Richard J. Zeckhauser, Ralph L. Keeney, & James K. Sebenius eds. Wise Choices: Decisions, Games, and Negotiations (Boston: Harvard Business School Press, 1996)

CHAPTER 14

WHAT CAN YOU LEARN FROM A GAME?

Edward A. Parson

This paper considers ways that simulation games can be used to inform difficult and complex problems of policy and decision. The use of games for serious purposes, as opposed to recreation, originated more than a century ago in the field of military planning; manual or map games served as devices to test battle plans, coordinate actions, train commanders, and predict the outcome of battles.[1] In the postwar period, interest in gaming has increased, driven partly by advances in game theory and computing power, and has broadened from military planning to the study of foreign policy crises and, occasionally, other areas of policy and decision.[2]

Despite several substantial contributions to the field, questions concerning the appropriate uses and goals of simulation gaming remain confused, and discussions of purposes seem to have had little influence on the practice of gaming or game development.[3] Several factors may have contributed to this confusion. Writers on game purposes have sought to construct extremely broad taxonomies including essentially all games and design methods, including those for recreational and commercial use, while the viability and implications of particular objectives likely depend strongly on the particular method and application. Categories of use have not been sharply drawn, so some applications of games fall into more than one, or none. Because most gaming experience is still in the military and security fields, discussions of uses have tended to emphasize those likely to arise in this field, while others less likely to do so have received less careful consideration. Finally, the relationship between purposes of games and criteria for their design and implementation has not been elaborated.

This paper seeks to elaborate and clarify a particular set of uses of simulation gaming applied to a particular class of problems: informing complex, difficult, and high-stakes policy and decision problems, particularly problems outside the field of military and security affairs. Problems that might merit investigation through simulation gaming are those with high enough stakes to merit substantial investment in knowledge or insight to guide decision making, but whose complexity, novelty, or perverse characteristics impose sharp limits on the usefulness of standard decision-making procedures, historical analogy, or conventional forms of analysis. In "informing" such problems we include all heuristic purposes of simulation gaming: uses of gaming intended to help actors with significant responsibilities to understand decisions they may face and their implications, or to understand the context in which continuing decisions take place.

The characteristics of these problems that make them so difficult, and that limit the usefulness of other forms of analysis, can be of several kinds. Key outcomes may depend upon the interacting decisions of multiple agents, which may be particularly complex when agents' interests are partly aligned and partly in opposition. Decisions may be drawn from choice sets that are ambiguous, poorly known, or changing. Large numbers of complex organizational routines may be required to work together.[4] Consequences may be poorly known, fall on unidentified or remote people, or occur far in the future. Or finally, knowledge about agents' interests and choices, and relevant properties of the world, may be seriously incomplete or contested.

Two classes of such problems are particularly salient: the management of rapidly developing crises, and problems involving major changes in laws, regulations, or institutions. Crisis management is the area of gaming application that most closely resembles its traditional use in military planning, but can also include, for example, disaster preparedness, energy supply interruption, and international macroeconomic or debt crises.[5] Problems of decision making under major changes of rules or institutions could include corporate strategy in industries in transition; domestic policy reform in health care or welfare; long-horizon policy issues such as public pension reform under demographic shifts; the fundamental redesign of institutions, policies, and laws, as has been under way for several years in Eastern Europe and the former Soviet Union; and international policy making on imperfectly understood but potentially grave global environmental threats such as climate change.[6]

Focusing attention on this limited set of applications of simulation gaming—applications to inform complex policy or decision problems—this paper discusses four models of how simulation gaming can contribute, seeking to assess the relative strength and weakness of each, and to sketch out the implications of each for the design and implementation of simulation games. The next section outlines the basic characteristics of simulation-gaming methods, and their central methodological challenges. Subsequent sections lay out and assess the four models, summarize the implications of each for simulation design, and offer concluding observations on the benefits and limits of informing complex policy and decision problems through simulation gaming.

Simulation Gaming: Basic Characteristics and Challenges

Simulation games, like other forms of simulation, are representations; they seek to represent a complex system by constructing a simpler one with relevant behavioral similarity.[7] The behavioral similarity to the real system of concern makes a simulation useful, for if it is realized, one can learn about the real system by observing, and manipulating, the simpler simulated one.

Simulation gaming differs from other kinds of simulation in how it represents human decision making and problem solving. While other kinds of simulation may exclude human agency from the system studied, or represent it entirely through formal or computer models, simulation gaming represents it through the behavior of human participants. It seeks to represent the information processing, cognition, negotiation, and decision making of senior decision makers, organizations, or governments, by engaging people to participate in a simulation, giving them roles, information, tasks, and responsibilities, and placing them in a vivid, demanding, realistic situation in which they must act.

