieve their goals in the face of obstacles. Learning is not always beneficial, however, as it tends to overgeneralize, parochial, maximizing local, short-term objectives at the expense of long-term success.
- **Characterized by trade-offs.** Time delays in feedback channels mean the long-run response of a system to an intervention is often different from its short-run response. Low-leverage policies often generate transitory improvement before the problem grows worse, whereas high-leverage policies often cause worse-before-better behavior.
- **Counterintuitive.** In complex systems, cause and effect are distant in time and space, whereas we tend to look for causes near the events we seek to explain. Our attention is drawn to the symptoms of difficulty rather than the underlying cause. High-leverage policies are often not obvious.
- **Policy resistant.** The complexity of the systems in which we are embedded overwhelms our ability to understand them. The result: many seemingly obvious solutions to problems fail or actually worsen the situation.
Public health systems are not immune to time delays: the belief that we are helpless victims of forces over which we have no control. Policy makers work hard to do so in a way that empowers, rather than reinforces, a community norm of blaming others, “the decision to build) and its effects on the state of the system. As a result, decision makers often continue to intervene to correct apparent discrepancies between the desired and actual state of the system even after sufficient corrective actions have been taken to restore equilibrium. The result is overshoot and oscillation: stop-and-go traffic, drunkenness, and high-tech boom and bust cycles.29

Public health systems are not immune to these dynamics, from oscillations in incidence of infectious diseases, such as measles30 and syphilis,31 to the 2004–2005 flu vaccine fiasco, with scarcity and rationing followed within months by surplus stocks.32

**Time delays**

Time delays in feedback processes are common and particularly troublesome. Most obviously, delays slow the accumulation of evidence. More problematic, the short- and long-run impacts of our policies are often different (smoking gives immediate pleasure, while lung cancer develops over decades). Delays also create instability and fluctuations that confound our ability to learn. Driving a car, drinking alcohol, and building a new semiconductor plant all involve time delays between the initiation of a control action (accelerating/braking, deciding to “have another,” the decision to build) and its effects on the state of the system. As a result, decision makers often continue to intervene to correct apparent discrepancies between the desired and actual state of the system even after sufficient corrective actions have been taken to restore equilibrium. The result is overshoot and oscillation: stop-and-go traffic, drunkenness, and high-tech boom and bust cycles.29

**FIGURE 1—Sources of policy resistance.**

Note. Arrows indicate causation, e.g., our actions alter the environment. Thin arrows show the basic feedback loop through which we seek to bring the state of the system in line with our goals. Policy resistance (thick arrows) arises when we fail to account for the so-called “side effects” of our actions, the responses of other agents in the system (and the unanticipated consequences of these), the ways in which experience shapes our goals, and the time delays often present in these feedbacks.

**Greener Vehicles**

California’s Air Resources Board seeks to reduce air pollution from transportation.33 True, zero emission vehicles need no tailpipe. But the plants required to make the electricity or hydrogen to run them do generate pollution. California is actually promoting displaced emission vehicles, whose wastes would blow downwind to other states or accumulate in nuclear waste dumps outside its borders. Air pollution causes substantial mortality, and fuel cells may prove to be an environmental boom compared with internal combustion. But no technology is free of environmental impact, and no legislature can repeal the second law of thermodynamics.

Second, people have poor intuitive understanding of the process of accumulation. Most people assume that system inputs and outputs are correlated (e.g., the higher the federal budget deficit, the greater the national debt will be).34 However, stocks integrate (accumulate) their net inflows. A stock rises even as its net inflow falls, as long as the net inflow is positive: the national debt rises even as the deficit falls—debt falls only when the government runs a surplus; the number of people living with HIV continues to rise even as incidence falls—prevalence falls only when infection falls below mortality. Poor understanding of accumulation has significant consequences for public health and economic welfare. Surveys show most Americans believe climate change poses serious risks, but they also believe that reductions in GHG emissions sufficient to stabilize atmospheric GHG concentrations can be deferred until there is greater evidence that climate change is harmful.35 Federal policy makers likewise argue that it is prudent to wait and see whether climate change will cause substantial economic harm before undertaking policies to reduce emissions.36 Such wait-and-see policies erroneously assume that climate change can be reversed quickly should harm become evident, underestimating immense delays in the climate’s response to GHG emissions. Emissions are now about twice the rate at which natural processes remove GHGs from the atmosphere.37 GHG concentrations will therefore continue to rise even if emissions fall, stabilizing only when emissions equal removal. In contrast, experiments with highly educated adults—graduate students at MIT—show that most believe atmospheric GHG concentrations can be stabilized while emissions into the atmosphere continuously exceed the removal of GHGs from it.35 Such beliefs are analogous to arguing that a bathtub filled faster than it
drains will never overflow. They violate conservation of matter, and the violation matters: wait-and-see policies guarantee that atmospheric GHG concentrations, already greater than any in the past 420,000 years, will rise far higher, increasing the risk of dangerous changes in climate that may significantly harm public health and human welfare.

