THE EPISTEMOLOGY OF WAR GAMING

Robert C. Rubel

Anyone who has conducted or has studied actual warfare knows well its massive complexities.[1]

These complexities do not relieve humans from the responsibility for making decisions—difficult decisions—aimed at navigating their organizations successfully through campaigns, be they in a theater of war or in the halls of the Pentagon. Minds must be prepared beforehand, both in their general, educated functioning and in the specific, sophisticated understanding of conflict and the competitive environments they face. This preparation must be predicated on the internalization of "valid" knowledge about the conflict environment. There are many ways of gaining such knowledge: the study of history and theory, practical experience, and exposure to the results of various kinds of research and analysis. Each of these methods of developing knowledge has its own particular epistemology—formally, a "theory of the nature and grounds of knowledge, especially with reference to its limits and validity," or more practically, rules by which error is distinguished from truth. War gaming is a distinct and historically significant tool that warriors have used over the centuries to help them understand war in general and the nature of specific upcoming operations. The importance of war gaming demands serious examination of the nature of the knowledge it produces.

Professor Rubel is chairman of the Wargaming Department in the Naval War College's Center for Naval Warfare Studies. Before retiring in the grade of captain, he was a naval aviator, participating in operations connected with the 1973 Yom Kippur War, the 1974 Cyprus crisis, the 1980 Iranian hostage crisis, the TWA flight 847 crisis, and DESERT SHIELD. He commanded Fighter Attack Squadron (VFA) 131 and served as the inspector general of U.S. Southern Command. He attended the Spanish Naval War College and the U.S. Naval War College, where he served on the faculty before his present appointment. He has a BS degree from the University of Illinois, an MS in management from Salve Regina University in Newport, Rhode Island, and an MA in national security and strategic studies from the Naval War College (1986).

Naval War College Review, Spring 2006, Vol. 59, No. 2

Before going farther, it is worthwhile to define exactly what we mean by "war game." Peter Perla provides as good a definition as any: a war game is "a warfare model or simulation whose operation does not involve the activities of actual military forces, and whose sequence of events affects and is, in turn, affected by the decisions made by players representing the opposing sides."[2] War gaming, rightly considered, is inherently a method of research, regardless of how people apply it. The essence of war gaming is the examination of conflict in an artificial environment. Through such examination, gamers gain new knowledge about the phenomena the game represents. The purpose of a game is immaterial to this central epistemological element. Moreover, the gaining of knowledge is inherent and unavoidable, whatever a game's object. The real question is whether such knowledge is valid and useful. This question is all the more important because of the growing reliance on gaming techniques in an increasingly complex world.

This article will attempt to initiate a professional dialogue on the underlying logic structure of gaming by examining the epistemological foundations of gaming in general and ways in which the knowledge gained from specific games can be judged sound.

Perhaps the most compelling reason to conduct such an inquiry is the possibility of insidious error creeping into war games. War gaming, even after centuries of practice, is still more a craft than a discipline, and it is quite possible for rank amateurs, dilettantes, and con artists to produce large, expensive, and apparently successful but worthless or misleading games for unsuspecting sponsors. There is little incentive to apply incisive criticism to games in which heavy investments have been made, and persons or organizations inclined to do so are hampered by lack of an established set of epistemological theory and principle. This does not mean that the majority of games are fatally flawed; it does mean that there is no accepted set of criteria to determine whether they are or not. Judgment as to the success and quality of a war game, especially one of high profile and consequence, is too often the result of organizational politics.

EPISTEMOLOGY

Some elaboration of the meaning of this somewhat esoteric term is essential. To avoid getting sidetracked by philosophical complexities, we can adopt a convention based on current thinking. One widely accepted branch of modern epistemological theory holds that knowledge results from the building of simplified mental models of reality in order to solve problems. The "validity" of a model (or knowledge) emanates from its utility in problem solving.[3] This approach seems sufficient for our purposes. Knowledge is a practical human response to the challenges of our environment. Valid knowledge is that which has

sufficient practical correspondence to our environment to be useful for problem solving.

Readers with knowledge of modeling and simulation will immediately find resonances in this definition with widely used definitions of computer simulation validity—for example, "substantiation that a computerized model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model."[4] Thus we are not so much concerned with the validity of knowledge in an absolute sense as with the practical utility of knowledge emanating from a game relative to the projected warfare environment in which it will be applied. Most war games are oriented in some way to the future, either explicitly or inherently; accordingly, the predictive value of knowledge emanating from a game is critical. At this point many veteran gamers will cry foul, as it is widely accepted that war games are not predictive (although there are some who will disagree). To untangle this knot, let us go back to our baseline definition of valid knowledge—that which is useful for problem solving. This presupposes that the environment can to some degree be shaped by decisions. If it were not, war gaming—in fact, any decision-support tool—would be irrelevant. If the environment is malleable, however, there are "right" and "wrong" decisions available to the decision maker.[5] Ignorant decision makers would be at the mercy of chance; their decisions would be shots in the dark, or worse. An informed decision maker—one who possesses valid knowledge about the environment and the potential consequences of alternate choices—could do better than that in a future situation. Valid knowledge is predictive to that extent. However, since life in general and war in particular are influenced by thousands of little happenstances that are beyond the control of any single decision maker (a true definition of Clausewitz's "friction"), "right" decisions do not guarantee success. If they did, war would be formulaic and gaming unnecessary. For that reason, although valid knowledge of the environment is inherently predictive—in that it indicates potentially valid cause-and-effect relationships through which decision makers can bring about their intent—a war game can never be truly predictive.

Setting aside, for now, arguments about certain war games in history that have seemed in some way predictive, we are left with the uncomfortable question of what games are good for if they cannot truly predict. Indeed, why do we game at all?

