2017

Changing DoD's Analysis Paradigm

John T. Hanley Jr.

Recommended Citation

Hanley, John T. Jr. (2017) "Changing DoD's Analysis Paradigm," *Naval War College Review*: Vol. 70 : No. 1 , Article 5.
Available at: https://digital-commons.usnwc.edu/nwc-review/vol70/iss1/5

CHANGING DOD'S ANALYSIS PARADIGM

The Science of War Gaming and Combat/Campaign Simulation

John T. Hanley Jr.

War gaming and military modeling have a well-documented history covering over two centuries, a period that coincides with the inception and evolution of formal professional development for military officers.[1] The term *war game* used here refers to "a warfare model or simulation that does not involve the operations of actual forces, in which the flow of events affects and is affected by decisions made during the course of those events by players representing opposing sides."[2]

Beginning with the early-nineteenth-century Prussian creation of war colleges to augment operational experience, professional military education involved a combination of the study of history and international law, the study of theorists who had written on the nature of war and strategy, practical exercises, and theoretical analysis as the means for understanding and developing military art and science. Carl von Clausewitz's *On War* and Antoine-Henri de Jomini's *The Art of War* competed for attention. Whereas Clausewitz treated war as a social phenomenon, rooted in the age of reason, Jomini believed in the existence of immutable principles of warfare, akin to Newtonian mechanics.

As war gaming became a routine part of Prussian military education, the Prussians attempted to create rigid rules for calculating the outcomes of engagements. Major powers around the world

John T. Hanley Jr. earned a doctorate in operations research and management science at Yale University, writing his dissertation on war gaming. A former USN nuclear submarine officer and fleet exercise analyst who employed military modeling to conduct campaign analyses, he used gaming extensively during his service with the first eighteen Chief of Naval Operations Strategic Studies Groups as an analyst, program director, and deputy director. He also served as special assistant to the Commander in Chief, U.S. Forces Pacific, in the Office of the Secretary of Defense (Offices of Force Transformation, Acquisition, Technology and Logistics, and Strategy), and as deputy director of the Joint Advanced Warfighting Program at the Institute for Defense Analyses. After serving as director for strategy at the Office of the Director of National Intelligence, he retired from government in 2012 and is now an independent consultant.

believed that war gaming contributed notably to the Prussians' success in 1866 and 1870. However, as the popularity of war gaming spread following the Prussian victories, semirigid and free-form adjudication based on the game director's judgment became more popular.[3]

War colleges used war gaming as a basis for both practical exercises and theoretical analyses. Both war colleges and military staffs used war gaming to develop strategy. In addition, in the early twentieth century, quantitative military modeling outside of war gaming was adopted more widely. New techniques were formulated, such as Lanchester equations, which Frederick W. Lanchester published in 1916.[4]

During World War II, the United States and the United Kingdom instituted operations evaluation groups, consisting of scientists, to quantify the outcomes of military practices and seek improvements. These groups observed operations, collected data, and created models of military operations analogous to the models they used in scientific endeavors. Following World War II, the U.S. government established federal contract research centers to continue this practice in peacetime.[5] The Navy transformed its Operations Research Group into an Operations Evaluation Group that became the Center for Naval Analyses. The Air Force established RAND. The Army established its Operations Research Office at Johns Hopkins University, which became the Research Analysis Corporation. The Joint Chiefs of Staff (JCS) founded a Weapons Systems Evaluation Group that became the Institute for Defense Analyses.[6] Initially these organizations provided mechanisms for contracting university professors; eventually, they developed permanent staffs.

In the long-term competition with the Soviets, the emphasis shifted from operations research to systems analysis: operations research focuses on analyzing operations to support commanders; systems analysis focuses on supporting the Pentagon's policy and procurement bureaucracies by attempting to quantify the effects of proposed platforms and weapons systems employing advancing technology. An expansion of the practice of quantification to optimize operations spread from the military to industry, leading to the creation of operations research as a discipline.

In 1961, coming from Ford Motor Company, Secretary of Defense Robert S. McNamara established the Pentagon's Planning, Programming, and Budgeting System and a Systems Analysis Office to oversee the selection of military systems and force allocation and determine how much was enough to invest in defense.[7] Alain C. Enthoven founded the Systems Analysis Office on well-intentioned tenets.[8] However, competing interests and divisions in staff responsibilities within and among the Office of the Secretary of Defense (OSD), the Joint Staff, and the

services and the need to align analytical processes with Pentagon staff procedures and budget cycles resulted in these tenets never being followed fully.

Computers rapidly expanded the scale of problems addressed in the 1960s and '70s. Computer-based campaign simulations that strung together and iterated sets of equations modeling combat became the primary method the Pentagon procurement bureaucracy used to undergird arguments for selecting one military platform or technology over another. As the Department of Defense (DoD) expanded its use of contractors to conduct analyses in the 1970s, a sizable industry emerged to support and embed Pentagon analytical practices. "Unfortunately, the trend over the last decades has been for DoD studies to become more focused on standard scenarios and big [computer] models."[9] On 8 May 2015, Deputy Secretary of Defense Robert O. Work and Vice Chief of the JCS Admiral James A. Winnefeld Jr. called for initiatives to renew war gaming within DoD.[10]

Scientific methods form the foundation for operations research. A frequent criticism of war gaming is that it is less scientific, and thus less useful for prediction, than computer-based combat/campaign simulation. This article examines war gaming and combat/campaign simulations against scientific standards to explore their usefulness and limitations and how they complement each other. Computer-based campaign simulation involves much larger uncertainties and indeterminacy than generally realized. Both campaign simulation and war gaming require the use of additional analytical techniques to validate and extend their findings.

Operations research is rooted in an interactive cycle of observing fleet/field operations, collecting data, modeling, collecting more data, proposing changes, then cycling through those results again. The original operations research groups involved interdisciplinary teams of scientists employing models and paradigms from their respective disciplines to understand military operations well enough to predict effects. DoD needs to overhaul its current analysis paradigm and its focus on individual major defense acquisition programs, weaning itself off large, computer-based campaign models. It should adopt analysis campaigns and cycles of research to meet growing security challenges within limited budgets.[11]

SCIENTIFIC INQUIRY

The System and Its States

Bernard O. Koopman begins his study of the logical basis of combat simulation with the following:

> Basic to any scientific examination of nature is the concept of the *system:* the set of interacting things considered. In a military action, the system is the totality of men[/women] and weapons involved, together with their environment: the medium in which the action occurs and which affects its course. And equally fundamental is

the concept of the set of *states* that the system can be in, just one at any given time. . . . In each case, the state of the system includes its physical state: positions and velocities of the units, condition of armaments, data-gathering status, and all the meteorological specifications. But how far into the mental state of the commanders must one go in defining the "state" of the system? This can only be settled by asking a second question, that of *the evolution of the state of the system with the passage of time.*

Classical physics has traditionally considered that the state of a system is only adequately described if, once the state is given, all later states are *determined:* Given any two similar systems in the same initial states, all their later states will be the same— provided that their environmental influences (external forces) continue the same. Thus, in Newtonian mechanics, the full and exact knowledge of the positions and velocities of the parts of a material system determine its whole future motion. But it is only in the simplest military operations that such an order of determinateness exists.

In far more cases, it is not feasible to specify the state of a system so that its subsequent evolution is determined. What is far more common is to have only *statistical determinateness:* in a large number of similar systems starting in the same state, the same *proportion* will go into any given later state.[12]

The premise of combat/campaign simulation is that the evolution of the states in some future combat can be determined adequately statistically. In war gaming, the state of the system evolves move to move through adjudication of player decisions. Keeping in mind the concept of states helps us consider the scope and limits of computer-based combat/campaign simulation and war gaming.

Scientific Standards

"Standards of scientific excellence, though they may occasionally be self-defeating, on the whole and in the long run make for success."[13] However, one must stipulate carefully what one intends when posing scientific standards, lest they become straitjackets. "The emphasis by historians and philosophers of science is that there is no such thing as *the* scientific method. The more realistic danger is that some preferred set of techniques will become identified with scientific method as such."[14]

As systems analysis took hold in DoD, those seeking to determine "how much is enough" sought to create models using equations that allowed quantitative comparisons to predict the costs and benefits of alternative systems. As computers became more powerful, DoD turned to quantitative combat/campaign simulations as a basis for major decisions, regarding them as more objective, rigorous, and useful than less-formal analytical techniques, such as war gaming. Such simulations were considered to be

- more objective, in the sense that computer models would support major decisions based on explicit criteria of national interest, not on compromises

among institutional forces, and provide open and explicit analysis (including transparent data and assumptions) available to all parties

- more rigorous, in the sense that computers would provide quantitative answers to support choices among explicit, balanced, feasible alternatives and allow reproducible runs for comparing alternatives

- more useful, in the sense that computers would allow more systematic analysis to predict the effects of decisions[15]

Therefore, objectivity, rigor, and usefulness provide the set of scientific standards used in this examination of combat/campaign simulation and war gaming.