**BARRIERS TO LEARNING**

Just as dynamics arise from feedback, so too all learning depends on feedback. As we perceive discrepancies between desired and actual states, we take actions that (we believe) will cause the real world to move toward the desired state. New information about the state of the world causes us to revise our perceptions and the decisions we make in the future. When driving, I may turn the steering wheel too little to bring the car back to the center of my lane, but as visual feedback reveals the error, I continue to turn until the car returns to the straight and narrow. Such single-loop learning is shown in the top of Figure 2.

Information feedback about the real world is not the only input to our decisions. Decisions are the result of applying a decision rule or policy to information about the world as we perceive it. These policies are conditioned by institutional structures, organizational strategies, and cultural norms, which, in turn, are shaped by our mental models (see the bottom of Figure 2). Single-loop learning is the process whereby we learn to reach our current goals in the context of our existing mental models. Single-loop learning does not result in deep change in our mental models—the time horizon we consider relevant—nor in our goals and values.

Deep change in mental models, or double-loop learning, arises when evidence not only alters our decisions within the context of existing frames, but also feeds back to alter our mental models. As our mental models change, we change the structure of our systems, creating different decision rules and new strategies. The same information, interpreted by a different model, now yields a different decision. Systems thinking is an iterative learning process in which we replace a reductionist, narrow, short-run, static view of the world with a holistic, broad, long-term, dynamic view, reinventing our policies and institutions accordingly.

For learning to occur, each link in the single- and double-loop learning processes must work effectively, and we must be able to cycle around the loops faster than changes in the real world render existing knowledge obsolete. Yet these feedbacks often do not operate well. Each link in the learning loops can fail (Figure 2).

**Limited information and ambiguity**

We experience the real world through filters. No one knows the current incidence or prevalence of any disease. Instead surveillance systems report estimates of these data on the basis of sampled, averaged, and delayed measurements. The act of measurement introduces distortions, delays, biases, errors, and other imperfections, some known, others unknown and unknowable.

Above all, measurement is an act of selection. Our senses and information systems select but a tiny fraction of possible experience. We define GDP so that medical care caused by pollution-induced disease adds to the GDP, whereas the production of the pollution itself does not reduce it. Because the prices of most goods do not include the costs and consequences of environmental degradation and resource depletion, these externalities receive little weight in policymaking.

Of course, the information systems governing the feedback we receive can change as we
learn. Figure 2 also shows feedback between mental models and the information feedback available to us: seeing is believing and believing is seeing. Through our mental models, we define constructs, such as GDP, and design systems to evaluate and report them. We conflate what is salient, tangible, and familiar with what is important. As we measure these things, they become even more real, whereas the remote effects of our decisions, the unfamiliar, and the intangible fade like wraiths. Thus we confuse the military budget with security, GDP per capita with happiness, and the size of our houses with the quality of our home life.

The self-reinforcing feedback between expectations and perceptions has been repeatedly demonstrated. Sometimes the positive feedback assists learning by sharpening our ability to perceive features of the environment, as when an experienced naturalist identifies a bird in a distant bush where the novice sees only a tangled thicket. Often, however, the mutual feedback of expectations and perception blinds us to the anomalies that might challenge our mental models and lead to deep insight.

As one of many examples, consider the history of ozone depletion by chlorofluorocarbons (CFCs). The first evidence describing the ability of CFCs to destroy atmospheric ozone was published in 1974. Industries dependent on CFCs argued that uncertainty in the evidence warranted inaction. Despite a ban on CFCs as aerosol propellants, global production of CFCs remained near its all-time high. It was not until 1985 that evidence of the Antarctic ozone hole was published. As described by Meadows, Meadows, and Randers:

The news reverberated around the scientific world. Scientists at NASA scrambled to check readings on atmospheric ozone made by the Nimbus 7 satellite, measurements that had been taken routinely since 1978. Nimbus 7 had never indicated an ozone hole.

Checking back, NASA scientists found that their computers had been programmed to reject very low ozone readings on the assumption that such low readings must indicate instrument error.