WHY GAME?

If we accept the notion that war gaming is inherently a research tool (a definition that includes the produced effects of education, training, experimentation, and analysis) and one that generates potentially valid knowledge, we must ask under

what conditions, or for what problems, it can have validity. Can it be used validly in lieu of other tools, or does it occupy a unique relationship to a class of problems for which it is the only valid tool?

Perhaps the deepest treatment of this question is that of John Hanley, who relates the inherent nature and structure of war gaming to the amount and kind of "fuzziness" (indeterminacy) attending a problem. Indeterminacy comprises those things we do not know about either the initial conditions of relevant elements of the problem or about the effects of our potential attempts to solve it. Hanley posits a spectrum of indeterminacy, as follows:

- *No indeterminacy.* The elements of the problem are known and amenable to engineering solutions.
- *Statistical indeterminacy.* The initial set of conditions is a random variable whose statistics we know, and the effects of our actions upon it can be determined. For instance, the chances of a submarine being in a particular area of ocean could be calculated from intelligence, and our search efforts would be shaped thereby.
- *Stochastic indeterminacy.* The initial set of conditions may be known, but the process by which new states of affairs (for instance, battle outcomes) are produced by our actions is subject to statistical variation—the "roll of the dice."
- *Strategic indeterminacy.* The initial set of conditions is known, but there are two or more competing "players" whose independent choices govern the end state.
- *Structural indeterminacy.* Significant elements of the problem are so little known or understood that we cannot define the problem in terms of the other forms of indeterminacy. Such elements might be "indeterminacy in current conditions, the kinematics of the process, acts of nature, the available response time, and the perceptions, beliefs and values of the decision makers."[6]

Hanley describes war gaming as a weakly structured tool appropriate to weakly structured problems.[7] Such problems are those so complex or poorly defined as to require a tool that can accommodate their considerable imprecision. Warfare in general and many of the problems subsumed within it are certainly weakly structured—that is, marked by structural indeterminacy. This adds up to the first part of the answer to our question: We war-game because we must. There are certain warfare problems that only gaming will illuminate.

This imprecision, or lack of solid structure, characterizes both the problem and the tool, and therefore governs the nature of the knowledge produced by a

war game. That knowledge is not in the form of a solution to an engineering problem. It is commonly said that war games produce insights, not proofs. This conventional wisdom is correct insofar as it goes, but it is not sufficiently developed to stand as an epistemological principle. Following Hanley's line of thought, we can say that the knowledge emanating from a game is also weakly structured, meaning that such knowledge is conditional and subject to judgment in application. Our confidence in the structural calculations for a bridge can be very high if we combine accepted engineering formulae, accurate measurements, and building materials of the predicted quality. In contrast, however, our confidence in answers produced by population sampling cannot be 100 percent; further, any answers produced by game theory for a particular conflict situation must be understood to be conditional on the scope for free choice enjoyed by the opponent. Answers produced by war games are yet more conditional, due to the wide scope of significant variables attendant to warfare, whether or not incorporated into the game. Perhaps the best way to characterize this conditionality is to say that knowledge produced by war games is *indicative*—that is, at its best it can indicate the possibilities of a projected warfare situation and certain potential cause-and-effect linkages.

Insiders have a term for nongames masquerading as games: BOGSAT ("Bunch of Guys Sitting around a Table").

Indicativeness is no mean thing when dealing with a very complex or weakly structured problem. The primary mechanism through which war games produce such knowledge is visualization. Games allow players and observers to see relationships—geographic, temporal, functional, political, and other—that would otherwise not be possible to discern. Seeing and understanding these relationships prepares the mind for decisions in a complex environment. This holds true whether the purpose of the game is education or research.

While weak problem structure is a compelling reason to war-game, there are other equally compelling reasons, each of which has epistemological implications. A common reason for mounting a war game is socialization, either of concepts or people. Many organizations within the U.S. government sponsor games in order to get a wide and diverse set of stakeholders to "buy into" a set of concepts or doctrine. Military "Title X" games (that is, Title Ten, referring to the federal statute that directs the armed services to raise, maintain, and train forces) frequently have this as at least a tacit purpose. Knowledge emerging from such games is less conditional than in other settings, at least with respect to the consensus they are meant to generate. A recent joint war game revealed that none of the military services had invested sufficiently in the suppression of enemy air defenses to support an aggressive airborne assault early in a particular scenario.

That revelation was more than just indicative—it was usable intelligence. Such knowledge could be used to alter budgets or even service roles and missions.

Some games are used to acquaint organizations with each other. This has been an important aspect of homeland security gaming in the wake of 9/11. For instance, in a recent homeland security game, a state emergency management agency learned that it had formally to request federal assistance in a disaster, not just expect it to show up. That knowledge was not in the least conditional; the game provided to key officers of a state agency concrete knowledge of federal requirements.

SIMULATION

War games are inherently simulations of reality. By this we mean that they are simplified representations of a potential future (or perhaps past) warfare situation. Simulation has epistemological implications all its own. Most fundamentally, simulation is a calculation technique, and as such it is coupled to the phenomena it seeks to represent along Hanley's spectrum of indeterminacy. For instance, physicists use simulation techniques to explore subatomic interactions. They can do this with high confidence because the problem set they are dealing with contains no more than statistical indeterminacy. Naturally, then, simulation of war is less closely coupled to its parent phenomenon because of the high degree of structural indeterminacy involved. In other words, it is far less likely that any warfare simulation would be "valid" due to all the imponderables that are necessarily distilled out.