While some simulations may employ a single human participant interacting with a complex mechanical system—e.g., flight simulators used for pilot training, practice, and testing control systems—simulation gaming involves multiple participants and focuses on their collective decisions and the interactions among them. While the interactions of decisions can be structured in many ways, two characteristics are generally deemed essential to making a simulation a game: rules structuring the interactions among participants' decisions; and separate centers of thought or decision (teams or individuals) that do not observe each others' planning and deliberation, and so must interpret and respond to statements and actions that they did not compose.[8]

The use of multiple human participants in a structured interaction without full information defines a simulation game; other design elements, which support and constrain participants' decisions, can vary widely. Simulation games may include rich narrative "scenarios" that provide context for, and identify the essential elements in, participants' decision problems;[9] some combination of an expert control team, formal models, or fixed rules that determine the consequences of participants' joint decisions;[10] or general information resources or formal planning and analysis tools to support participants' deliberations.[11]

Lines of simulation gaming work include political-military exercises, crisis gaming, policy exercises, adaptive environmental assessment and management, and the experimental negotiation studies pioneered by Howard Raiffa.[12] Since the broadening of interest in gaming in the 1950s, the history of simulation gaming has been marked by great promise, cyclic periods of enthusiasm and reaction, and instances of both excessive claims by advocates and impossible standards imposed by detractors.[13]

Achieving the necessary degree of "relevant behavioral similarity" to the real problem to be investigated is the essential problem of designing and using a simulation-gaming exercise. How much of what kind of similarity is required, and how persuasively it must be demonstrated, depend on the particular application and objectives for the exercise.

Possible requirements for achieving it may include participants who sufficiently resemble the decision makers they represent; reasonably realistic data to define the scope and context for decision in the simulation; and roles that are both sufficiently accurately portrayed, and sufficiently intense and vivid, that participants act on them, exert themselves, and worry about the consequences of their actions. It is essential, for example, that participants' interests in the social context of the simulation itself—their desire to win, to make mischief, or to be a good citizen—do not overwhelm or trivialize the interests of their roles or teams.[14] Finally, certain unavoidable artifacts of simulation—arbitrary time, the heightened salience of the issue studied, and participants' limited information-processing ability compared to the organizations or governments they represent—must not so dominate the simulation that all significant insights are attributable to them.[15]

Proposed Models for the Use of Simulation Gaming

This section lays out four possible models of how simulation games can help inform complex decision and policy problems. These differ in the kind of things to be learned, by whom, and through what mechanism, and they imply different design criteria for the development and implementation of simulations. Two of these models are problematic, while two show substantial promise.

Simulations as Experiments

It is sometimes suggested that simulations can serve as experiments, or quasi-experiments, to test hypotheses about the behavior of people, negotiations, organizations, decision makers in crisis, governments, or the international system. This capability, if realized, would represent a great increase in our power to understand the world. Simulations could serve as artificial worlds to test conjectures for which history failed to offer sharp empirical cases, and for which real-world experimentation would be impossible, too costly, too risky, or morally prohibited. The potential benefits for both enhanced understanding and better decisions and policies would be vast.[16]

The claim of experimental validity imposes stringent standards on the correspondence of a simulation to reality. The simulation must embody enough behavioral similarity to the situation investigated that hypotheses about the real situation imply expected behavior in the simulation, and hence that running the simulation exposes the hypothesis to the risk of falsification that can suggest the limits of its credibility, and increase our provisional willingness to accept it and act on its implications.[17]

In meeting this standard, the simulation itself, like a theory, must be validated; its "relevant behavioral similarity" to the real system must be demonstrated with sufficient confidence. This test can in part be met by validation of the simulations' components: participants enough like the real decision makers, and roles, structures, and rules that embody well verified relationships. But simulations are so complex by standards of experiments that full correspondence cannot even be approximately achieved, and how much similarity is enough is judgmental, not obvious, and sometimes contested. Simulations can fail to achieve the required similarity in many ways: observed behaviors may be artifacts of the particular individuals participating, of arbitrary details of simulation design, or of the unavoidable artificiality inherent in any simulation.

For example, how much and in what ways must simulation participants resemble real decision makers? It is frequently argued that simulation validity requires participants who are real experts, equivalent in knowledge, skills, authority, resolve, and temperament to those who would make decisions in the real system, lest ludicrous behavior occur.[18] Others, also reporting simulation experience, contend that if psychological profiles are well matched then high-school students can in some respects reliably replicate behavior representative of generals and cabinet ministers; that there are only insignificant differences between the performance of middle and top-level officials, and between military and civilian officials; and that if tasks are appropriately defined and intensely demanding, then any participants sufficiently smart, skilled, and motivated will generate instructive and relevant behavior, even if they know little about the particular decision makers they represent or their institutional setting.[19]

Since the validity of a simulation cannot be determined absolutely from the representational accuracy of its components, validation of a simulation must depend on controlled replication. Simulations, to persuade, must generate behavior that resembles the situation of interest, in a manner that persuades knowledgeable and skeptical people that the similarity is solid, reliable, and not obtained through cheating. Replication must be conducted with different participants, to verify that behavior is not an artifact of a particular group; and with variation in design details, to verify that behavior is not an artifact of some particular arbitrary design feature. Except for intentional variation of the factors being investigated, this replication must be closely controlled.

The need for controlled replication strongly shapes simulation design. Sparser decision environments are preferable, to reduce the number of potentially confounding factors. Moreover, while the decision setting might appear quite open-ended, participants' license to redefine the problem or change the rules must in fact be quite limited, lest their simulation become noncomparable with others. Authority over the simulation remains strongly with the experimenters, while participants serve largely as experimental subjects.