Objectivity. "That is objective which insists on its own rights regardless of *our* wishes, and only experience can transmit its claims to us. Experience is ultimate because it confronts us with a continuous ultimatum. For a man to by-pass experience in the pursuit of truth is to make himself God. . . . The subjectivist lives in a fool's paradise."[16]

Objectivity equates to "the intersubjectivity of findings independent of any one person's intuitive judgment."[17] Demanding intersubjectivity requires that "a scientific observation could have been made by any observer" and "testifies that the observation is uncontaminated by any factors save those common to all observers."[18] "For an enterprise to be characterized as *scientific* it must have as its purpose the explanation and prediction of phenomena within its subject-matter domain and it must provide such explanation and prediction in a reasoned, and therefore intersubjective, fashion. . . . While precise predictions are . . . preferred to vague ones, a discipline which provides predictions of a less precise character, but makes them correctly and in a systematic and reasoned way, must be classified as a science."[19]

Concepts lead to observations, which then lead to theories and laws. Laws have counterfactual force, carry explanatory force, and support prediction. They serve as standpoints from which we can survey for exceptions. They provide the basis for broader theories that advance the understanding of complicated and complex phenomena. A definition of an expert is one who knows what context must hold for a law to apply.[20]

Basic Newtonian physics involves laws strictly determining the relationships between actions and their effects. But even physics requires statistical laws to explain quantum phenomena, thermodynamics, etc., and cannot predict the behavior of many multibody problems and other chaotic systems. Statistical laws permit probabilistic explanations for phenomena involving statistical indeterminacy.[21] Similarly, systems involving human behavior admit quasi laws or tendency laws.[22] "In order for the [quasi] law to be valid, it is not necessary that no apparent exceptions occur, it is only necessary that, if an apparent exception

should occur, an adequate explanation should be forthcoming."[23] Statements such as "fear, honor, and self-interest are the fundamental causes of war" qualify as quasi laws.

Both war gaming and combat/campaign simulations are pseudoexperiments: experiments carried out on a model instead of in reality.[24] The person or team designing the experiment reduces a substantive problem to a conceptual model on the basis of the perception of what is relevant to the problem. This conceptual model is a *world,* defined as the object or system about which a person is concerned. A *state* of the world is a description leaving no relevant aspect undefined. A *true* state of the world is a state that does in fact obtain, i.e., the true description of the world.[25] The conceptual model is reduced further to physical and semantic (quantitative and relational) models, each equating to a theory of behavior of the subject matter, employed in the analysis to determine the true state.[26] If the experiment serves its purpose, this system of models produces an outcome that can be generalized by induction to advance a substantive conclusion.[27]

The character of military (and civil) operations involves both "an evolving physical system, and . . . an unfolding set of plans, intentions, reasoning and counterreasoning of the men [and women] engaged in the action, the commanders."[28]

War gaming addresses the plans, intentions, reasoning, and counterreasoning of the roles represented in the game. It highlights "predictions regarding the behavior of human organizations inasmuch as the latter can be simulated most effectively by having experts play the roles of certain members of such organizations and act out what in their judgment would be the actions, in the situation simulated, of their real-life counterparts."[29] Outcomes result from the interacting decisions and actions of the role players, as adjudicated by game umpires and game-control oversight.

> Epistemologically speaking, the use of an expert as an objective indicator . . . amounts to considering the expert's predictive pronouncement as an integral, intrinsic part of the subject matter, and treating his[/her] reliability as part of the theory about the subject matter.[30] Our information about the expert is conjoined to our other knowledge about the field, and we proceed with the application of precisely the same inductive methods which we would apply in cases where no use of expertise is made. Our "data" are supplemented by the expert's . . . valuations and by his[/her] judgments of relevance . . . , and our "theory" is supplemented by the performance of experts.
>
> In this manner the incorporation of expert judgment into the structure of our investigation is made subject to the same safeguards which are used to assure objectivity in other scientific investigations. The use of expertise is therefore no retreat from objectivity or reversion to a reliance on subjective taste.[31]

Computer-based combat/campaign simulations focus on physical aspects of combat. Human decisions are present and have a substantial impact on the

output, but are embedded in the simulation construction and the choice of inputs (data and models) rather than the decisions of combatants. To encompass human decision in statistical determinateness, one might turn to doctrine or, absent clear doctrine or future systems, query commanders for their expert opinions regarding decisions they would make given each possible state of the system. To be practical, this approach requires a world with few states. One also might assume that each commander is attempting to do maximum harm and seeks a course of action to minimize the harm to his/her forces, using the *minimax* convention of game theory.[32] "A more general method of this sort is for each commander to maximize his[/her] own value function—not necessarily the negative of his[/her] opponent's."[33] This approach to combat/campaign simulation assumes that once the statistics of human decision are incorporated into the model, what remains is the statistically determinate evolution of the military system. But separating the human from the physical model often leads to erroneous conclusions. Barry Watts's research indicates that, rather than having been let down by their radars and missiles, 80–90 percent of the pilots shot down in Vietnam and Korea never saw their attackers until it was too late to react.[34]

By virtue of the statistical determinateness, the basic process is stochastic. That is, there is a definite probability—the transition probability—that if the system is in state x at time t it will be in state x' at time t'. *"Evidently, if the values of the transition probabilities a(x, t; x', t') were all known, the probabilities of every outcome of the battle would be known—and this for every assumed starting state"* (italics in original). Thus, the whole problem of the quantitative study of military operations is that of finding the transition probabilities from knowledge that can reasonably be obtained. "[A]ll the standard analytical models, Monte Carlo simulations, etc., fit into this scheme."[35] Clearly, one also must have knowledge of the transition rates to specify at which time t' the new state x' obtains.[36]

In practice, analyzing stochastic processes also employs the Markovian assumption, which holds that, faced with the same state, the transition probabilities for the system remain constant throughout the process. In the context of human decision, this means that no learning from previous states, no history, affects the process.
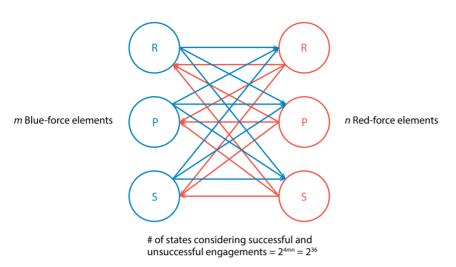
> Of course, when methods of computer simulation are made in the usual way they depend for their validity on the Markov property, but when this does not apply . . . the numerical results, however realistic they may appear, are without logical basis—at least until they are *proved* to give an acceptable degree of approximation. The act of simplifying and still retaining the Markovian character—as well as operational realism—is an art as well as a science. Success is more apt to be achieved by limiting the objective of the study to the answer of a precise question rather than a diffuse multitude.[37]

In summary: to assume that such a use of machines gives even approximately valid information about the military operation is to assume the following:

- The human uncertainties have been removed.

- The combat situation involves a system that is, at any time, in an objectively describable state (presumes transition probabilities and rates are known).

- The situation's state transitions are Markovian.

- Its stochastic equations can be satisfactorily approximated by difference equations without losing their Markovian character.

- The repetition of runs gives, by the law of large numbers, satisfactorily accurate and reliable values of the desired probabilities.[38]

At this point, the number of states involved in a combat simulation is worth considering.[39] Consider an engagement involving m units on the Blue side and n units on the Red side.[40] Indicate that a Blue unit has engaged a Red unit by drawing a blue line between the two units. Similarly, use a red line for a Red unit engaging a Blue unit. "The resulting colored graph indicates the state of our system. How many different graphs are possible? Of the *mn possible* ways of drawing the blue lines, any one can actually be drawn or not. Hence, there are 2^{mn} possibilities for the blue lines; and similarly for the red. Consequently, there are 2^{2mn} possible colored graphs."[41] See the accompanying figure for a depiction of the case for a combined-arms rock-paper-scissors contest in which all "units" could engage simultaneously or in any order. The number of states of this world for a single battle is 2^{18}. If we consider whether each engagement is successful, we double the number of states to 2^{36}. Each additional consideration enlarges the exponent for computing the state space.

We can perform a mind experiment to estimate how large the state space would be for a battle that a "perfect" parallel computer the size of the universe, given the time of the universe, might compute. In this parallel computer, the processors are as small as protons, they operate at the speed of light, and they are packed densely into the volume of our universe. Each processor is assigned a distinct engagement to calculate, can compute the outcome instantaneously, and can fetch a new engagement in 10^{-23} seconds, an approximation of the time it takes light to go the diameter of a proton. Given 10^{45} processors per cubic meter, 10^{81} cubic meters in the universe, 10^{23} calculations per second, and 10 seconds as an epoch a bit longer than the age of the universe, this computer could perform 10^{168} calculations, or about 2^{558}.[42] If $4mn = 558$ and we examine the same number of force elements on both sides, *this "perfect" computer could calculate the states for an engagement with just less than twelve force units per side*. Note that this

COMBINED-ARMS ROCK-PAPER-SCISSORS

m Blue-force elements

n Red-force elements

# of states considering successful and
unsuccessful engagements = $2^{4mn} = 2^{36}$

Source: Koopman, "A Study of the Logical Basis of Combat Simulation," pp. 871–72.

formulation of the engagement does not consider the timing of engagements, which would vastly increase the possible states of the "world."