A war game is an artificial representation—that is, simulation—of war that is used to learn more about a particular situation. A common misconception is that computer simulations are war games. Computer programs are not in themselves war games, although they are frequently referred to as such; war games require human players, who may employ computer programs to assist them. In a broad sense, simulation is the attempt to represent reality to the degree necessary to explore the warfare phenomena in which we are interested. Thus when we talk of simulation in this article, it is in the general sense of war-game design and not the narrower sense of computer software.

Following Hanley, we can attack the issue of warfare simulation by establishing a vertical spectrum of sorts, based on the degree of fidelity a simulation possesses. At the bottom of the spectrum exist such games as Go and chess. These games are abstractions; all that is retained of reality is the essence of conflict. That does not mean that valid knowledge cannot be gained from these games; many wise generals have extolled their virtues in preparing the mind for actual battle. At the top of the spectrum are detailed simulations, attempts to capture as much reality as possible. In between exist what we will call "distillations"—

games in which significant simplifications of reality are made for specific purposes. In a sense, all simulations are distillations, because a perfect representation of reality would *be* reality. To put it more practically, exact simulation of real warfare is not possible. Admiral Arleigh Burke illustrated the matter well when he said, "Nobody can actually duplicate the strain that a commander is under in making a decision during combat."

This distilling process has epistemological implications for simulation. Pursuing farther the logic we have been following, we could easily conclude that the knowledge produced by highly distilled games is more conditional and less predictive than that from simulations having greater fidelity. Such reasoning would force us to conduct nothing but elaborate and expensive games. Fortunately, such an epistemological blind alley can be avoided by linking purpose to predictiveness. All war games have explicit purposes, and rarely are these purposes so holistic as to demand unsparing investment in fidelity. Bringing the purpose of a game into focus leads quite naturally to distillation; many games are able to set aside significant aspects of reality. To the extent that distillation promotes clarity, highlighting relationships in the aspect of warfare we are studying, the epistemological damage of failure to include all possible factors is counterbalanced. Since knowledge gained from a war game is in the eye of the beholder (player or analyst), obfuscation caused by excessive comprehensiveness is at least as damaging as the omission of some significant element.

Epistemologically speaking, we conclude that a war game should be designed with as much fidelity as possible without including factors that, because they are not clearly related to its purpose, risk diluting or masking valid knowledge that might legitimately be gained.

There is another implication of simulation that must be addressed: the common wisdom holding that war games are not experiments, as they cannot prove anything. This is clearly true, in terms of John Hanley's logic, since knowledge emerging from games is conditional. The proposition is confirmed also by the nature of warfare simulation; the lack of close coupling with its parent phenomenon due to structural indeterminacy makes it always incomplete and defective in some, possibly unknown, way.

Nevertheless, there is an aspect of war gaming that can accommodate experimentation. Some war games focus on command and control. In them, players are organized into cells, each of which represents a command or perhaps an element of a staff organization. These cells are provided with communications devices (most recently networked computers) and command and control (C2) doctrine. The war game provides a venue in which command and control processes can take place. The point here is that within the context of the game, actual—not simulated—command and control occurs. Thus, knowledge gained from this

activity can be treated like experimental data, subject to all the epistemological principles and injunctions of the scientific method. One caveat is that war games are most commonly one-time affairs, so the data cannot be treated with the same confidence as that gained from experiments run a number of times. On the other hand, simple and appropriately distilled games have been used as substrates within multiple-run C2 experiments, the output of which constitutes valid statistical data.[8] However, in games featuring a significant command and control focus, information gained from the underlying simulation must be treated differently than that derived from the command and control "layer."

GAME ARTIFACTS

Games can easily produce information that is invalid. Commonly, such information is produced by what are termed "game artifacts," defects of simulation that corrupt a game's cause-and-effect relationships. If, for instance, a Control umpire somehow used the wrong weapons-effects table to look up the outcome of a tactical engagement, subsequent player decisions based on that assessment would be tainted. Similarly, defects in display may cause players to be artificially misled as to where units are. Simply ascribing such defects to the "fog of war" and allowing them to be folded into the game's flow is as much an epistemological mistake as assigning too much significance to game outcomes.

It is entirely reasonable to build the fog of war into a game, which can be done in various ways. These devices, such as revealing to players only that information which their reconnaissance assets could "see," normally place bounds on the nature of misinformation that may crop up. Players may, for instance, make unwarranted assumptions about the location of enemy forces due to a lack of information; they might equally do so in the real world, and such imperfection of information does no violence to the intellectual validity of cause and effect or critical analysis. However, if a computer-generated operational picture through some system defect placed a "Red" unit far out of position and thereby affected "Blue's" decision making, we cannot explain it away as the result of a Red computer attack or some sophisticated deception. Nor can it be chalked up to equipment failure that might happen in real life; unless it is known that the game's designers provided for this real-world factor, it cannot be assumed to be a part of the simulation.

A game artifact that is perhaps easier to understand but more difficult to detect or avert is invalid decision making by players. It is a fundamental, if tacit, assumption of war gaming that players will make the best decisions they can. They need not be the right decisions—after all, somebody has to lose—but they must not be capricious or negligent. Players are expected to try to win, or at least to carry out doctrine in a faithful way. When they do not, as a result of alienation,

inattention, or malice, the game's results are contaminated. This can happen all too easily. In some games, Red is constrained by Control, in order to shape the game in some needed way, from certain otherwise reasonable actions it wants to take; if Red players react with disillusionment or cynicism, they may "mentally disengage" from the game and make very different decisions than if they were properly immersed and motivated. Another source of defective decision making is ignorance or improper training among players. If the goal of the game is to examine the efficacy of a particular concept or doctrine but the key players do not know or understand the material, the game results cannot be accepted.