Scientists’ preconceptions about “normal” ozone concentrations led them to design a measurement system that made it impossible to detect evidence that might have shown that belief to be wrong. Fortunately, NASA had saved the original, unfiltered data, and later confirmed that ozone concentrations had indeed been falling since the launch of Nimbus 7. By creating a measurement system immune to disconfirmation, the discovery of the ozone hole and resulting global agreements to cease CFC production were delayed by as much as 7 years.

**Bounded rationality and the misperceptions of feedback**

Humans are not computers, coolly assessing possibilities and probabilities. Emotions, reflex, unconscious motivations, and other nonrational or irrational factors all play a large role in our judgments and behavior. But even when we find the time to deliberate, we cannot behave in a fully rational manner (that is, make the best decisions possible given the available information). As marvelous as the human mind is, the complexity of the real world dwarfs our cognitive capabilities. Herbert Simon articulated these limits in his famous principle of “bounded rationality,” for which he won the Nobel Memorial Prize in economics in 1978:

The capacity of the human mind for formulating and solving complex problems is very small compared with the size of the problem whose solution is required for objectively rational behavior in the real world or even for a reasonable approximation to such objective rationality.

Faced with the overwhelming complexity of the real world, time pressure, and limited cognitive capabilities, we are forced to fall back on rote procedures, habit, rules of thumb, and simple mental models. Although we sometimes strive to make the best decisions we can, bounded rationality means that we often systematically fall short.

Bounded rationality is particularly acute in dynamic systems. Experiments show that people do quite poorly in systems with even modest levels of dynamic complexity; for example, creating business cycles, bankrupting their companies, depleting renewable resources, and delaying medical treatment while (simulated) patients sicken and die. These misperceptions of feedback are robust to experience and financial incentives.

Among the most damaging misperceptions is the tendency to attribute the behavior of others to dispositional rather than situational factors; that is, to character and especially character flaws rather than the system in which they are embedded—the “fundamental attribution error.” The atrocities at Abu Ghraib were blamed on a few bad apples, whereas decades of research, from Milgram’s obedience studies and the Stanford prison experiment on, demonstrate that “it’s not the apples, it’s the barrel.” Despite overwhelming evidence that our behavior is molded by pressures created by the systems in which we act, problems such as the failure of patients to stay on their medications, recidivism among drug users, and childhood obesity, are persistently attributed to the undisciplined personal habits, poor attitude, or low intelligence of these “others.” The focus becomes scapegoating and blame, and policy centers on controls to force compliance. Blame and attempts to control behavior provoke resistance and patient drop-out, strengthening the erroneous belief that these people are unreliable incompetents requiring still greater monitoring and control. Recognizing the power of system structure to shape behavior does not relieve us of personal responsibility for our actions. To the contrary, it enables us to focus our efforts where they have highest leverage—the design of systems in which ordinary people can achieve extraordinary results.

**Poor inquiry skills**

Learning effectively in a world of dynamic complexity requires dedicated application of scientific method. Unfortunately, people are poor intuitive scientists. We do not generate alternative explanations or control for confounding variables. Our judgments are strongly affected by the frame in which the information is presented, even when the objective information is unchanged. We suffer from overconfidence in our judgments (underestimating uncertainty), wishful thinking (assessing desired outcomes as more likely than undesired outcomes), and confirmation bias (seeking evidence consistent with our preconceptions). Scientists and professionals, not only “ordinary” people, suffer from many of these judgmental biases. Some argue that, while people err in applying the principles of logic, at least they
appreciate the desirability of scientific explanation. Unfortunately, the situation is far worse. The scientific worldview is a recent development in human history, and remains rare. Many place their faith in what Dos- toyevsky’s Grand Inquisitor called “miracle, mystery, and authority” for example, astrology, creationism, Elvis sightings, and cult leaders promising Armageddon. The persistence of such superstitions is strongly self-reinforcing: during his career with the Boston Red Sox, Hall of Fame hitter Wade Boggs ate chicken every game-day for years because he once played particularly well after a dinner of lemon chicken. While on the chicken diet, which he came to loathe, Boggs won five batting championships, decisively proving the “chicken theory.”