Although simulations such as those of one-on-one air or naval combat might be reduced to a computable number of states, force-on-force combat and campaign simulations quickly exceed the number of states that admit of brute-force computation. So, how are these simulations implemented? By using a combination of shortcuts (heuristics) and clever analysis. These heuristics are essentially quasi laws whose application requires the contribution of experts who understand well the scope of those laws' applicability. Combat/campaign simulations often use expected-value models to determine what would happen "on average," rather than Monte Carlo simulations. Increasing the number of runs does not increase statistical prediction by the law of large numbers in these simulations, as the expected value provides a determined outcome for each run. Lanchester equations—developed to help predict the outcome of naval and land battles—most often use expected values, but can employ Monte Carlo techniques.[43] Varying the exponent used in Lanchester equations between square and linear laws essentially reflects the command and control and operational concept employed in the engagement. The complexity of ground models results in heuristic techniques such as weighted effectiveness indices / weighted unit values or qualitative judgment models to calculate engagement outcomes. All these approaches involve subjective judgments and the insights of the analyst/team developing and using the model. In a combat/campaign simulation, the analyst/team must use subjective judgments to anticipate every interaction represented in the simulation,

supplement missing data, and create models that have not been validated in actual operations or exercises.

The works of Wayne P. Hughes, Glenn A. Kent, Bernard O. Koopman, and Paul K. Davis, among others, suggest clever approaches to overcoming computational limitations of brute-force calculations and appropriate forms of analysis.[44] With the development of complexity sciences, computers came to be used to simulate cognitive and other processes, rather than to solve equations. As Deep Blue and AlphaGo have demonstrated, in games of finite size with well-specified rules, computers can use artificial intelligence (AI) techniques to top human performance.[45] However, current DoD computer-based campaign simulations use brute-force calculations. They have yet to incorporate agent-based models, automatons, fitness landscapes, genetic algorithms, or other techniques from complexity science. RAND incorporated some AI techniques into campaign simulations in the 1980s, but DoD chose not to employ those features in the simulations it adopted, instead staying with the types of deterministic and stochastic models Koopman addressed.[46]

So, how do we assess objectivity, given the logics of combat/computer simulation and war gaming?

Guidelines for the practice of operations research, although written with military modeling in mind, apply equally to war gaming and to combat/campaign simulation.[47] Significant distinctions between good operations research practice and other scientific inquiry include a presumption of the existence of a client (sponsor) and the complications presented by security classification and proprietary work. Close cooperation with the client in framing the analysis is good practice common to any technique of analysis.

In war gaming, a design and development team develops the scenario and reference materials (e.g., commander's intent, task organization, subordinates' missions, orders of battle, unit locations, weather) to establish the world and its initial state and develop prebriefings to immerse players into the game. The team identifies the number of competing sides, the scope of disciplines required, the command echelons represented, the bureaucratic verisimilitude desired, and the number and expertise of role players needed to accomplish the game objectives. Team members also design the information conditions: the information available to each side and its flow, the communications techniques and their verisimilitude to accustomed formats, the physical arrangements, the move structure, and the game rate to arrive at a desired culmination point.[48] To facilitate decision making, they construct move forms and provide for feedback among the participants. For adjudication, they select methods and models (quantitative and qualitative) used to change the world state resulting from each game move, and the qualifications

and organization of game controllers and umpires. They also anticipate control inputs of plausible events (usually wild cards, such as rogue actions or accidents to initiate conflict) to shape player decisions to achieve game objectives.[49]

Combat/campaign simulations similarly frame the world for the purpose of the pseudoexperiment and provide input data to establish its initial state. Whereas war-game design involves detailed considerations of context for role-player decisions—with particular attention to information conditions (who knows what and acts when), as discussed above—combat/campaign simulations remove human decision. Information conditions are embedded in the combat models. The models selected incorporate a theory of command and control and the concept of operations in their code—consciously or not. The analysts / team members develop or choose models and techniques they judge appropriate to the study, on the basis of their expertise. "A fundamental truth in analysis is that scenarios drive the answers. Thus, much effort should go into conceiving and tuning the scenarios used and specifying uncertainty ranges. This should be a deeply analytic affair rather than the result merely of creative people spinning stories that raise interesting issues."[50] Whereas game scenarios are necessarily rich, to provide the context essential for expert role playing, the world of the combat/campaign simulation employs sparse scenarios, with only the data needed to perform the calculations.

In war gaming, a control team and umpires run the simulation. They execute the game design, adjudicating changes in the "true" state of the world using the decisions of the role players, their quantitative models, and their judgments, taking into account the game's objectives. In computer-based combat/campaign simulations, the computer computes the state transitions and the analyst decides what constitutes a stopping point or state for ending the computer run. Both war gaming and combat/campaign simulation also involve analysts who observe, record, analyze, and report on the pseudoexperiments. Costs and time available to design, develop, and run the simulation and subsequent analysis constrain both types of simulation.

Both war games and combat/campaign simulations involve clients, designers, developers, and analysts employing informal reasoning processes and subjective judgment in creating their theory of the world under study. The totality of the participants, models, and data employed in these simulations and the relationships among them represent the theory of the war game or combat/campaign simulation. In the case of games, this includes the role players, umpires, and control team, in addition to any quantitative models used in adjudication. In combat/campaign simulation, it involves the treatment of human decision and the concepts and information conditions embedded in the models, as well as the flow of outcomes from one process into the next (e.g., who attacked whom first).

The motivations, expertise, tastes, beliefs, and reliability of all human participants involved in the pseudoexperiment are thus integral, intrinsic parts of the subject matter, and therefore parts of the theory expressed in the war game or combat/campaign simulation.

Given the subjective judgment involved in defining the world and assessing its true state in both forms of pseudoexperimentation, objectivity comes from intersubjectivity. For combat/campaign models, this involves techniques such as the use of models that have shown value in actual combat (e.g., those developed using combat data in war) or that have been verified in field/fleet exercises employing actual forces. A weaker, but essential, form of verification for assessing the objectivity of scenarios, models, and data is to open them to debate and review—realizing that "sunlight is the best disinfectant"[51]—while recognizing the pitfalls that may result from political logrolling. Interpreting the structure of relationships in and among models and how to sequence these models in pseudoexperiments relies on the subjective judgment of the developers. It also requires developers who know what factors are indeed relevant to the world under study; e.g., attacking air forces on the ground can be a way to gain dominance of the air.[52]

Operational experience is useful in developing the expertise to make such judgments. Gaming has an advantage in this regard: "In operational gaming, the simulated environment is particularly effective in reminding the expert, in his[/her] role as a player, to take *all* the factors into account . . . that are potentially relevant; for if he[/she] does not, and chooses a tactic or strategy which overlooks an essential factor, an astute 'opponent' will soon enough teach him[/her] not to make such an omission again."[53] "People sensitive to a variety of responsibilities collaborate, applying the criteria that are relevant to their own interests, making estimates that reflect their own kinds of knowledge, and putting themselves in a mood to worry about probabilities rather than just a list of possibilities. They really live through a simulated crisis and not only learn things about their plans and their predictions but learn something about the nature of crisis."[54]

Gaming allows all participants—role players, control team, umpires—the right of *reclama* when they need additional information for a decision or question the adjudication of a move. Manual games are particularly useful in this regard. In manual games (which may employ computer calculation in adjudication), players must make decisions from one turn to the next, taking into account the current situation; and procedures used to evaluate the consequences of the players' decisions must be quite clear to the players—simple enough for them to understand.[55]

Gaming achieves objectivity by allowing all experts involved to share both their formal and informal reasoning explicitly and openly. Deficits in knowledge and both consensus and dispute are evident, thus providing a foundation for further inquiry. Critiques provide ways to improve the games continually. By

Hanley: Changing DoD's Analysis Paradigm

contrast, the details of combat/campaign simulations are evident only to the analyst/team that developed the simulation or to someone willing to conduct a detailed study of the data and models used in the simulation.