Another player artifact, one that is harder to account for, crops up in games as well: players tend to be more aggressive than they would be in the real world with real lives at stake. There are several inherent reasons for this. First, it is just a game, and therefore real lives are *not* at stake. Second, depending on the extent of the simulation, there are no tactical commanders screaming bloody murder if the operational-level player puts them in a unnecessarily dangerous situation. One of the most common misfortunes to attend Blue players in Cold War games was the loss of amphibious groups because the Blue players had let them sit in exposed positions. Third, since every game has a defined end point or specific set of victory conditions, there is no "tomorrow" to be provided for by players after the last move. Game designers must therefore understand these tendencies and attempt to structure their games to minimize the likelihood and intensity of this player artifact.

THE WAR GAME AS MILITARY HISTORY

We have seen that knowledge gained from war games is conditional—that its validity is ultimately dependent on its effects on decisions made in real-world operations. But analysts examine games after the fact, and all participants have the opportunity to learn from their findings. How should this information be handled, sorted, and considered? How can it be converted into valid knowledge? Because it is not scientific data, it cannot be statistically reduced or otherwise treated in ways appropriate for "hard" data. Perhaps information produced by war games is best considered artificial military history. Game data can then be approached with the full array of methods available to the historian. Moreover, the trap of treating mere discussions as games can be avoided. Insiders have a term for nongames masquerading as games: BOGSAT ("Bunch of Guys Sitting around a Table"). If the data derived from an event consists solely of what participants said, it was not truly a war game, and its results should not be accorded the stature that knowledge gained from a real game should have.

Perhaps the best commentary on converting military history into useful knowledge is to be found in the writings of Carl von Clausewitz. Clausewitz

regarded history as a real-life laboratory of war, one that can be mined for information useful for preparing the minds of future commanders. His approach was what he called *Kritik,* or critical analysis: researching the facts, tracing effects back to their causes, and evaluating the means employed.[9] This process (which emerges from a close reading of Book Two, chapter 5, of his classic treatise *On War*) is as valid today as it was in Clausewitz's time. These three steps constitute more than a method; they establish a criterion for the extraction of valid knowledge from a war game. It is not enough simply to list the facts of what happened in the game; these are meaningless in themselves, because the game was a simulation. We must examine why these events occurred—the combinations of player decisions and umpire determinations that produced them.

Clausewitz himself, however, acknowledges the limits of the method: at some point, results must be allowed to speak for themselves. The critic, "having analyzed everything within the range of human calculation and belief, will let the outcome speak for that part whose deep, mysterious operation is never visible."[10] In other words, war cannot be completely understood in its full complexity; ultimately criticism must recognize that there are factors at work whose functioning can be revealed only by the actual victories or defeats of a commander being studied. This is perfectly reasonable with respect to real warfare. It might also be true for war games, but its usefulness is limited by the fact that they are simulations. For example, a common method of introducing uncertainty into battle-outcome calculations is rolling dice to represent the probabilistic nature of certain phenomena, like sonar or radar detection. Beyond this narrow use of stochastic indeterminacy, game designers frequently aggregate complex interactions of large combat forces with a combination of dice rolls and structured combat-results tables. Here the die simulates the effects of a wide range of variables that are not explicitly modeled.

A common reason for mounting a war game is socialization, either of concepts or people.

It would be easy enough, lacking any other good explanation of the cause-and-effect relationships between player decisions and outcomes, to sense here the presence of invisible factors. But if such "deep, mysterious" elements exist in war games, they are not those of which Clausewitz speaks. A roll of the dice is simply that. To say it simulates unmodeled portions of reality is going too far. The most one can say is that there are physical forces at play on the die itself that players cannot calculate and therefore cannot predict. This is different from admitting one does not understand all the complexities of a real battlefield. Thus, we cannot approach the results of a war game as a military critic would the outcome of a real battle or campaign. Results of a war game cannot be used to fill in

analytical blanks in the way Clausewitz describes, nor can theory or judgment be derived from them in the way historians do from real events.

Nevertheless, we can ascribe a certain significance to war-game outcomes. If the game is run according to a specific set of rules and those rules constitute a valid distilled simulation of reality, outcomes of individual "moves" or entire games can yield useful knowledge. To understand when this can be the case, we need to understand the difference between *rigidly assessed* and *freely assessed* war games. We describe as "rigidly assessed" those games that proceed strictly according to rules governing movement, detection, and combat. Such games produce situations governed by player decisions, the rules, and combat-results tables (manual or computerized). Assuming the absence of artifacts and within the limitations of dice rolls, we can in such a case ascribe significance to game, or even move, outcomes. The game goes where the rules take it; if the rules and the combat-resolution tables are good representations of reality, the outcome constitutes artificial military history, and one can usefully work backward from outcomes and look for reasons. This would be so whether the game is played by hand around a board or at computer workstations. Inputs are generated, and these, by means of a known system, produce results that cannot be predicted or influenced. The game goes where it goes.

Freely assessed games are somewhat different epistemological animals. In these, the flow of the game is governed by umpires and game directors. Instead of following game rules, players make plans and decisions as they would in real life, more or less, and umpires, collecting the interacting moves of all the players, translate them into force movements, detections, and combat results. The umpires may be aided by computers. The key difference is that the game's progress, including move results, are governed by the objectives of the game's sponsors, the time available, and sometimes the conflicting interests of stakeholders. Control may determine that a certain set of conditions must occur at a specific point if the game's objectives are to be met. This is most commonly the case in educational games, but it can also occur in research games. In such a case, Control defines in operational game terms the needed conditions, looks at the situation at the end of the previous move, and then figures out what—within the bounds of plausibility, given the players' new moves—*must have happened* in order to get from that situation to the desired condition.