This replication requirement is a daunting task, but one characteristic of simulation slightly offsets it. Because of simulations' richness and complexity, their outputs contain huge amounts of information. The main narrative observed in a simulation, and participants' observations, admit detailed examination at both large and small scales, and so include many opportunities to demonstrate either realism or unrealism. Thus, simulations provide many more opportunities to establish validity than if each simulation run were treated as a single datapoint.[20]

Despite this partially offsetting factor, there are two basic difficulties with the model of simulations as experiments, one conceptual, and one practical and often decisive.

The conceptual difficulty is a basic conflict between the presumed novelty and uniqueness of the policy problem to be investigated, and the pursuit of generalizable hypotheses bearing on it. Simulations treated as experiments seek to identify and test regularities that generalize across instances. To this end, their focus even in novel situations is on aspects of the situation that are potentially generalizable and

subject to abstraction. This inevitably leads them to concentrate on regularities in the normal course of policy and decision (e.g., general principles of foreign policy and international relations),[21] or on small components of the situation that can be abstracted, and that might occur in many situations. Relevant examples include studies of simple abstract negotiations, prisoner's dilemmas or commons problems, and simple coalition exercises, plus a variety of psychological research in individual decision making and communication. Such studies are more easily replicable because they are less elaborate, costly, and time consuming, and because they depend less on domain-specific expertise and so admit a much wider pool of potential participants.

While simulations as experiments tend to address those aspects of a situation that are most normal and replicable, the essential difficulties associated with complex policy problems concern their uniqueness and novelty. Those aspects that are liable to treatment by simulations as experiments are thus likely to be peripheral to the main difficulties for which insights are required. Of course, no issue, decision, or policy problem can be unique in all its dimensions, and it may be extremely valuable to decompose novel, unique problems into familiar or routine components, to the extent possible. But important elements of the real problem are likely to remain unaddressed, and even if well validated insights into component pieces are attained, the experimental approach cannot recombine the component insights to yield aggregate judgments about the issue. Even if simulation experiments lead us to understand the separate 10 elements that comprise a problem, we have no information about how to combine these insights, and make judgments about the relative influence of the different factors, in the aggregate problem.

The practical difficulty with simulations as experiments concerns the low likelihood of sufficiently replicating an experimental simulation with the right participants, or to control it tightly enough, to achieve a persuasive level of validity. To the extent that participants must be expert in the issue studied, the required participants are likely to be few in number, busy, not substitutable, and likely unwilling to participate in an exercise that casts them as passive experimental subjects. The people who must participate for the exercise to have validity as research are unlikely to participate in an exercise tightly designed for research. The more unique and novel the problem to be studied, the sharper this difficulty is likely to be.

In summary, the model of simulation games as experiments is deeply problematic as a description of how such exercises can help to inform complex policy and decision problems. Simulations can act as experiments, but principally for simple abstract repeatable situations that may occur as components of complex policy problems but are unlikely to represent their most important elements. It may happen that a particular insight arises cogently enough, in enough varied simulations, and with enough external basis for confidence in it that it can come to be accepted as generalizable; but this cannot describe the normal use of simulation games to inform policy.[22]

Simulations to Instruct Decision Makers

Every discussion of simulation games highlights their value as vehicles for instruction. When simulations are applied to complex policy problems, several forms of instruction may be relevant. Of these, several are persuasive and potentially important, but normally tangential to the core of what makes the problem investigated difficult. Instruction in the essential character of the problem itself may occasionally be valid, but is usually deeply problematic.

Simulations can offer instruction on many levels. Because of their vividness, they can be rapid vehicles for learning large volumes of prosaic facts about an issue, an organization, or a location. Because they involve intense collective work by teams of diversely skilled and knowledgeable experts, they can be powerful devices for the transfer of skills and knowledge among participants. They can instruct in such generic skills as processing large volumes of information, clear communication, negotiation, group problem solving, and decision making under uncertainty. These are all important, potentially valuable forms of instruction. All exploit the vividness and concreteness of simulation to provide a learning experience more salient, durable, or workable than may be available from other means. But none of these bears on the core of the complex problem being investigated. Indeed, these forms of instruction depend very little on the representational accuracy of the simulation. It need only be persuasive enough to get participants engaged in their tasks, and be supported by accurate background information. The simulation structure may be little more than a pretext to bring the right set of participants together.

Seeking to use simulation gaming to instruct decision makers in the essential character of a complex policy problem is a much more problematic enterprise. This mode of instruction presumes that the simulation structure embodies insights into the essence of the problem that are true, important, relatively enduring, and not dominated in practical settings by other factors that the participants understand better than the analysts, and consequently that decision makers should understand them and would act differently if they did.

These conditions are likely to be very seldom attained. Specific pieces of policy-relevant scientific knowledge will seldom warrant this

form of instruction, partly because they are liable to major revisions. (For example, recent advances in understanding the radiative effects of chlorofluorocarbons and sulfate aerosols have sharply altered the policy debate around global climate change.) The insights that may merit such instruction will typically be deep insights into the basic structure of a problem. Examples might include the basic logic of nuclear deterrence, or the implications of the long time-horizon of global climate change.