In all science, good practice calls for independent review. However, in this field the practice is difficult to follow owing to the additional costs and the limitations that security and proprietary concerns impose. Clients often use "need to know" as an excuse to hold details of the pseudoexperiments close. Often, when a contractor performs the analysis, the details of the pseudoexperiment are proprietary. The Operations Research Society of America promulgated "Guidelines for the Practice of Operations Research" as a consequence of a dispute in testimony to Congress over two studies of ballistic-missile defense that supported conflicting recommendations. The guidelines conclude as follows:

> The analyst, *as analyst,* must restrict his[/her] analysis to the quantifiable and logically structural aspects of the problem only. In complex problems, perhaps the most valuable thing the analyst can do is to point out to his[/her] client that there are uncertainties deriving from such factors as:
>
> - Lack of agreement on means of evaluating the worth of complex systems.
>
> - Uncertainty about the technical capabilities and costs of systems yet unbuilt.
>
> - Uncertainty about environmental and operational factors that influence performance.
>
> - Uncertainty about the future capabilities or intentions of possible opposition.
>
> The analyst should be prepared to engage in dialogue with the client and other advisors to consider how other value systems, assumptions, and conditions might influence conclusions. . . . The analyst's job, especially in tough policy questions, is to analyze and help illuminate, and this means having the qualities of humility and openness necessary to participate in open dialogue with the client and other advisors.[56]

Subjective judgment enters once more in deciding what actions to take as a result of the war game or combat/campaign simulation. Here again, gaming has an advantage in that those who will decide what actions to take (or those on their staffs) have participated in the experiment—the decision makers learn directly from the game experience. In contrast, in the case of a combat/campaign simulation, decisions on actions to take depend on how the analyst/team used subjective judgment to frame and report the results, adding another layer of interpretation to the decision process.

A critique of games is that the subjective judgments of the experts involved make them irreproducible. A question for combat/campaign simulation is whether, given the same subject matter, independent teams would select the same

scenarios, models, data, structures, and relationships, among them producing the same results and the same analysis on the basis of those results. A 1973 General Accounting Office (GAO) report following promulgation of the Operations Research Society of America's "Guidelines for the Practice of Operations Research" found shortfalls in independent checks to ensure the accuracy, timeliness, consistency, and overall quality of the data—about 18 percent of the models were considered generally transferable for use by another person or another site—and "[t]he choices of scenarios, equipment performance, and personnel operations are based somewhat upon unknowns and uncertainties. The extent that the model reflects the real-world situation depends on the accuracy of the model-builders' judgment."[57]

Relying on intersubjectivity generates concerns centering on the role of bias in forming belief. Critiques of limitations on human judgment and decision are legion. Irving L. Janis and Leon Mann provide a framework for how people make decisions (unconflicted adherence, unconflicted change, defensive avoidance, hypervigilance—as with a crowd heading for the exits in an emergency—and vigilance).[58] Even vigilant decision making may be subject to cognitive, egocentric (self-serving motives), or affiliative (organizational or social acceptability) constraints.[59] Charles Pierce provides a set of methods for fixing belief similar to those above, including tenacity (such as a child hears from its mother), authority (the will of an institution), apriority (the adoption of self-evident assumptions that are clear to the user, but to no one else), and finally the method of science.[60] Humans are exceptionally poor at assessing subjective probabilities.[61] "When we pit [political] experts against minimalist performance benchmarks—dilettantes, dart-throwing chimps, and assorted extrapolation algorithms—we find few signs that expertise translates into greater ability to make either 'well-calibrated' or 'discriminating' forecasts."[62] Humans make decisions on the basis of their tastes (preferences) and beliefs (subjective probabilities).[63] They persist in even discredited beliefs. The Central Intelligence Agency provides four reasons for this persistence: "We tend to perceive what we expect to perceive; mind sets tend to be quick to form but resistant to change; new information is assimilated to existing images; and initial exposure to blurred or ambiguous stimuli interferes with accurate perception even after more and better information becomes available."[64]

However, "[w]hether a distortion common to all can nevertheless be said to yield something objective is a philosophical question that has no bearing on the conduct of the human enterprise of science. The methodological question is always limited to whether what is reported as an observation can be used in subsequent inquiry even if the particular observer is no longer part of the context."[65]

Thus, for objectivity, the analytical team / rapporteurs in war gaming should note the assumptions and design choices that went into the game, arguments

both for and against a particular course of action by teams making their decisions, and what outcome the team hoped to achieve, capturing both consensus and disputes. Objectivity in combat/campaign simulation involves using models validated by observation of operations or field/fleet exercises, employing data collected from those exercises. Studies done in advance of actual operations should be compared with what transpired and why.[66] Analysts of both war games and combat/campaign simulations should keep in mind the motivations and beliefs of the participants and should extract from the experiment that which subsequent inquiry could verify or refute.

Rigor. Aristotle said, "A well-schooled man is one who searches for that degree of precision in each kind of study which the nature of the subject at hand admits." Kaplan goes on to note, "Another failing of models—more accurately, of model builders—consists in an undue emphasis on exactness and rigor."[67] Used in this way, rigor too often is equated to precise quantification, usually in the form of increasing the number of significant figures relative to a decimal point. However, the tests of rigor are whether (1) the analytical techniques used are appropriate to the subject matter, (2) we can articulate clearly the details of the method used and how we arrived at conclusions, and ultimately (3) we can state what valid lessons the study produced. Employing analytical techniques that provide overly exact answers that do not reflect the uncertainties and indeterminacy inherent in the subject matter are not rigorous.

Rigor is related closely to objectivity. It demands careful attention to the design of a war game or combat/campaign simulation to achieve the objectivity described above. It also requires efforts to understand the quality of data used in quantitative models, estimating the range of uncertainty in quantitative results, and framing conclusions in quasi law–like statements that reflect the consensus and disagreement of those involved in the pseudoexperiment. New understandings of chaos and complexity also raise questions regarding the treatment of human action in combat simulations.

A first test of rigor is the data used in quantitative and semantic modeling. The most reliable data are collected during operations or exercises that are essentially the same as those represented in the model. Operations research originated during World War II with the collection of data, then the use of those data to develop models of the operation under study.[68] Many of those models and the modeling techniques have persisted, but sustained efforts to collect data at sea or in the field are now rare. The 1973 GAO study found that in 85 percent of the cases submitted to the researchers, DoD activities used data obtained from sources other than field exercises or actual experience.[69]

Beginning with its World War II experience involving malfunctioning torpedoes, the U.S. submarine force began collecting data on every torpedo fired.

When tasked with creating an antisubmarine warfare (ASW) capability in 1949, Submarine Development Group 2 developed a process of designing exercises to test technology and tactics, collecting data on system and platform (including crew) performance during those exercises, using the submarine approach and attack manual to standardize the data, and reconstructing the exercises to quantify the results.[70] Using this process, the submarine force went from having essentially no ASW capability in 1949 to having the world's premier ASW capability in 1969. The process led to continual improvement of the search and combat models used in war games and combat/campaign simulations. At-sea exercises discovered and corrected errors in search models implemented on computers.[71] The Navy used a similar approach in its Tactical Development and Evaluation Program and some equipment-development programs in the 1970s and '80s.[72] However, oddly enough—given accelerating demands for data—as computer simulation became more popular in the Pentagon for platform and weapon systems analysis (what DoD calls program analysis), emphasis on prototyping equipment and collecting data on processes and performance at sea and in the field waned. Structured operational testing and evaluation for systems in development largely replaced mission-oriented operations analysis involving all aspects of the system's use and its effectiveness as one of a suite of systems.

Recent efforts to return to the roots of operations research have encountered difficulties. During Operation IRAQI FREEDOM, initiatives to put analytical teams into the field were severely limited by commanders' concerns about protecting the analysts—and controlling the data. Although we have been fighting in that region for a decade and a half, data on processes and performance from the field have not been the source for modeling and experimentation that they were in World War II.

The majority of friendly-force data used in computer-based combat/campaign simulation come from structured operational testing and evaluation of system performance (which may or may not reflect its performance in actual field/fleet use, with different concepts of employment) or from expectations of future system performance based on key performance parameters used for design. However, data from structured tests have not proved reliable. In World War II, "experimental results overestimated the casualty production rate for tanks by a factor of two; for artillery duels by a factor of three; and for pure infantry actions by a factor of seven."[73] Given the human penchant for survival and the fog and friction of war, structured tests provide overly optimistic estimates.