Such foolishness aside, there are more disturbing reasons for the prevalence of these learning failures. Human beings are more than cognitive information processors. We have a deep need for emotional and spiritual sustenance. But from Copernican heliocentrism through relativity, quantum mechanics and evolution, science has stripped away ancient and comforting beliefs, placing humanity at the center of a world designed for us by a supreme authority. For many people, science leads not to enlightenment and empowerment, but to existential angst and the absurdity of human insignificance in an incomprehensible vast universe. Others believe science and technology are the shock troops for the triumph of materialism and instrumentalism over the sacred and spiritual. These antiscientific reactions are powerful forces. In many ways, they are important truths. They have led to many of the most profound works of art and literature. But they can also lead to mindless new-age psychobabble and radical fundamentalism. Readers should not conclude that I am a naïve defender of science as it is practiced nor an apologist for the real and continuing damage done to the environment and to our cultural, moral, and spiritual lives in the name of rationality and progress. On the contrary, I have stressed the research showing that scientists are often as prone to error and bias as lay people. It is precisely because scientists are subject to the same cognitive limitations and moral failures as others that we experience abominations such as the Tuskegee experiment. Systems thinking requires us to examine issues from multiple perspectives, to expand the boundaries of our mental models, to consider the long-term consequences of our actions, including their environmental, cultural, and moral implications.

**Defensive routines and implementation failure**

Learning by groups can be thwarted even if participants receive excellent information and reason well as individuals. Argyris and Schön, Janis, and others document the defensive routines people rely on, often unknowingly, in interpersonal interactions. We use defensive routines to save face, make untested inferences seem like facts, and advocate our positions while appearing to be neutral. We make strong attributions not grounded in data. We avoid publicly testing our beliefs, tacitly communicating that we are not open to having our mental models challenged. Defensive routines often yield group-think, as members of a group mutually reinforce their current beliefs, suppress dissent, and seal themselves off from those with different views or possible disconfirming evidence.

Even if a team were united in recommending the proper course of action, the implementation of their decisions is often distorted by asymmetric information, private agendas, and game playing by agents throughout a system. Obviously, implementation failures can hurt an organization. Imperfect implementation can hinder learning as well, because the managers evaluating the outcomes of their decisions may not know the ways in which those decisions were distorted, delayed, or derailed altogether by other actors in the system.

Finally, because error is often costly and many decisions are irreversible, the need to maintain performance often overrides the experimentation needed to learn. It’s important for pilots to learn how steep a dive their aircraft can handle in case an emergency requires rapid descent. But no pilot would try a maximum dive on the 10 o’clock to Chicago just to learn. Even when the consequences of experiments are mild, however, the fear of failure, of appearing to have made a mistake, often stifles innovation. Voltaire advised that we “love truth and pardon error,” but the desire to avoid embarrassment regularly suppresses deviations from standard practice that might reveal opportunities for improvement.

**VIRTUES OF VIRTUAL WORLDS**

To learn effectively in a world of dynamic complexity, we must attend to all these impediments (Figure 3). The figure features a new feedback loop created by the use of virtual worlds. Virtual worlds are models or simulations in which decision makers can conduct experiments, rehearse decisionmaking, and play. They can be physical models, role-plays, or computer simulations. In systems with significant dynamic complexity, computer simulation will typically be needed.

Simulations provide low-cost laboratories for learning. The virtual world allows time and space to be compressed or dilated. Actions can be repeated under the same or different conditions. One can stop the action to reflect. One can make decisions that are infeasible or unethical in the real system. Participants can receive perfect and immediate outcome feedback. In an afternoon one can gain years of simulated experience. In contrast to the real world, which, like a black box, has a poorly resolved structure, virtual worlds can be open boxes whose assumptions are known and can be modified by the learner. Often, pushing a system into extreme conditions reveals more about its structure and dynamics than incremental adjustments to current practices. Thus a great deal of the time that pilots spend in flight simulators is devoted to extreme conditions, such as engine failure. In the virtual world, you can find the maximum dive angle—“crashing” hurts no one, and you walk away every time, better prepared for a real emergency.

Dubbed the “third branch” of science (after theory and experiment), simulation is now an essential tool in research on problems from galaxy formation to protein folding to epidemiology. Virtual worlds for learning and training are commonplace in the military, pilot training, power plant operations, and other tasks. The use of virtual worlds and simulation models in public policy and management is more recent and less widely adopted. Yet these are precisely the settings in which dynamic complexity is most problematic.
learning feedbacks are least effective, and the stakes are highest.

**PITFALLS OF VIRTUAL WORLDS**

To be effective, a virtual world must capture those aspects of the real system of concern to the decision makers with sufficient fidelity for their purpose. In addition, the user interface must enable people to learn from the model. The most insightful model accomplishes nothing if the interface is obscure and the protocol for its use ineffective. The converse is worse: a poor model embedded in a potent interface may teach harmful lessons more effectively than ever before. Effective virtual worlds require both substantive fidelity and a productive learning process that enables people to challenge and improve their mental models.