Simulations to Promote Creativity and Insights

Simulations can be powerful devices to provoke participants to think creatively about the problem, and hence to promote the generation of new ideas and insights. Simulations can achieve this through their pressure, relevance, and distance. A well-designed simulation game is an intense, demanding experience for the participants, and should resemble their real responsibilities closely enough to engage their professional skills and knowledge, but not so closely that they rely on their standard habits, heuristics, and bargaining positions. The experience of uncertainty that comes from teams' not knowing what other teams plan or intend may also serve to push people away from habitual ways of thinking.

The insights that may arise can be of many kinds. They might include changed views of the relative importance of different aspects of a problem, including seeing that something previously overlooked is important; new ideas for negotiating stances, policy design, institutions, or responses to specific contingencies; and ideas about plausible consequences of specific proposals, including unanticipated potential pitfalls. They may include practical ideas for officials, and hypotheses for researchers. Thomas Schelling reports that a participant realized that the supply of jet fuel in Teheran was 10 times larger than had been thought, because kerosene (an acceptably close substitute) was used for domestic cooking.[23] Parson reported that when participants sought to revise an international environmental regime, the new regime only attained stable levels of compliance when it had been adopted unanimously.[24]

Of course, it is not certain that novel, valuable, or useful ideas will emerge in any particular simulation, but the likelihood can be increased by the appropriate selection of participants and careful management of their decision context to maintain required pressure, relevance, and distance.

Even if novel insights appear in a simulation, it is not certain how relevant or applicable they will be to the real problem being investigated. Nor is this necessarily easy to determine. The simulation itself does not offer such a test, for the same reasons that simulations are not viable for hypothesis testing, but the more closely the simulation design has succeeded in capturing essential behavioral features of the real system, the more likely the simulation insights are to be relevant.

The participants' collective judgment provides the first test for the relevance, generalizability, and practicality of the supposed insights generated in the heat of the simulation. Consequently, this testing is much more reliable, and the practical value of simulations greatly enhanced, when participants are real experts. Granting participants the standing to critique the relevance of simulation insights, and consequently the simulation design that conditioned them, implies an equality of standing between simulation participants and designers that is not present in the experimental or the instructional model. Participants are not just experimental subjects or recipients of instruction; nor is the simulation reduced to a mere pretext for bringing them together.

Taking effective advantage of participants' expertise in this way imposes two requirements on simulation design. First, the rules and structure of the simulation should not be rigidly fixed, but should be open to challenge by participants and renegotiation among participants and designers. Second, a sober, critical postsimulation debriefing is an essential component of the learning effect of the simulation. It is here that the significance and legitimacy of the problem posed is explored, potential implications and consequences of decisions taken and plausible alternatives not taken are explored, and the practical applicability and generalizability of strategies and insights from the simulation is tested against participants' knowledge and experience. If these conditions are adequately met, the model of simulation games as generators of new ideas and insights is persuasive and promising.

Simulations for the Integration of Knowledge

To advance understanding and policy on global environmental change issues, the most important knowledge need is increasingly asserted to be "integrated assessment," the synthesis and organization of knowledge across domains to serve the needs of practical understanding, policy making, and decisions of a variety of actors.[25] The required integration or synthesis can be across a variety of dimensions: across research fields or disciplines, and also across degrees of formalism and confidence, and across people holding knowledge. Three purposes are normally identified for integrated assessment: to bound the importance of particular problems to determine which issues merit attention; to assess the potential consequences of specific policy or decision opportunities; and to identify and prioritize knowledge needs to help inform decisions. Meeting these needs for global environmental change issues and other complex policy problems can require integration or synthesis because disciplinary knowledge is not normally motivated or prioritized by its utility for decision needs, does not normally attend to the linkages between disciplinary domains that can be essential for drawing practical conclusions about the consequences of decisions, and does not normally provide any basis for accommodating informal or intuitive expert judgment.

To inform complex policy problems such as global environmental change, the breadth of integration potentially required is vast. Even if the issues are conceived narrowly, as assessing the environmental consequences of specified, fixed human activities, the breadth of knowledge about the physical and biological world that must be synthesized can be vast. For example, understanding the consequences of specified patterns of acid emissions requires integrating knowledge of atmospheric transport, chemistry, and deposition with knowledge about responses of lakes, forests, and soils under multiple forms of environmental change and other stresses. But since no single decision maker has the authority to determine the level of human emissions, assessing the consequences of particular feasible decisions or policies also requires knowledge of the behavioral, social, economic, political, and organizational factors that determine whether policies are feasible, how much they cost, and what effects they have on the human behavior that is ultimately of concern, plus understanding of ways of valuing the complex, multi-attribute environmental changes that will result.

Two standard methods attempt such integration of knowledge: formal integrated models, and multidisciplinary advisory panels. While these can each be informative and helpful, both suffer from basic limitations, concerned with limits on the breadth of the relevant knowledge that they are able to synthesize, and limits on their ability to clarify preferences or values at stake in the issue considered, or to accommodate diverse preferences.[26] Alternative methods have been proposed to seek aggregation of expert knowledge and opinion without requiring full consensus, such as the Delphi method and expert elicitation techniques.[27] Some of these methods show substantial promise for assessing collective expert views of well-posed uncertain quantities, but they are of little help in the problems of making integrated knowledge serve decision needs, or of accommodating diverse preferences and values that bear on policy choices.