Lest you think we are better off now with modern computers and powerful algorithms built into our best models, here is a more recent example. The U.S. Navy depends mightily for defense of the fleet on the Aegis missile system. Using data from controlled experiments at sea, one may calculate that if you shoot two missiles at an

incoming missile and they are operationally and statistically independent of each other, and if you also add some point defense, you can expect to shoot down 90 percent or more of the attacking anti-ship cruise missiles. What is the combat record? First, in battles at sea warships of other states have averaged around 75 percent success in defending themselves. On the other hand, all of their success must be attributed to soft kill and point defense weapons, not to surface-to-air missiles [SAMs]. Second, there are several instances of warships that might have defended themselves but did not, illustrated by the recent successful missile attack on the Israeli missile ship *Hanit*. Navy analysts will also remember the Exocet hits on the defendable USS *Stark* and HMS *Sheffield*. Third, in the entire record of over 220 missiles fired on ships at sea starting in 1967, only one anti-ship missile has been shot down by a SAM.[74]

Models predicted the United States would incur thirty thousand casualties in Operation DESERT STORM, not the roughly three hundred that actually occurred; and half of those casualties did not occur in battle.[75] Models for casualty estimates almost never include friendly fire. Even when friendly-force data are available in a combat model, factors such as the reliability and effectiveness of allied and adversary weapons, the proficiency of an adversary in using counterfire or countermeasures that depend on the adversary's training, etc., must be estimated. Key data disputes "often center around order of battle, unit effectiveness, munitions quantities, chemical warfare performance degrade values, advance rates, sortie rates, and concepts of operation [CONOPS]. More time is spent instantiating and refining CONOPS information than systems performance data. Hence the obvious utility of wargames to understand CONOPS and the flow of the warfight."[76]

In World War II, the operations evaluation groups determined that a simple estimate of the error in a model is *the individual percentage error of the data times the square root of the number of data elements*. For a model with five thousand data entries and a tight error range of 10 percent, this equates to a factor of seven.[77] The 1973 GAO report found that 27 percent of the models they examined had over ten thousand coded instructions. Campaign models that DoD currently uses typically have on the order of one hundred thousand data elements and hundreds of equations and semantic models establishing the relationships among the data elements. Mistakes in the internal validity of computer models resulting from treating continuous functions as discrete and stipulating relationships for which no theory or data exist to allow computation compound the errors in the final calculation.[78] Adding detail to a combat/campaign simulation may or may not improve the rigor, but it surely will increase the uncertainty of the calculation.

Understanding this principle, the members of the World War II Operations Evaluation Group used a hemibel (half a decibel, or a factor of about three) rule. If they could not demonstrate factor-of-three improvements in a recommended change, they were uncertain that they had sufficient accuracy to merit

the recommendation, particularly considering the time and costs involved in changing operational practices. At a recent Military Operations Research Society workshop, a section leader informed me that military operations research no longer uses the hemibel rule. Why not is unclear.

The use of combat models to adjudicate war games is subject to the same concerns as is their use in campaign simulations. However, employing models that participants can question and umpires can explain adds both objectivity and rigor to the enterprise.

Whereas combat/campaign simulation requires the analyst/team to represent all indeterminacy as statistical, war gaming specifically addresses strategic and structural indeterminacy. Strategic indeterminacy means that the outcome largely is determined by the interaction of role-player decisions and the adjudication of control/umpires (who may be considered additional actors). Structural indeterminacy involves uncertainties in appropriately bounding the subject under study, determining which elements are relevant to include in characterizing the state of the world, and understanding the relationships among those elements. Manual games are good for the following:

- study of partially understood dynamic processes

- study of partially understood force interactions

- building of players' backgrounds for future study and analysis

- continual game improvement on the basis of players' criticisms[79]

Where the fundamental character of the subject under study involves strategic and structural indeterminacy, war-gaming techniques are more appropriate than combat/campaign simulation. Adding the data and formalities needed for computation detracts from, rather than adds to, rigor.

Usefulness and Value. The final criterion for science under exploration is the value or usefulness of the study or, in our case, the pseudoexperiment. Usefulness is the ability to use the experiment to take appropriate action. It presumes objectivity and rigor.

DoD turned to computer-based combat/campaign simulation because it desired methods that could produce rapid, objective, rigorous simulations to examine contingencies involving different adversaries to predict force requirements, study strategic/operational concepts, and compare costs and effects of alternative new platforms or weapons systems. DoD found these simulations useful in providing a common basis for making comparisons on a timeline consistent with annual program and budget development.

However, the "method of Monte Carlo [or any other form of combat/campaign simulation] has one particular value: its *educative* or *intuition-building*

effect on those who behold the actual performance of the process. It allows the results of experimental variations of certain factors of the situation to be perceived in a direct and life-like way. This appearance of realism is so great that it has often led observers to forget that they were not in fact observing nature directly: a disastrous error."[80]

The predictive value of a large-scale, complicated combat/campaign computer simulation depends on how the analyst/team represents the results. Good, scientific analysis of computer-based campaign simulation can support quasi laws such as the identification of governing factors, but not strictly statistical or deterministic answers. Also, the premise that changing the characteristics of one system while leaving the rest of the world the same can determine an outcome assumes no feedback between the change and the rest of the system (e.g., that a change in combat capability will not influence commanders' decisions and CONOPS). However, DoD's use of computer-based simulation seeks to predict outcomes rather than to develop deep understanding of the factors governing the outcomes of battles and campaigns. Rarely do reports address governing factors or attempt to quantify the uncertainties inherent in the simulation.[81]

When DoD clients are facing a decision, telling them that their simulation identified topics that require future study is rarely what they want to hear. However, failure to identify unresolved issues from the pseudoexperiment obfuscates important uncertainties that should be considered. Science values the so-called heuristic fertility of studies rich in implications for further observations, experiment, or conceptualization.[82]

Making predictions from games presents challenges similar to making predictions from combat/campaign simulations, with the added proviso that although there is widespread skepticism about accepting any prediction of human behavior —much less quantified predictions—from a game, predictions derived from computer models are widely accepted. Yet although experts making stand-alone predictions are unreliable, "[e]xperience has shown that people often tend to adopt the same solutions to similar problems. Insofar as this is true, a realistic war game may predict the future, or at least some aspects of it[,] quite accurately."[83] Where games have preceded military battles and campaigns, they have demonstrated value in anticipating adversary tactics and courses of action and the many governing factors needed to prosecute battles and campaigns successfully. Examples include the following:

- Naval War College (NWC) games anticipating tactics in the Russo-Japanese War

- battle of Tannenberg gaming by both the Russian and German general staffs

- German general staff gaming of the Schlieffen Plan before World War I

- gaming different strategic approaches for a war with Japan at the College between the world wars

- Japanese gaming of the battle of Midway

- NWC gaming of the naval mining campaign against the Japanese in World War II

- German and Russian general staffs gaming the German invasion of Russia (Operation BARBAROSSA) in World War II

- Israelis' gaming before their operations

- U.S. Joint Staff gaming in anticipation of North Vietnam's Tet offensive

In almost all these cases, the games accurately predicted factors driving the success of future operations. However, in many cases the military system was unable to adapt in a timely fashion or the games had no effect on the political leadership conducting the war. Sometimes senior military leaders rejected game results.[84]

The Chief of Naval Operations Strategic Studies Group (CNO SSG) conducted a game exploring the implications of an Iraqi invasion of Kuwait in February 1990, before the actual invasion in August. Although the game had Iraqi forces advancing into Saudi Arabia toward the oil fields, otherwise it accurately anticipated a need for nontraditional coalitions, challenges in strategic lift, and the inadequate numbers of precision weapons on deployed Navy forces, among other things.[85] Yet many senior officials briefed on the game in March 1990 expressed no interest, viewing Iran rather than Iraq as the adversary of concern. Requests for game documents increased as Iraq conducted the invasion.

"Gaming is a powerful method for simultaneously mastering complexity, enhancing communication, stimulating creativity, and contributing to consensus and a commitment to action."[86] Thomas C. Schelling found the following: "First, the games are intensely stimulating; people are very active; ideas and conjectures get tossed around and analysed by a highly motivated group of people; a great deal of expertise is collected in a single room, expertise that is not often collected together; and people discover facts, ideas, possibilities, capabilities, and arguments that do not in any way depend on the game but nevertheless emerge in it." Players discover important facts that may never have occurred to them or are counter to what they understood (e.g., unprecedented acts excite attention, jurisdictional seams, and overlaps), and ways that players not represented in their usual thinking affect the feasibility and acceptability of possible courses of action.

> [T]he game, as a social and intellectual occasion, tends to be highly productive of little things of this sort. . . .

> Second, people . . . learn more . . . about a country, by going through a game . . . than by any cram course [of equivalent time]. . . . If somebody were going to be responsible for some operations in the Pacific Islands, or were going to be Deputy Chief of Mission in Finland, or going to run an [Agency for International Development] program in Cyprus, just putting him[/her] into a game for three days focused on the area he[/she] is going to would teach him[/her] more than he[/she] could get by any kind of briefings, lectures, reading program, or other program of self-improvement.

Third, acquaintance is made with people with whom one might have occasion to work in the future involving intense common experience in joint problem solving. These by-products are just preliminary to costs. People can spend the other 362 days of the year pursuing other forms of analysis and learning. "All analytical techniques, all research methods, all stimulants to the imagination are dangerous. This includes games. But games are not much worse in this regard than the other techniques."[87]

A critique of current professional military education is that it does not give officers a detailed appreciation of military geography in theaters of interest or of adversaries' weapons systems and their concepts for using them. Theater-level games are valuable for learning geography, including the military geography of basing; the kinds and ranges of adversary and allied forces that may come into play and the complications they represent; and the logic of adversary concepts, as represented by Red teams. At the tactical level, war games are good for teaching junior officers the capabilities of adversary forces in an experiential way that tends to stick better than reading intelligence reports.[88] As the Prussian and German militaries recognized, games are exceptionally useful for developing an appreciation of command relationships and skills in writing orders and in working through control of forces in complicated situations.