That is, the umpires deduce tactical outcomes, the necessary inputs, by working backward from a set of desired results. This fact does not negate the validity or value of the game, but it does mean that its *outcome* does not have the same analytical weight as that of a rigidly assessed game. Freely assessed games can be valuable for discovery purposes—perceiving relationships or finding defects in

plans—but they cannot be used to see "who would win." Similarly, they cannot be regarded as artificial military history to the same extent as rigidly assessed games.

MONTE CARLO VERSUS DETERMINISTIC COMBAT RESULTS

A Naval War College elective course on war-gaming theory and practice recently designed and played an instructional board game. In the course of it, a Blue player exclaimed in frustration, "This is a dice game, not a capabilities game!" His observation was trenchant as well as accurate. In the game—which combined various types of dice and combat-results tables—a small Red force had just hammered a larger Blue fleet after four or five very lucky die rolls. The rules had attempted to reflect lower Red strength by awarding hits only on rolls of one or two on a ten-sided die, but five consecutive rolls of one or two now produced a David-slaying-Goliath result. How does one deal with such an outcome?

As we have seen, there are several reasons to roll dice—that is, to use Monte Carlo methods to produce uncertainty in outcomes. Perhaps the best reason is to simulate real-world phenomena that are in fact probabilistic. Some good examples are certain types of radar detections and the reliability of weapons systems. Epistemologically, there are few reasons to object to such an application of probabilistic simulation.

Another reason to roll the dice is to represent the aggregate performance of complicated systems that are at least partially dependent on human performance. If, for instance, we assign an 80 percent probability of a hit by an antiship missile and its purely mechanical reliability is on the order of 99 percent, the other 19 percent of uncertainty would consist of such things as operator error and, perhaps, brilliant maneuvering by the target ship. Here, epistemologically speaking, we start to get a bit uneasy, because the moment probability enters into the picture, we introduce the possibility of very-low-probability occurrences, such as the string of lucky rolls by Red just mentioned. Could such a thing happen? Of course it could—anything is possible—but we must ask ourselves if such an ascription of exceptional human incompetence or brilliance has any place in the intellectual architecture of game objectives. On some level, we may accept the validity of the knowledge produced by such simulation methodology, but the student's complaint haunts us: Is it a dice game or a capabilities game? To put it differently, does the introduction of Monte Carlo methodology distort the intellectual structure of the game?

We have previously asserted that it is not valid to substitute dice rolls for unmodeled aspects of reality. Here we see one reason why—that luck in dice rolling is a special phenomenon in itself. The actual likelihood of unmodeled factors all lining up in a way that would be represented by rolling five ones or twos in a row is likely to be far smaller than the roughly three-in-ten-thousand

odds of such a string of rolls. It would be different if we contemplated a hundred or even a thousand iterations of the game; by looking at the most frequent outcomes, we might then place the "outliers" in their proper perspective. This is done in campaign analyses via computer simulations; scenarios are iterated very many times at high speed to produce a population of results that are subject to statistical reduction. However, most war games are conducted once, and thus the impact of outlying results arising from the peculiarity of Monte Carlo methods must be considered. What validity should we ascribe to a web of human decisions impacted by quirky dice rolls? From this point of view, it appears that invalid Monte Carlo methods can produce game artifacts.

The obvious alternative to Monte Carlo simulation is deterministic calculation, using algorithms. Playing pieces are assigned numbers to represent their capabilities on offense, defense, and perhaps other aspects of combat power. Combat-result tables based on some predetermined formula are consulted to determine outcomes. One simply compares offensive points to defensive points to find a ratio and enters the table with that ratio to look up the result. Every time that ratio arises, the same result ensues. For this methodology, game validity is a function of the accuracy with which the embedded algorithms describe real combat interactions. In a deterministic game, neither human idiocy nor brilliance exists, below the level of the game player; the impact of player decisions is sharply highlighted. This leads us back to the axiom that games should model reality with as much fidelity as possible without masking the phenomena we are trying to elucidate.

STRATEGY AND EFFECTS

Clausewitz extended his *Kritik* from the tactical and operational levels into the realm of strategy through the device of concentric analytic rings. He undertook to analyze and critique the decision of Napoleon Bonaparte (then a general in the field, under the French Directory) to make the peace of Campo Formio by examining the wider strategic context in stages, working from narrower to wider views. In other words, he examined the context for Napoleon's northern Italy campaign to ascertain whether the latter's decision to make peace with the Austrians when and where he did was justified.[11] Such analysis might be possible in war games, but the analyst must decide whether the strategic context of the game was established with sufficient detail and realism to stand as a criterion for judgment. Operational-level war games are frequently accompanied by unrealistic or truncated strategic contexts, in order to allow the fighting called for by game objectives to take place. Assessments of operational decision quality or utility based on such strategic criteria are likely to be invalid.

As an example, the Naval War College's Global War Game series (played annually from 1979 until 2001) focused on rapid, operational-level decision making, supported in later years by an advanced, networked collaboration environment and computer-analysis tools.[12] In 2000 the scenario featured a brink-of-war situation in which Blue players had to generate high "speed of command" in the conflict's first exchanges in order to avoid catastrophic casualties. The national-level command apparatus was played by Control, which assigned the role to a small cell of subject-matter experts. Pressure from the game's directorship resulted in quick, streamlined, and aggressive decision making by this cell (also recall the player aggressiveness artifact mentioned previously), allowing operational-level players to preempt and gain a smashing victory. The postgame judgment was that network-enabled speed of command was a very good thing.[13] However, in fact, the strategic-level command apparatus context had been so unrealistic as to invalidate any such assessment. In any case, games that incorporate detailed play at both the strategic and operational levels are uncommon, for a number of reasons, including the practical matter that free play at the strategic level tends to constrain or disrupt operational-level processes.