Simulation-gaming methods can help fill some of the important gaps that are left by present assessment methodologies. In pursuing this goal, simulation methods have three salient advantages. First, their open structure permits them to focus and bring together into a transparent, structured, and common forum knowledge from an extremely wide variety of sources—the participants' knowledge, skills, and intuition; technical knowledge embedded in formal models; and research results and disciplinary knowledge from a variety of fields made available within a simulation in text or data. Because of the breadth of knowledge sources that can be represented, simulations can permit a comprehensive treatment of sources of uncertainty in an issue that arise from all sources: limitations of knowledge about the world, as well as strategic uncertainty about the preferences, choices sets, and resources of other actors.

Second, simulations require simulated decisions, and if well designed are sufficiently intense that participants take their decisions very seriously. The pressure on participants, and the requirement for practical action, can force participants to use the available sources of knowledge more aggressively, integrate them more broadly, and organize them more coherently to support their required decisions than they would do in a lower-pressure setting such as a panel meeting or a workshop. Among other benefits, this orientation toward practical decisions can force a sharper consideration than other assessment methods of what current knowledge is most valuable for informing present decisions, and what new knowledge would most help inform required future decisions.

Third, simulations can help to focus and elaborate the criteria by which the consequences are assessed, and the implications of different preferences and values bearing on the issue. Simulations can promote this clarification both by requiring participants to live through the simulated consequences of their choices, and by promoting the airing and elaboration of disparate views through team discussions and interteam negotiations.

The broad implications for simulation design of taking integrated assessment as a major objective are relatively clear. To realize the benefits of the ability to integrate from a wide set of knowledge domains, such knowledge must be made available within the simulation. Two design tradeoffs are likely to arise. First, where the amount of potentially relevant information is vast, there may be a choice between prescreening what is made available, hence the risk of introducing bias in what participants choose to use, and providing a vast ocean of information that could overwhelm the participants or make the simulation even longer and more unwieldy than such exercises already inevitably are (and hence, make it less likely to attract the required senior and expert participants). Second, where there are important gaps in present policy-relevant knowledge, there is a design tradeoff between responsibly characterizing the present state of knowledge, and keeping the decision environment of the simulation sufficiently simple, consistent, and vivid, which may argue for some fictitious extensions or filing of gaps in current knowledge. This second tension is more salient for simulations set in the future, when more knowledge may plausibly be available. This tension may be at the heart of the often denounced data problems and inaccuracies that plague simulation games.[28]

To realize the benefits of the practical decision-orientation of simulation, the exercise must require specific decisions and provide some plausible method of generating consequences, rather than, say, merely consisting of participant teams developing and elaborating scenarios. Such exercises may also have substantial value, but they do not help elaborate the information needs of policy choices; the two forms might complement each other, perhaps by alternating between the two.[29]

Simulation gaming may promise to be a more effective device for the broad integration of knowledge in support of decision and policy needs, in particular the integration of technical and strategic knowledge, than other currently available methods of "integrated assessment" such as formal integrated models. They are not the only effective integrated assessment methods, but they may have key advantages in certain aspects of integration, in particular in integrating scientific with strategic, behavioral, and judgmental knowledge, and in clarifying and exploring the implications of diverse preferences and values. The price is likely to be a lower-resolution treatment of scientific knowledge; while simulation methods can integrate such knowledge, with limited time and attention these are likely to be treated in less detail. Which forms, dimensions, and degrees of integration are likely to be the highest priorities will depend on the issue. Multiple forms of integrated assessment are likely to bring complementary benefits.

Though several commentators on simulation gaming have identified knowledge integration as potentially one of its strongest contributions, few exercises have yet taken this as a primary purpose.[30] Consequently, this application of simulation-gaming methods, while promising, remains relatively undeveloped.

The Value of Simulation Games

This paper has presented four models of how simulation gaming can inform decision making on difficult, complex policy problems, and has argued that two of these models—simulations as experiments, and simulations to instruct decision makers—are either flawed or irrelevant in application to problems of this kind, while two—simulations to promote creativity and insights, or to integrate knowledge—are persuasive and promising. In sum, the potential contributions to better understanding and decision making that could result from the wider use of simulation methods according to the two persuasive models are great enough to justify their more widespread use.

Certain risks and conditions limit this endorsement; they do not negate the potential contribution of simulation, but do require vigilance. These risks are principally associated with the ambiguous status of predictive claims ascribed to simulations. Predictions, whether derived from simulations or other methods, are difficult, usually wrong, rarely done well, and not much improved by analytic complexity or size of modeling effort.[31] Still, analysts continue to make predictions—some through recognition that intellectually honest development of social theory requires taking the risk of making specific, falsifiable predictions,[32] and some for profit, in a flourishing industry of predictors ranging from the responsible and qualified through the enthusiastic, naive, and disreputable.[33]

The risks of inappropriate or excessive claims of predictive power take two forms: bias, and generalization from small samples. Bias can arise in the simulation's design or structure, in suggestions participants may unintentionally be given about how to act, or in the basic specification of what is relevant and important that is essential to simulation design.[34] It may be accidental, or introduced intentionally.[35] The risk of too-confident generalization arises because the rich narrative experience of a simulation lies somewhere between history and experiment—less rich and more replicable than the former, but much richer and less replicable than the latter. With few simulation runs, each yielding a complex narrative outcome, much of what happens is inevitably nongeneralizable variance.