Between the world wars, the German army (Wehrmacht) conducted field exercises during the summer and gamed when in garrison the rest of the year. During winter, each echelon, from the general staff to the company level, gamed their roles in the operations contemplated, then took what they gamed to the field the next year, beginning with company-level exercises and culminating, usually in August, in as large-scale an exercise as they could manage. With the army restricted in size by treaty, the games aimed to teach each rank, career enlisted and officer, how to perform at two ranks senior so the army could expand quickly. During war, these games became rehearsals for upcoming operations and occasionally continued as battles were being fought. The games were of great value to the Wehrmacht for developing concepts such as the blitzkrieg, and for developing its operational competence when it had sufficient forces to retain the initiative.[89]

The interaction of experts trying to achieve opposing aims within the context provided in the scenario helps ensure that relevant factors are not overlooked.

Games provide a basis of shared experience and a common vocabulary.[90] Whereas creativity of the analyst in combat/campaign simulation is reflected in the coding and analysis, the Markov assumption does not allow for learning during the game. Including learning algorithms (e.g., Bayesian calculations) in the code further complicates analysis of the results. In war games, the role players adapt to each state of the world, as provided by game umpires and control. Courses of action that do not provide desired results lead to reexamining possible approaches and objectives. New ideas that do work become apparent to all participants, contributing to the consensus needed to generate commitment to a course of action. Concerns over the appearance of realism in gaming represent the same risks and unintended consequences as those resulting from combat/campaign modeling.[91]

The scope of issues amenable to war gaming exceeds that of combat/campaign simulation. Manual war gaming is uniquely suited to increasing our understanding of and appreciation for the information dimension of warfare.[92] Ultimately, military operations are about influence: deterring or compelling change in others' actions inconsistent with one's political aims, while reassuring and encouraging others' actions that are consistent with one's political aims. The critical feature of a game, as opposed to computer modeling or any other forms of one-sided analysis,

> is that at least two separate decision centers are involved, neither of which is privy to the other's planning and arguing, neither of which has complete access to the other's intelligence or background information, neither of which has any direct way of knowing everything that the other is deciding on. . . . What this mode of organization can do that can not otherwise be done is to generate the phenomena of understanding and misunderstanding, perception and misperception, bargaining, demonstrations, dares and challenger's [sic], accommodation, coercion and intimidation, conveyance of intent, and uncertainty about what each other has already done or decided on. . . .
>
> . . . If I draw a face with a hidden picture there is no way for me to tell how hard it is to see the face except to show the picture to somebody. . . .
>
> It is the peculiar element of collaboration, communication, and bargaining, that is involved in any crisis game, that cannot be captured by "straightforward" unilateral analysis. . . .
>
> . . . [I]n arguments about the treasures or dangers that one may stumble on in games it is significant that there is at least *something* that games can do or generate that cannot be done or generated in any other way.[93]

Another value is that those who participate in a pseudoexperiment learn far more than those who receive a report of the study's findings. Few clients have the time or technical ability to understand the internal details of the combat/campaign simulation; they instead rely on their analytical teams to distill key findings

relevant to the objectives of the study. In contrast, war gaming facilitates participation by those who must make and implement decisions. Joint planning dictates that, ideally, "the individuals who were deeply involved in the development of the COAs [courses of action]" should participate in the gaming used to develop those COAs.[94] War gaming facilitates recognition-primed decision making that allows commanders and their staffs to adapt rapidly to emerging situations, using their experiences in games "demanding careful sequential analysis of plans, decisions, events, and intelligence."[95]

IMPLICATIONS FOR DOD ANALYSIS AND A WAY AHEAD

The principal implication of this assessment is that DoD should overhaul its analytical paradigm that began with the Systems Analysis Office and evolved with the development of computers. DoD should rely on talented analysts and not again make the mistake of attempting to create universal answer machines through standardized processes and techniques. The focus of analysis for acquisition and force development should shift from individual weapons systems to capabilities to conduct sets of missions. DoD should reinvigorate the examination of warfare and military operations to develop an appreciation of fundamental questions to focus analysis, balancing a marketplace of ideas and approaches with the instincts of its hierarchy to centralize planning. It then should employ analysis campaigns, using cycles of research focused on top decision makers' concerns, that incorporate the following:

- war gaming

- DoD's investment in large-scale campaign models, to develop intuition and help identify factors governing combat outcomes

- field/fleet operations analysis

- intelligence collection

- campaign analysis

- quantitative modeling using simple, understandable models that incorporate only governing factors derived from observation and analysis (as opposed to creating computer code for each combat process and adding more code to already complicated models to address new technologies and phenomena)

- the study of history and recent advances in complexity sciences, and complementary analytical techniques based on advances in artificial intelligence and cognitive and social sciences

- review of study results against actual operations

No new analysis paradigm can meet scientific standards without addressing the roadblocks created by the abuse of need-to-know strictures and proprietary control of analyses.

Avoiding Past Mistakes

With the recent policy to make more use of war gaming, the first principle for a way forward should be to avoid mistakes of the past. Efforts to use large-scale computer modeling to create universal answer machines were misguided. In its search for systematic analysis routines, the natural tendency of the Pentagon will be to create similar standardized systems of war gaming that would allow those developing procurement programs and strategists to "turn the crank" to address issues as they arise. However, even the most objective and rigorous efforts in the past have not produced the desired results, as the following examples indicate.

RAND Strategy Assessment System. In the 1980s, concerns over the ability to analyze a possible war between NATO and the Warsaw Pact leading to a nuclear exchange motivated the OSD Office of Net Assessment to sponsor RAND in developing the RAND Strategy Assessment System (RSAS). The approach was to combine the best features of war gaming and analytical modeling in a comprehensive, farsighted framework for comparing views rigorously and moving toward some conclusions. RAND formed a stellar team to do the work, led by Paul Davis.

To this effort, war gaming provided the following:

- the contextual richness of complete scenarios

- interaction of political and military factors

- operational constraints

- often-ignored features of real war (e.g., unconventional attacks against command-and-control communications)

- asymmetries in objectives and perceptions

- asymmetries in national forces, doctrines, and styles

- relatively realistic descriptions of military campaigns

- action and reaction among the nations involved in the conflict

Analytical modeling provided the following:

- clarity of assumptions and causality

- reproducibility

- logical structure and rigor

- efficiency, permitting many war games (multiscenario analysis)

- depersonalization, by laying issues out on paper logically[96]

To make gaming more efficient and rigorous, the RSAS approach used AI techniques to replace human teams. To make the process transparent, the design permitted human interaction at all levels, with the exception of some core model and execution coding. The intent was not to eliminate the role of expert judgment but "to capture most of the human-expert contribution in background research reflected in the models."[97] Computer code was written so that analysts knowledgeable in the subject matter did not need to have extensive experience to read and program decision rules.[98] The team intended that analysts and senior decision makers would be able to get definitive explanations and have the opportunity to change assumptions readily.

Departures from traditional analysis included automated game-based simulation to permit multiscenario analysis, heuristic rule-based modeling to make explicit the key assumptions on which outcomes depend, structured military campaign analysis, and interactive force-operations modeling. This would enable the analysis to treat interrelationships among strategic and nonstrategic forces; cut across theater boundaries, military services, and types of warfare; and reflect the effects of special phenomena such as unconventional warfare and failures in command and control.[99] The aim was not to predict outcomes but to understand what affected outcomes most.

In 1986, government agencies received the first installations of RSAS. An RSAS Steering Group, consisting of sponsors, developers, and users, approved requests to use the system. Although the RAND team intended that actual decision makers use the system for policy analysis, it proved too complicated to be of use in evaluating immediate operational situations, and high-level decision makers turned to their own analysts. RSAS was open to review, critique, and improvement. The challenge was that it was akin to an engineering library. One could investigate any subject, but only the developers could comprehend the whole system.[100]

As a spin-off from RSAS, RAND developed the Joint Integrated Contingency Model (JICM). It designed the model to be modular for transparency and to avoid needing to add hundreds of thousands of input variables. "As the model [JICM] was used in later years, however, the optional simplicity fell into disuse as users focused on getting the detailed databases 'right' (meaning agreed upon) for running standardized cases."[101]

Although RSAS and JICM were as objective, rigorous, and comprehensive as was practical, the limited interests and capacity of the DoD bureaucracy defeated RAND's sophisticated efforts to meet exacting standards of science.

Joint Warfare System. In a subsequent effort to allay concerns over the services using their own scenarios, models, and data, in the 1990s OSD began funding the Joint Warfare System (JWARS) to "support multi-billion dollar resource

allocation decisions and critical operational planning." JWARS was "a closed-form analytic simulation" using deterministic and stochastic models, including information operations, and "high-level abstractions of sensor and communications systems, the related information flows, imperfect perception of the battlespace, and command decision making."[102] The aim, as with individual service campaign simulations, has been to create a simulation to determine the effects of varying the characteristics of a system or concept by turning a crank, leaving the rest of the simulation unperturbed.