How can the substantive quality of a model be assessed? System dynamics emphasizes a multifaceted process for testing models, identifying errors, and comparing model assumptions and behavior to data. The process of model testing and improvement is iterative. Discrepancies between mental models, formal models, and data stimulate improvements in each.\(^{64,65}\) For the testing process to be effective, models must be fully documented so that independent third parties can replicate the results, carry out sensitivity analysis, try alternative theories, and subject the model to extreme conditions. Space does not permit a full treatment of these tests; readers should consult the extensive literature for principles and examples.\(^9\)

Although simulation models may be necessary for effective learning in dynamically complex systems, they are not sufficient to overcome the flaws in our scientific reasoning skills and group processes. Although a virtual world enables controlled experimentation, most policy makers lack training in scientific method and the design of experiments. A commonly observed behavior in virtual worlds is the “video game syndrome” in which people play too much and think too little. People often do not take time to reflect on the outcome of a simulation, identify...
discrepancies between outcomes and expectations, formulate hypotheses to explain the discrepancies, and then devise experiments to discriminate among competing theories. Defensive routines and groupthink can operate in the learning laboratory just as in real organizations. Indeed, protocols for effective learning in virtual worlds such as public testing of hypotheses can be highly threatening, inducing defensive reactions that prevent learning. Managers unaccustomed to disciplined scientific reasoning in an open, trusting environment will have to build these skills before a virtual world can prove useful.

CONCLUSIONS

Policies to promote public health and welfare often fail or worsen the problems they are intended to solve. Evidence-based learning should prevent such policy resistance, but learning in complex systems is often weak and slow. Complexity hinders our ability to discover the delayed and distal impacts of interventions, generating unintended “side effects.” Yet learning often fails even when strong evidence is available: common mental models lead to erroneous but self-confirming inferences, allowing harmful beliefs and behaviors to persist and undermine implementation of beneficial policies. When evidence cannot be generated through experiments in the real world, virtual worlds and simulation become the only reliable way to test hypotheses and evaluate the likely effects of policies. Most important, when experimentation in real systems is infeasible, simulation is often the only way we can discover for ourselves how complex systems work. Without the rigorous testing enabled by simulation, it becomes all too easy for policy to be driven by ideology, superstition, or unconscious bias. The alternative is rote learning on the basis of the authority of an expert, a method that dulls creativity and stunts the development of the skills needed to catalyze effective change in complex systems.

When humans evolved, the challenge was survival in a world we could barely influence. Today, the hurricane and earthquake do not pose the greatest danger. It is the unanticipated effects of our own actions, effects created by our inability to understand the complex systems we have created and in which we are embedded. Creating a healthy, sustainable future requires a fundamental shift in the way we generate, learn from, and act on evidence about the delayed and distal effects of our technologies, policies, and institutions. The reductionist program of ever-finer specialization is no longer sufficient. Though often leading to deep and useful knowledge, it contributes to policy resistance by narrowing the boundaries of our mental models. As leaders in public health, you do not face medical problems, financial problems, technical problems, and community relations problems. You just have problems. Some boundaries are necessary and inevitable: all models must simplify the overwhelming complexity of the world. But all too often ignoring what lies outside familiar walls cuts critical feedbacks and breeds arrogance about our ability to control nature and other people—and we solve one problem only to create others.

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 tools to understand complexity, stocks and flows, feedback, and time delays. It requires the use of virtual worlds and simulations to augment the evidence generated by experiments in the real world. It requires an unwavering commitment to the rigorous application of scientific method, and the inquiry skills we need to expose our hidden assumptions and biases. It requires crossing boundaries between departments and functions in an organization, between disciplines in the academy, between the private and public sector. It requires breaching barriers of culture and class, race and religion. It requires listening with respect and empathy to others—then using these systems thinking capabilities to act in consonance with our long-term goals and deepest aspirations.

About the Authors
The author is with the MIT Sloan School of Management, Cambridge, Mass.
Requests for reprints should be sent to John Sterman, MIT Sloan School of Management, 30 Wadsworth Street, Room ES5-351, Cambridge Massachusetts 02142 (e-mail: jsterman@mit.edu).
This article was accepted June 1, 2005.

Acknowledgments
Financial support was provided by the Project on Innovation in Markets and Organizations at the MIT Sloan School of Management.
This paper builds on and extends the argument in Chapter 1 of Sterman (2000). I thank Jack Homer, Scott Leischow, Bobby Milstein, Mary Northridge, Kim Thompson, and the referees for helpful suggestions.

Human Participant Protection
No human participants were involved in this study.

References
15. Tengs TO, Ahma S, Savage JM, Moore R, Gage E. The AMA proposal to mandate nicotine reduction in...