As are other methods of analysis and assessment, simulation gaming is liable to tempt its users and audiences to ascribe stronger predictive power to the exercise than is appropriate, neglecting the risks of bias and generalizing excessively from tiny samples. Simulation methods may be more liable to fall into this trap than other methods, because the prediction of battle outcomes was their earliest objective and because they generate narrative outcomes of a vividness and specificity that can mislead the enthusiastic or naive, but formal analytic methods are also at risk. Formal models are more opaque, while simulations are more vivid and engaging; which approach is more likely to mislead no doubt will vary from case to case. Indeed, decision makers' unique experiences of real-life events also create a significant risk of excessive generalization from samples of one.[36]

Despite these risks, the potential contributions of simulation games, as of other methods for thinking about important decisions with enduring consequences, does depend on attaining a limited kind of predictive power. This consists of identifying contingencies and consequences that are sufficiently likely that they are relevant to consider in decision making, and of drawing broad comparative judgments regarding relative salience, likelihood, and consequences of particular choices.[37]

Consequently, the merit of simulation depends on sustaining a claim for a limited, qualified predictive power that, with qualifications and subject to corroboration through other means, provides a sufficient basis to justify some action under some conditions. The characteristics of this limited predictive power have resisted precise definition. Consequently, advocates of simulation and of other methods have sought to claim this ambiguous middle ground, asserting that their work makes valuable contributions to decision making and planning that do not depend upon specific (or even probabilistic) predictions. Both modelers and gamers tend to urge their audiences to disregard their exercise's specific results while taking very seriously its deeper and more general insights.[38]

Since these limited claims to quasipredictive power are not amenable to strong verification or refutation, they turn on other means of judging their plausibility and relevance that are necessarily informal and consensual. These other means can include correspondence with real system behavior, and the sober, retrospective reflection of participants and other experienced and knowledgeable people on the relevance, generalizability, and practicability of insights and ideas derived from the simulation.

In fact, inappropriate generalization may be more avoidable in simulation than in other methods or in inference from real events, for two reasons. First, simulations have a capability for replication, albeit a highly limited one. There might be particular value in running heterogeneous sets of simulations, with the same participants having different experiences, playing different roles, and engaging different specific problems, all within the same problem area. Second, simulation's risks are mitigated in the debriefing, when participants can question the simulation's relevance and generalizability, present other possibilities, and reflect critically on their experience, at some distance from the experience they have just lived through.

One must approach complex policy problems with some humility, whatever means is used to study them. While the potential contributions of simulation gaming are large, and complementary to those of other methods, it is undeniable that despite 40 years of experience, simulation-gaming methods have not yet lived up to their perceived promise. The reasons are not clear. Garry Brewer and Martin Shubik, reviewing the experience with both simulation games and formal models in military planning, cite a variety of reasons including limited publication and limited forums to develop and support professional standards.[39] Outside security, the sparser experience with simulations makes these problems more severe. Current technological advances in communications, computing power, interfaces, and modeling tools are substantially advancing what can be achieved in simulation, and making feasible alternative forms of simulation that would no longer require assembling a large number of expert participants for a week. How these new tools could advance the standards of simulation gaming as presently practiced, and the merits and limits posed by such new fora as distributed simulation, are not yet explored.

Notes

1. Alfred H. Hausrath, *Venture Simulation in War, Business, and Politics* (New York: McGraw-Hill, 1971), pp. 1–5.
2. Garry D. Brewer and Martin Shubik, *The War Game: A Critique of Military Problem-Solving* (Cambridge, MA: Harvard University Press, 1979).
3. See, for example, Martin Shubik, *The Uses and Methods of Gaming* (New York: Elsevier, 1975), Table 2.1, p. 28; K. C. Bowen, *Research Games* (London: Taylor and Francis, 1978), pp. 11–17; and William M. Jones, *On the Adapting of Political-Military Games for Various Purposes* (Santa Monica, CA: RAND Corporation, N-2413-AF/A, March 1986).
4. Paul Bracken, "Gaming in Hierarchical Defense Organizations," in *Avoiding the Brink: Theory and Practice in Crisis Management*, eds. Andrew C. Goldberg, Debra Van Opstal, and James H. Barkley (London: Brassey's, 1990), pp. 81–98.
5. See, for example, Margaret A. Thomas, *An Energy Crisis Management Simulation for the State of California* (Santa Monica, CA: RAND Corporation, R-2899-CEC, August 1982).
6. See, for example, Edward A. Parson, *Negotiating Climate Cooperation: Learning from Theory, Simulations, and History*, doctoral dissertation, Harvard University, 1992; Robert J. Lempert and William Schwabe, "Transition to Sustainable Waste Management: A Simulation Gaming Approach" (Santa Monica, CA: RAND Corporation, European-American Center for Policy Analysis, 1993); Jill Jager et al., "The Challenge of Sustainable Development in a Greenhouse World: Some Visions of the Future" (Stockholm: Stockholm Environment Institute, 1991); and Laura Cornwell and Robert Costanza, "An Experimental Analysis of the Effectiveness of an Environmental Bonding System on Player Behavior in a Simulated Firm," *Ecological Economics* 11 (1994): 213–226.
7. Shubik, *The Uses and Methods*, p. 7. Other forms of simulation include physical models such as wind tunnels, formal mathematical representations such as predator-prey models, and computer models.