Given the expansiveness of the state space, the use of models and data based on judgment rather than observations from operations or exercises, and the likely feedback among systems characteristics and concepts, this approach involves large uncertainties that are difficult to quantify. As Koopman stated, "Rightly employed, it [combat simulation] gives a useful indicator in evaluations; it can *never* be relied on to predict the future."[103] JWARS was expensive, yet could not accomplish the vision of those who conceived and advocated for it.

Analytic Agenda / Support to Strategic Analysis. Given the expense and challenges of JWARS, in 2002 Secretary of Defense Donald H. Rumsfeld created an Analytic Agenda (now called Support to Strategic Analysis—SSA) to transform DoD's analysis system supporting strategic and programmatic decision making. The Analytic Agenda was a set of activities designed to do the following:

- Articulate, through scenarios, the secretary's guidance to the department about the missions, environments, and threats for which the future force should be prepared.

- Apply joint concepts to future missions depicted in planning scenarios.

- Produce standardized, accessible, transparent data and common assumptions for department-wide use in analysis.

- Design and conduct major joint analyses to support decisions on force structure, investments, and capability trade-offs.[104]

This effort did result in scenarios for analysis approved by DoD leadership, and it created conferences at which the services met to agree on common datasets they would use in their analyses. Each service was assured of having one of its preferred scenarios included. The services also used their preferred "all-purpose" campaign simulations for their capability-development processes, incorporating data beyond that in the common datasets as needed. However, few of these data came from detailed analyses of operations and exercises. These efforts have had little impact on cross-service force structure investments or capability trade-offs.

The details of studies done using these simulations are classified and proprietary, limiting opportunities for review of their objectivity and rigor. OSD, the

Joint Staff, and the services should take care not to create a similar, highly structured set of expensive, complex, proprietary war games.

No defense problem is specified well enough that an optimum can be calculated without employing subjective judgment to establish values. The large campaign simulations used for SSA result in large sets of feasible courses of action. Expecting large combat/campaign simulations or war games to resolve conflicting preferences among institutional forces within the military-industrial-congressional enterprise that drive the defense program and budget is illogical.[105] Improvements to JWARS or the SSA are incapable of providing the precise predictions for resolving complicated and complex defense issues that those who misunderstand scientific rigor expect. "As one goes up the scale of complexity, the personal qualities of the analyst shift from scientific to artistic and his[/her] model from precise to abstract. That is why asking me which model to buy is asking the wrong question. Instead, ask me which analysts and modelers to hire."[106]

Capabilities-Based Planning

DoD's acquisition system, which consumes the vast majority of the Pentagon's attention and analytical effort, focuses on major defense acquisition programs—platforms and systems that involve the commitment of billions of dollars.[107] Under Secretary Rumsfeld, DoD attempted to introduce capabilities-based planning as a means of putting the development of individual weapons systems in context. Capabilities-based planning has received rough treatment in recent reviews for being tied to the revolution in military affairs and force transformation, focusing on concepts such as net-centric warfare rather than on strategy to defeat the strategies and forces of identified potential adversaries. These critiques largely miss the mark.[108]

The usual driver for acquisition is that an aircraft, vehicle, or vessel is reaching the point where it is expensive to maintain or upgrade with new technology, and a military service proposes to replace that platform with a new one incorporating the latest generation of technology. A 1992 study of the cost growth of DoD Major Force Program categories since Secretary McNamara instituted them in 1962 demonstrated that DoD needs 7 percent growth in its budget to maintain its force structure if it continues attempting to replace each platform with the latest generation on a one-for-one basis.[109] Using the rule of 72, this means that a 4 percent growth in defense budgets results in halving the force roughly each quarter of a century.[110]

Following the 2006 Quadrennial Defense Review, DoD made an effort to institute "strategic and tactical" acquisition reform.[111] A major part of the reform involved pilot Evaluation of Alternatives on topics such as integrated air and missile defense as a basis for resource allocation, rather than conducting an Analysis

of Alternatives for each major defense acquisition program. The effort demonstrated promise, but failed when key leaders departed. Also, weapons program managers wanted to know what the study would show before providing their data for analysis, despite direction from higher authorities.

If DoD is to overcome its accelerating mismatch between limited budgets and growing challenges, it requires a new analysis paradigm and a culture focused more on national security than on protecting parochial service and program priorities by withholding knowledge and data.

Asking Essential Questions and Selecting Appropriate Methods

Adoption of a new analysis paradigm will involve some time before the paradigm becomes institutional practice within DoD, and will incur transition costs. DoD should ensure that initial efforts focus on substantive issues. In the 1950s and '60s, federally funded research centers led the way in understanding the implications of nuclear weapons for warfare and deterrence. RAND employed Bernard Brodie, Herman Kahn, Thomas C. Schelling, Albert J. Wohlstetter, and Roberta M. Wohlstetter, among many other highly talented intellects, to explore fundamental questions of war in the nuclear age, strategy and games, and many other topics. Now, federally funded research and analysis centers have become principally an extension of Pentagon staff studies. Funding for independent research on fundamental questions has been eliminated in favor of studying the issue du jour, which eliminates many fundamental distinctions between federally funded research centers and for-profit defense contractors. In addition to making better use of its Office of Net Assessment, which under the leadership of the recently retired Andrew W. Marshall (who came to OSD from RAND in the 1970s) had a long history of searching for the right questions, DoD should return to the former model and mission for federally funded research centers, having them help DoD's leadership understand the questions they should be asking and the issues they should analyze.

DoD should realize that the principal value of good analysis is in eliminating infeasible or unsuitable courses of action, and that no analyses can provide point solutions to complicated problems. Prevailing concepts and political power among those involved will determine the final trade-offs in defense policy and plans within the space of feasible and suitable solutions. Centralized processes that give too much power to one institution, such as OSD or the Joint Staff, are likely to generate more mistakes than a messier analytical competition among concepts, methods, and proposed solutions. The Secretaries of Defense must earn their pay.

That said, different subjects call for different analytical approaches. In turning to war gaming, DoD should avoid the law of the instrument.[112] To improve

rigor, the Military Operations Research Society should assist DoD in developing guidelines for analysts to align analytical techniques with the fundamental characteristics of subjects under study.

The most appropriate action from pseudoexperimentation, whether war gaming or combat/computer simulation, is exploring the validity of the findings using other techniques. Analysis campaigns involve using a variety of techniques to address important issues. Cycles of research emphasize the interaction among these techniques as progress in one investigation informs others and is in turn informed by them.

Learning from RSAS and decades of experience in defense analysis, Davis recommends analysis campaigns. "The analysis campaign should provide for breadth with a mix of models, human gaming, historical analysis, trend analysis, and collaboration with experienced operators," and should consider multiple objectives. The approach is to conduct first-cut analyses to narrow the world under consideration, then to conduct detailed analyses. "Campaign models, for example—when used with large negotiated databases for only some standard case—are poor decision aids but are excellent for integration, for understanding the many facets of a successful large operation, and for building analyst expertise that is valuable in answering specific questions quickly, often with simpler models."[113] As an example of first-cut analysis considering multiple objectives, Hughes recommends examining alternative futures.

> For example, in determining the best naval forces to influence China and our Asian allies, it is essential to remember that the same American ships and aircraft, many of which are built for 30 and even 40 years of combat life, must serve our interests whether the China-American international relationship at any given moment is one of cooperation, competition, crisis containment, or conflict at different levels of intensity. By testing our fleet's utility in each circumstance we can judge how and where risks are involved with different fleet compositions and deployment patterns. The OSD Office of Net Assessment found that looking at alternative futures by region or economic circumstance was powerful. One did not make predictions about which future was most likely to come to pass. Instead [one] looked for common forces, solutions, deployments and negotiating positions that were suited for every future.[114]

Scenario planning has proved an effective technique for resolving structural indeterminacy.[115] Davis provides a comprehensive matrix of instruments (techniques) assessed by important attributes to be considered in an analysis campaign.[116] The discussion below represents the author's appreciation of techniques essential to cycles of research.

War Gaming and Combat/Campaign Simulation. War gaming and combat/campaign simulation are complementary to each other. Both provide insight to

participants on factors governing the contingency under study and issues and data needing further study. War games are particularly valuable for helping those employing DoD's large, computer-based campaign models to understand CONOPS and the flow of campaigns.[117]

Fleet/Field Operations Analysis. Games and combat simulation should tie directly to field/fleet exercises experimenting with new concepts, using prototype systems designed to address capability enhancements, and carefully collecting data to inform important areas of ignorance and assumptions used in plans, games, and campaign simulation.