Strategic games have a long history, and they can produce knowledge as valid as that from games at the operational and tactical levels. It is possible to explore the strategic conflict environment in order to discern relationships between factors, including the structure of incentives that influence players. Sometimes these games are used as background for subsequent operational-level games. If so, consistency must be achieved between the scenarios, orders of battle, and player assumptions of the various games, or it will not be possible to relate their outcomes to each other—they will be "apples and oranges." Moreover, analysts must rigorously identify artifacts in the first game in order to prevent them from affecting player decisions or analysis in following games.

There is yet another issue related to strategic context and critical analysis that must be considered—"effects-based operations," or EBO. This concept, which is permeating the U.S. military lexicon today, has been an aspect of war gaming for the last few years. EBO focuses on the second- and higher-order effects of military actions, with an eye toward making these actions more effective and avoiding adverse side effects, in terms of broader purposes. At the tactical and operational levels, the prediction of battle effects is reasonably straightforward, at least in the physical realm. Consequently, assessing war-game move outcomes when players are using EBO planning methods is fairly straightforward. Even "moral" effects at these levels are possible to assess; for instance, units that are outflanked tend to lose cohesion, and generals faced with the cutting of main supply routes can be expected to withdraw their forces to avoid encirclement.[14]

However, at the strategic level, the degrees of freedom proliferate, and assessment of possible effects on populations and on national leaders is highly problematic.[15] If it is difficult in real war, as has been proven time and again, it is doubly hard in war games, which look to an uncertain future.

There is an epistemological solution. It lies in understanding that while war games are not crystal balls, they can highlight the relationships between factors. We could, for example, decide to explore the political terrain of war termination under given mind-sets or policies of the enemy leadership. Game designers would "script" a set of presumed conditions faced by enemy leadership—personal proclivities, influence distribution among top leadership, and the like—establishing a "moral context" for strategic decision making. Players would role-play and umpires assess strategic effects strictly within this context. Such a game would have a chance at generating indicative information concerning, say, the relationship between the course of one's own offensive operations and the willingness of an enemy leadership to negotiate. Iterative gaming involving different internal enemy conditions would at the very least prove educational.

COMPARING WAR GAMES

A large military organization with a mission of experimentation and concept development once developed a system for synthesizing the data gained from multiple war games so that it could capitalize upon the considerable investment in gaming by the services. The key to the system was correlation; the more frequently a particular result emerged, the more weight was ascribed to it. Epistemologically, there is potential validity to this approach, but it was implemented in a way that had serious defects. First, the system essentially captured and digested the comments of senior and experienced subject-matter experts who participated in the games and interpreted their results. However, that in effect reduced games to BOGSATs; the system processed people's opinions, not game results (i.e., plans, decisions, and move assessments). Second, since the same senior folks tend to be invited to games, one after another, an expert with a particular outlook or agenda is likely to make very similar comments at each game, thus lending these "findings" artificial weight. It is easy enough to pick apart such a correlation system, but less easy to establish a sound way of comparing results of different war games.

Experienced gamers, for instance, quite naturally on the basis of running many games, derive rules of thumb and gaming techniques; also, a number of phenomena tend to occur in similar and consistent ways even in games of very different kinds. One example is the tendency of players to "fight the scenario"—that is, to object to certain aspects of the game's story line, structure, or orders of battle and use these objections to hedge against the possibility of "losing." Such

underlying commonalities with respect to game process can lead gamers to assume that equivalent commonalities exist in terms of game substance. They believe that they can derive on that basis, in an essentially correlative way, synthesized lessons from the substantive outputs of multiple games. But such an attempt is intellectually unsupportable, on several grounds.

First, unless games are specifically designed to be analyzed in conjunction with other games, there are almost certain to be differences in objectives and design so fundamental as to prevent it. For instance, imagine two games producing results that, taken together, point to an apparent vulnerability of the littoral combat ship (LCS)—in both games several of that ship type are sunk. Closer scrutiny reveals, however, that whereas in one game the objective was indeed to examine the utility of the LCS in littoral warfare, with consequent close attention in move assessment to ship defenses, the other was meant to explore maritime command and control processes, with assessments focusing on the handling of various kinds of reports and orders by the C2 system. In the latter game, umpires in fact imposed ship losses specifically in order to generate reports and command responses. To attach significance to the fact that several LCSs were lost in both games distorts conclusions, since in the second game at least some of the losses were "artificial." This example is a bit contrived, in order to define the issue clearly; in reality, many games appear to offer numerous opportunities for comparison, because their methods and outputs appear comparable. Even then, however, there can exist subtle, disabling differences.

War gaming is still more a craft than a discipline, and it is quite possible for rank amateurs, dilettantes, and con artists to produce apparently successful but worthless or misleading games.

A second reason why correlation of seemingly similar events in different games fails at the substantive level (even inside the scenario) arises from the very nature of gaming. Games are not reality, and players are likely to do things they simply would not do in reality. A common manifestation, as previously discussed, is inadvertently leaving important forces unprotected, to be knocked off by the enemy. Controllers and umpires, however, rarely identify such instances, making it almost impossible to go back after the game and determine when this tendency was in play.