8. Thomas C. Schelling, "An Uninhibited Sales Pitch for Crisis Games," Internal Research Memorandum, RAND Corporation, circulated Summer 1964 and published in "Crisis Games 27 Years Later: plus c'est deja vu" (Santa Monica, CA: RAND Corporation, P-7719, 1964).
9. Harvey DeWeerd, "A Contextual Approach to Scenario Construction," *Simulation and Games* 5 (1975): 403–414; and Peter de Leon, "Scenario Designs: An Overview," *Simulation and Games* 6 (1975): 39–60.
10. Lincoln Bloomfield, "Reflections on Gaming," *Orbis* 27 (1984): 783–790; William M. Jones, *On Free-Form Gaming* (Santa Monica, CA: RAND Corporation, N-2322-RC, 1985).
11. Nihajlo D. Mesarovic, *Globesight: A System for Integrated Assessment of Climate Change* (Cleveland, OH: Systems Applications, 1994).
12. Howard Raiffa, *The Art and Science of Negotiation* (Cambridge, MA: Harvard University Press, 1982); Garry D. Brewer, "Methods for Synthesis: Policy Exercises," in *Sustainable Development of the Biosphere*, eds. William C. Clark and R. E. Munn (Cambridge: Cambridge University Press, 1986), pp. 455–473; Ferenc L. Toth, "Models and Games for Long-Term Policy Problems," paper presented to the 1994 Meeting of the International Simulation and Gaming Association (ISAGA), Ann Arbor, MI, 1994; Ferenc L. Toth, "Policy Exercises: Objectives and Design Elements," reprinted as International Institute for Applied Systems Analysis (IIASA) RR-89-2, *Simulation and Games* 19 (September 1988): 256–276; Ferenc L. Toth, "Policy Exercises: Procedures and Implementation," reprinted as IIASA, RR-89-2, *Simulation and Games* 19 (September 1988): 256–276; Ferenc L. Toth, "Policy Exercises: The First Ten Years," paper presented to the 1994 Meeting of ISAGA, Ann Arbor, MI, 1994; and C. S. Holling, ed., *Adaptive Environmental Assessment and Management* (Chichester: Wiley, 1978); Nicholas C. Sonntag, "Commentary," in *Sustainable Development of the Biosphere*, eds. Clark and Munn, pp. 472–475.
13. See, for example, Robert A. Levine, *Crisis Games for Adults*, Internal Research Memorandum, RAND Corporation, circulated Summer 1964 and published in "Crisis Games 27 Years Later: plus c'est deja vu"; Randall L. Schultz and Edward M. Sullivan, "Developments in Simulation in Social and Administrative Science," in *Simulation in the Social and Administrative Sciences*, eds. Harold Guetzkow, Philip Kotler, and Randall L. Schultz (Englewood Cliffs, NJ: Prentice-Hall, 1972); and Brewer and Shubik, *The War Game*.
14. Experimental economists, studying decisions whose most important outcomes are monetary, achieve this confidence by paying participants according to their outcomes. See, for example, Alvin E. Roth, "Laboratory

Experimentation in Economics," *The Economic Journal* 98 (1988): 974–1031. For decision problems whose important outcomes are not purely monetary, it is not clear that payments would achieve the intended effect.

15. Guetzkow has pointed out that in simulating foreign policy, the most central problems of realism are time and memory, in particular the absence in the accelerated simulated world of slow processes of learning, adaptation, reconciliation, cooling of tempers, and forgetting that occur over the time-scales of real government decision making. In crisis management, though, the relationship of real and simulated time-scales can be reversed; a major purpose of simulation can be to permit more leisurely or repeated "dress rehearsal" of organizational responses to crises that demand response too fast for reflection. Harold G. Guetzkow, "Six Continuing Queries for Global Modelers: A Self-Critique," in *Simulated International Processes: Theories and Research in Global Modeling* (Beverly Hills, CA: Sage Publications, 1981), pp. 331–358.
16. This aspiration for simulations has been stated most cogently by Harold Guetzkow and his colleagues. See, in particular, Harold Guetzkow, "Simulations in the Consolidation and Utilization of Knowledge about International Relations," in *Theory and Research on the Causes of War*, eds. Dean G. Pruitt and Richard C. Snyder (Englewood Cliffs, NJ: Prentice-Hall, 1969); Richard C. Snyder, "Some Perspectives on the Use of Experimental Techniques in the Study of International Relations," in *Simulation in International Relations: Developments for Research and Teaching*, eds. Harold Guetzkow et al. (Englewood Cliffs, NJ: Prentice-Hall, 1963); and the essays collected in Michael Don Ward, ed., *Theories, Models, and Simulations in International Relations: Essays in Honor of Harold Guetzkow* (Boulder, CO: Westview Press, 1985). In addition, Robert Mandel summarizes the use of political games to test hypotheses about systematic perceptual biases of decision-makers in crises, in "Political Gaming and Foreign Policy Making during Crises," *World Politics* 29 (1977): 610–625.
17. Karl Popper, *The Logic of Scientific Discovery* (New York: Harper and Row, 1968).
18. Levine, *Crisis Games for Adults;* and K. C. Bowen, *Research Games.*
19. Harold Guetzkow and Joseph J. Valadez, "Simulation and 'Reality': Validation Research," in *Simulated International Processes*, eds. Guetzkow and Valadez, pp. 253–330; Mandel, "Political Gaming"; and Schelling, "An Uninhibited Sales Pitch."
20. James G. March, Lee S. Sproull, and Michal Tamuz, "Learning from Samples of One and Fewer," *Organization Science* 2 (1991): 1–13.
21. Guetzkow and Valadez, "Simulation and 'Reality'"; and Mandel, "Political Gaming."
22. For example, Schelling reports a consistent difficulty in keeping a crisis boiling across many simulations with disparate settings, contexts, and participants, and advances a cogent explanation based on how teams manage internal differences over how forcefully they should act. Thomas C. Schelling, "The Role of War Games and Exercises," in *Managing Nuclear Operations*, eds. Ashton B. Carter, John D. Steinbruner, and Charles A. Zraket, pp. 426–444 (Washington, DC: Brookings Institution, 1987).

23. Schelling, "An Uninhibited Sales Pitch."
24. Parson, *Negotiating Climate Cooperation.*
25. Edward A. Parson, "Searching for Integrated Assessment," paper presented to the 3rd meeting of the CIESIN Commission on Global Environmental Change Information Policy, Washington, DC, February 17, 1994; Hadi Dowlatabadi and M. Granger Morgan, "Integrated Assessment of Climate Change," *Science* 259 (1993): 1813; Edward S. Rubin et al., "Integrated Assessment of Acid-Deposition Effects on Lake Acidification," *Journal of Environmental Engineering* 118 (1992): 120–134; and Brewer, "Methods for Synthesis."
26. See, for example, William C. Clark, "Themes for a Research Program," in *Sustainable Development of the Biosphere*, eds. Clark and Munn; and William C. Clark and Giandomenico Majone, "The Critical Use of Scientific Inquiries with Policy Implications," *Science, Technology, and Human Values* 10 (1985): 6–19.
27. Norman C. Dalkey, *Studies in the Quality of Life: Delphi and Decision-Making* (Lexington, MA: Lexington Books, 1972); M. Granger Morgan and Max Henrion, *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis* (New York: Cambridge University Press, 1990); David Keith and M. Granger Morgan, "Elicitation of Expert Opinion Regarding Climate Uncertainty," *Environmental Science and Technology*, forthcoming.
28. See, for example, Brewer and Shubik, *The War Game;* Martin Shubik, *Games for Society, Business, and War* (New York: Elsevier, 1972), esp. pp. 299–308; and Garry D. Brewer and Bruce Blair, "War Games and National Security with a Grain of SALT," *The Bulletin of the Atomic Scientists* (June 1979): 18–26.
29. William M. Jones, *On Free-Form Gaming* (Santa Monica, CA: RAND Corporation, N-2322-RC, 1985), p. 42; Robert Mandel, "Professional-Level War Gaming: A Critical Assessment," in *Theories, Models, and Simulations*, ed. Ward, pp. 483–500.
30. See Brewer, "Methods for Synthesis"; Uno Svedin and Britt Aniansson, eds., *Surprising Futures*, Report 87:1 (Stockholm: Swedish Council for Planning and Coordination of Research, 1987); Jager et al., *The Challenge of Sustainable Development;* and Toth, "Policy Exercises: The First Ten Years."
31. William Ascher, *Forecasting: An Appraisal for Policy-Makers and Planners* (Baltimore, MD: Johns Hopkins University Press, 1978).
32. Ithiel de Sola Pool, "The Art of the Social Science Soothsayer," in *Forecasting*

in International Relations: Theory, Methods, Problems, Prospects, eds. Nazli Choucri and Thomas W. Robinson (San Francisco: W. H. Freeman and Company, 1978).

33. Herbert L. Smith, "The Social Forecasting Industry," in *Forecasting in the Social and Natural Sciences*, eds. Kenneth C. Land and Stephen H. Schneider (Dordrecht: Reidel, 1987).
34. Harey DeWeerd, "A Contextual Approach to Scenario Construction," *Simulation and Games* 5 (1975): 403.
35. Paul Bracken, "Unintended Consequences of Strategic Gaming," *Simulation and Games* 8 (1977): 283–318.
36. Lloyd S. Etheredge, *Can Governments Learn?* (New York: Pergamon, 1985).
37. Garry D. Brewer, "Discovery Is Not Prediction," in *Avoiding the Brink: Theory and Practice in Crisis Management*, eds. Andrew C. Goldberg, Debra Van Opstal, and James H. Barkley (London: Brassey's, 1990), pp. 99–107.
38. Schwarz et al. observe, ironically, that researchers working in predictive methodologies never say "prediction," and only call other peoples' work "forecasts"; their own products are called "projections" or "scenarios." Brita Schwarz, Uno Svedin, and B. Wittrock, *Methods in Futures Studies: Problems and Applications* (Boulder, CO: Westview Press, 1982), p. 109.
39. Brewer and Shubik, *The War Game*, pp. 272–274.