What then can be gleaned from comparing multiple games? First, we must remember what games can reliably produce: knowledge about the nature of a warfare problem, such as potential flaws in a plan, the potential importance of geographic features, gaps in command and control, logistical needs, etc. The familiar metaphor of blind men feeling around an elephant tells us that multiple games, almost regardless of their individual methodologies, can contribute

incrementally to the understanding of a particular warfare problem. That problem may be a specific scenario, such as a war on the Korean Peninsula, or it may be a function, like close air support. If we avoid attaching significance to the number of times something happens, we can derive epistemologically sound knowledge. We can collect anecdotes of various game happenings, lessons learned, and analyses, to be pieced together into a more complete, qualitative understanding of the issue in which we are interested. In one game we may learn that command and control arrangements for close air support are flawed, in another that certain types of preferred weapons are in short supply. These specific outcomes can be combined to form a picture of the "elephant."

LISTENING TO WHISPERS

Our general thrust to this point has been to identify limitations on what can be said to have been learned from a war game. Still, there is an epistemological reason to wrest from a game all the valid knowledge it has to offer. If it is easy to overstate what was learned from a game, it is also easy to ignore what it did produce—all *too* easy, if that information or knowledge is either subtle or somehow threatening. Such information, being tempting to dismiss, might be called "whispers."

We have seen that the results of a war game are in the eyes of the beholder (player or analyst), because of conditionality. That is, game-generated knowledge, being merely indicative in itself, must be combined with judgment in order to have useful predictive value. But such application of judgment is rarely easy or straightforward. For example, in war games at the Naval War College in the 1920s and '30s, despite the repeated indications of the importance of the Mariana, Caroline, and Marshall island groups—then known as the Mandated Islands—as intermediate logistics bases in any campaign to relieve the Philippines and defeat Japan, it took many years for the U.S. Navy to abandon fully the idea of mounting a direct thrust on the Philippines from Pearl Harbor.[16] The games, apparently, were telling officers things many did not want to hear. Conditional knowledge can be a slippery thing. Games are complex affairs that almost always produce more information than their designers intended to generate. Moreover, game results are often equivocal, open to interpretation.

The subjective nature of game-produced knowledge is nowhere clearer than in games that generate information that is bureaucratically or politically threatening to players or sponsors. It is all too easy either to ignore or put a favorable spin on game events or results that do not fit comfortably into existing doctrines or accepted theories. A notable historical example of this phenomenon was a war game conducted by the Japanese Combined Fleet staff prior to the Midway operation. Historians have made much of the fact that the umpires resurrected a Japanese carrier that had been sunk by American aircraft operating out of

Midway, citing it as evidence of "victory disease." In fact, however, the Japanese umpires were perfectly justified—a dice roll had given a highly improbable hit to level-flying bombers (that is, as opposed to dive-bombers), which had proven generally ineffective in attacking ships. They were properly attempting to prevent a capabilities game from becoming a dice game. However, at another point during the game it was asked what would happen if an American carrier task force ambushed Vice Admiral Chuichi Nagumo's carrier force while it was raiding Midway, and that uncomfortable question seems to have been ignored. The existing plan was based on deception and surprise, tenets and war-fighting values dear to the Imperial Japanese Navy. To acknowledge the existence of an American task force northeast of Midway in a position to ambush Nagumo's carriers would have been to discount the possibility of surprise. The Japanese planners simply did not want to admit that—it would have negated their plans, and there was no time to start again from scratch. At the very least the game should have suggested more extensive searches in that sector, but the plan was not modified even to that extent. It was easier to ignore this particular game outcome.[17]

The "whispers" phenomenon has important implications for war-gaming policy. As the Japanese example shows, players and sponsors are almost never objective about their games. Games are played in a setting of institutional imperatives, such as budget justification, or the need to affirm a service's foundational theory and doctrine ("airpower is decisive," "the infantryman is the ultimate strategic weapon," and so on). Moreover, as in the Japanese case, games may be linked in some way to imminent deadlines. All of these factors tend to deaden ears to the whispers. But these whispers are frequently the most important outcomes of war gaming. How can an organization increase its ability to hear them?

The key is objective, disinterested sponsorship, or at least analysis. A sponsoring organization (the agency that "gives," or initiates, the game, as distinct from the facility that stages it) cannot realistically be relied upon, especially if constrained by time, political imperatives, or the dictates of theory and doctrine, to hear whispers from its own games. A frequent alternative is the use of civilian contractors; the difficulty is that contractors, paid for their services and generally hoping for follow-on contracts, have a built-in incentive, regardless of the talent or intellectual integrity of the individuals and companies involved, to tell sponsors what they want to hear, or at least not press them to hear whispers. Another option is academia. The service colleges frequently perform this role, and each has a war-gaming center. These facilities, however, must have a sufficient degree of autonomy—specifically, protection from firing of personnel or other sanctions for games that produce uncomfortable results. The gaming departments themselves must incorporate a culture of rigorous intellectual objectivity and commitment to the discipline of war gaming.

Finally, the results of war games must receive proper handling. Perhaps most importantly, the heads of sponsoring organizations must commit themselves to receiving game results directly and personally from gaming organizations, and not after filtering and sanitizing by their own staffs.

A GUILD OF WAR GAMERS

In professional war gaming the stakes are high. Not only do games cost money and time, but their results can influence important operational and programmatic decisions. This holds true for the business as well as military worlds. Many organizations conduct war games, and even more consume their results, but few if any individuals involved have rigorous understanding of whether the games produce valid knowledge. As we have seen, it is entirely possible for games to produce valid-looking garbage. It is not easy to distinguish error from insight; it can be accomplished only if game design, execution, and analysis are conducted with discipline and rigor, and according to principles like those outlined here. Even then, however, wheat cannot be sifted from chaff with consistency and confidence unless another step is taken.

War gaming is currently a craft. There are a few highly experienced and skilled game designers and directors "out there," and these individuals each operate by rules of thumb they have learned over the years. Approaches vary. A large war game might be proclaimed a success by sponsors but at the same time be criticized severely—in private—by players, observers, and analysts. Who is right? What is missing is a universal set of standards, an accepted body of knowledge, such as established academic disciplines possess. In the "hard" sciences, even the social sciences, there is less room for charlatanism and sloppiness. Practitioners there have frameworks for understanding their disciplines and becoming credentialed in them. War gaming needs the same if it is to warrant the resources invested in holding games and the confidence routinely vested in their results. Such a step is all the more important today in light of the changing nature of warfare and the concomitantly receding utility of traditional force-on-force gaming techniques. "Fourth-generation warfare" blends politics, mass media, global information flows, culture, and religion, with combat in a highly complex way; games attempting to simulate it can lead to catastrophic intellectual error if not conducted under the aegis of a sound, overarching framework.

The substrate for founding a gaming discipline exists. The nation's war and staff colleges all have war-gaming departments whose directors have professional contact with each other and with key figures in the wider war-gaming world. Certain academic institutions, notably the Naval Postgraduate School, teach courses in war gaming. These organizations could come together in a

"guild" of sorts to establish standards and promote the formalization and professionalization of a war-gaming discipline. This professional society, in effect, could draw members from outside the military, such as business and academia, whose contributions would universalize standards and add vitality. The society might publish a professional journal, with refereed articles. All this is necessary if war-game output is to merit a level of epistemological confidence commensurate with the uses made of it.

Valid knowledge can emerge from war games, but only if due diligence is applied. That diligence is considerably hampered today because war gaming is a craft or an art, not a true profession, a discipline. Much more work must be done. Those who believe in the value of games must now link up and work toward the goal of truly professional war gaming.

NOTES

1. For background on the theory and practice of war gaming, see Robert C. Rubel, "War-Gaming Network-centric Warfare," *Naval War College Review* 54, no. 2 (Spring 2001), pp. 61–74.
2. Peter P. Perla, *The Art of Wargaming: A Guide for Professionals and Hobbyists* (Annapolis, Md.: Naval Institute Press, 1990), p. 164.
3. F. Heylighen, C. Joslyn, and V. Turchin, eds., *Principia Cybernetica Web* (Brussels: Principia Cybernetica, 1995), available at pespmc1.vub.ac.be/EPISTEMI.html.
4. S.[tewart] Schlesinger et al., "Terminology for Model Credibility," *Simulation* 32, no. 3 (1979), pp. 103–104.
5. *Right* and *wrong* are not absolute terms. For the purpose of this discussion, "right" means a decision the likely outcome of which has envisioned benefits for the decision maker. Clearly, even "right" decisions could result in failure due to bad luck (statistically speaking) or the intervention of imponderable factors.
6. John T. Hanley, *On Wargaming* (dissertation, University of Michigan, Ann Arbor, Mich.; University Microfilms International, 1991), p. 13.
7. Ibid., pp. 19–25.
8. Peter Perla, Michael Markowitz, and Christopher Weuve, *Game-Based Experimentation for Research in Command and Control and Shared Situational Awareness,* CRM D0006277.A1/Final (Alexandria, Va.: Center for Naval Analyses, 2002). This document reports on the Naval War College's Scud Hunt experiment and offers some excellent prescriptions for achieving additional progress in game-based C2 experimentation.
9. Carl von Clausewitz, *On War*, ed. and trans. Peter Paret and Michael Howard (Princeton, N.J.: Princeton Univ. Press, 1976), p. 156.
10. Ibid., p. 167.
11. Ibid., pp. 159–61. The Treaty of Campo Formio of 17 October 1707 between France and Austria produced, aside from various territorial annexations and guarantees of support, the latter's retirement from the War of the First Coalition (1793–97, originally pitting Austria, Prussia, Great Britain, Spain, Sardinia, and the Netherlands against France).
12. For the Global games see Rubel, "War-Gaming Network-centric Warfare"; Kenneth Watman, "Global 2000," *Naval War College Review* 54, no. 2 (Spring 2001), pp. 75–88; Bud Hay and Bob Gile, *Global War Game: The First Five Years,* Newport Paper 4 (Newport, R.I.: Naval War College Press, 1993); and Robert H. Gile, *Global War Game: Second Series, 1984–1988,* Newport Paper 20 (Newport, R.I.: Naval War College Press, 2004).
13. *Global 2000 Network-centric Warfare: Gaming the Navy Capstone Concept for Operations in*

the Information Age (Newport, R.I.: Naval War College, December 2000). The report offers glowing endorsements of networked speed of command. The assessment of the national command authority play is that of the author, who was an observer during the game. See also Watman, "Global 2000."

14. Clausewitz, *On War*, p. 137. Clausewitz talks extensively and explicitly in *On War* about effects, except with much greater lucidity than is commonly found in the current literature, which is riddled with unsupported assertions and esoteric jargon.

15. Ibid., p. 178. A brief passage is referred to, but Clausewitz devotes considerable space to the difficulties of strategy, extolling its successful practitioners precisely because of the many imponderables at the strategic level.

16. Edwin Miller, *War Plan Orange: The U.S. Strategy to Defeat Japan 1897–1945* (Annapolis, Md.: Naval Institute Press, 1991), p. 168. Miller describes in this passage some of the Newport war games that indicated the folly of attempting to sail the U.S. fleet directly from Hawaii to the Philippines. However, despite these results, the "thrusters," who advocated such a strategy, held sway until the mid-1930s.

17. Mitsuo Fuchida, *Midway: The Battle That Doomed Japan* (Annapolis, Md.: Naval Institute Press, 1955), pp. 96–97.