The approach and attack manual served as a basis for data collection to advance U.S. submarine force capabilities rapidly, as did the coordination-in-direct-support (CIDS) fleet exercise guide for operational data on fleet communications. The analysis based on these data demonstrated that a CIDS concept for using submarines as an outer screen for aircraft carriers was infeasible. The fleet communications data, collected in ten fleet exercises over a two-year period in the late 1970s, provided the basis for the Warfare Environment Simulator, a simulation sponsored by Naval Electronics System Command (now the Space and Naval Warfare Systems Command) focused on command and control. Unfortunately, the Warfare Environment Simulator morphed into the Naval Warfare Simulation System, losing its focus on using fleet data and on command and control, instead becoming a large-scale campaign simulation.[118]

NWC war games served as the basis for developing new operational concepts to be explored at sea, both before World War II and during the 1980s and '90s. Fleet exercises in the 1920s and '30s turned concepts for amphibious and carrier air warfare and underway replenishment of naval task forces into key capabilities for the World War II effort. Fleet exercises in the 1980s translated operational concepts developed by the CNO SSGs (at the College) into capabilities to execute the 1980s Maritime Strategy.[119] Similarly, in the 1990s, the Navy Warfare Development Command (then collocated at the College) pursued fleet experimentation through a program called Sea Trial. However, the Navy did not sustain that effort. A debate exists over whether dedicated units are required to conduct such experimentation. The submarine force since 1949, the Navy Tactical Development and Evaluation Program in the 1970s, and U.S. Pacific Command around 2000 have made experimentation a matter of routine during fleet and joint exercises. Data collected from routine rather than structured exercises better represents what would occur in unstructured combat and operations.

As part of war-gaming initiatives, OSD, the Joint Staff, and the services should reinvigorate field/fleet experimentation and embed operations analysts in deployed battalions and carrier strike groups and on higher-echelon staffs to collect

data on operations and exercises. For large programs and issues, exercise and operations analysis guides using conceptual processes would provide consistent datasets for analysis and use in pseudoexperimentation. Those educated in engineering and the hard sciences are likely to perform in the field as well as or better than those educated in operations research curricula emphasizing mathematical programming (optimization) and stochastic processes.[120]

Cyber warfare should receive particular attention, given current challenges in creating operational models. Beyond Red teams, white hats should experiment in the field with what it would take to turn unmanned systems into kamikazes attacking their host forces, for example, before making large investment decisions.

Intelligence Collection. War games also should be tied to intelligence collection and analysis. While military intelligence naturally tends to focus on possible adversary technical capabilities (e.g., range and accuracy of weapons), war games require Red teams that understand adversary planning, training, ethos, and operational concepts. Similarly, war games also suggest adversary courses of action that would create difficulty for the Blue team. Therefore, war-game findings should play into intelligence requirements to determine whether adversaries have identified and are preparing to execute such courses of action.

Campaign Analysis. Rather than using war games or large campaign models that require significant amounts of time to set up, rapid, focused analyses on the eve of war have demonstrated value in anticipating important outcomes. Shortly before each war began, Captain/Professor Wayne Hughes gave Naval Postgraduate School (NPS) students seventy-two hours to analyze the Falklands War between the United Kingdom and Argentina, the wars in Afghanistan, and the wars in Iraq. These analyses all provided results that would have been valuable to the commanders involved.[121] The key is selecting appropriate measures for quantification. Selecting appropriate analytical measures begins with developing an appreciation for the principal factors governing outcomes, and often is not done well.[122]

What useful results reasonably can be expected from war gaming and rapid campaign analysis, since accurate results cannot be expected? At NPS, Hughes teaches the students in his joint campaign analysis course that these war-gaming and campaign analyses provide the following:

- patterns of activity, both tactical and operational; the reward of new tactics to accompany new technology

- a focusing by decision makers and their staffs on the important things— those most likely to influence the outcome and achieve "victory," or whatever the intended outcome is

- synthesized information about almost anything: the traffic, the places of concealment, the beaches, the mountain passes to block, the critical roads, or the vital bridges to protect or destroy; and, perhaps most important because it is calculable, the time to arrive on scene and the logistical support necessary to sustain operations

- advice to the decision maker that is quantitative, objective, informed, specific—and incomplete

- unexpected side benefits; for example, in designing a warship one might discover that it is not a good idea to put too many eggs in one basket if the ship can be lost while performing a dangerous task[123]

Observe that predicting outcomes, or even winners by some criterion, does not appear on the list. Hughes is a great proponent of campaign analysis and its value—if one does not claim too much predictive power from it. Decisions have to be made amid uncertainty, and informed decisions are better than those based on individual experience and personal predilections alone.[124]

Simple versus Large Combat Models. Good analysis derives from understanding those few essential features of the subject under study that govern an outcome.[125] Although using models to understand essential features is valuable, attempting to predict outcomes by adding ever more detail without considering the implications for additional uncertainty is antithetical to analysis. Campaign analyses and manual war games employing simple, focused combat models and rules that are understood and subject to question by all participants can expose the factors that govern success—i.e., those on which commanders and capability developers should focus.

Barring a more exact method for quantifying the uncertainty of a combat simulation, the analyst should estimate the typical error involved in the variables used in the models, multiply that times the square root of the number of variables, and present and report the result as the range of uncertainty in the quantitative findings. Although simulations are of great value in providing insights to analysts, analysts should be appropriately humble in recommending program or policy changes solely on the basis of the outcomes of their models.

Complexity Sciences. Advances in complexity sciences raise questions regarding current combat models and present new opportunities for defense analyses. The combat models used in war gaming and campaign simulations were developed before more recent improvements in understanding chaos and complexity. Chaos involves sensitivity to initial conditions on a space of measure zero. In a space of measure zero, no matter how precise an interval, area, or n-dimensional volume around an initial state, there exist points that will result in far different future

states of the system. A pendulum hung amid three magnets—such as Clausewitz described in explaining the pulls of government (reason), the population (primordial violence), and the military (chance) in war—is such a chaotic system. Classical physics and statistics, as discussed above, presume that describing the initial state allows prediction of future states, at least with probabilities. The foundations for statistics on spaces of measure zero are not well understood. Mathematics based on continuity does not apply in chaotic and in many complex systems.

Complexity involves power laws. Power laws have a mean, but unlike Poisson or Gaussian distributions, their standard distribution is infinity.[126] The law of large numbers does not apply to power laws. Power laws apply to phenomena such as earthquakes—and to much of human behavior that involves bursts of activity.[127] Historically, a small number of pilots and submarine commanders account for the most kills. Is this a power law? If so, how do combat models account for the distribution of talent among pilots and commanders? More broadly, how many events treated statistically in combat/campaign simulation involve chaotic and complex phenomena that make Monte Carlo processes and Markov assumptions inappropriate?

Warfare is renowned for extended periods of boredom followed by bursts of intense activity during battle. The outcome of battles is determined by tens to 10^7 motivated agents performing individual functions that are more difficult to represent than molecules in a liquid or gas. Agent-based models involve agents executing rules based on the local information they have. These models are known for demonstrating emergent behavior, such as the collapse of a line of troops when adjacent soldiers retreat.[128]

Fundamental features of warfare suggest chaos and complexity sciences may be more fruitful for understanding underlying phenomena than current models.[129]

History, Cognitive and Social Sciences, and Artificial Intelligence. The cycle of research for war gaming and combat/campaign simulation also extends to studying history and developments in social science, including experimental gaming on human behavior (such as in behavioral economics) and cognitive science studying developments in understanding the brain, etc., to explore human reasoning and dynamics.

AI has had recent success in defeating human champions in games such as chess and Go, and increasingly is embedded in computers and weapons. Having people who understand AI on a team conducting analysis campaigns will add considerable value to the effort.

Reviewing Previous Results. A final area of emphasis in a cycle of research is reviewing previous results.

> Clearly war gaming and campaign simulations are a blend of an objective, scientific approach and the artistry of human designers and participants. What can be done to evaluate how well individual studies, or a series of mutually reinforcing games, simulations, results, and conclusions have aided decision makers? One thing that is rarely done is to review "old" studies and evaluate their strengths and weaknesses after the projected future scenario year has passed. It is too much to ask, perhaps, for an evaluation of the study results and conclusion and it is exceedingly difficult to evaluate any study's impact on decisions it was to have enlightened.[130]

An objective examination of the scenario, the Red and Blue forces available, and the Red and Blue force combat capabilities after the fact can consider how well the study anticipated reality.[131] Independent review of key features of the analysis will contribute to objectivity and rigor and help to identify analytical techniques appropriate to the subject matter.

The extent to which pseudoexperiments, whether war games or combat/campaign simulations, are scientific depends wholly on the character of their execution. "Electronic computers, game-theoretic models, and statistical formulas are but instruments after all; it is not *they* that produce scientific results but the investigator who uses them."[132] Neither type of simulation is inherently more scientific than the other. The principal difference is that combat/campaign simulation is analytical—reducing the problem to constituent pieces—while war gaming emphasizes synthesis—ensuring all relevant factors are considered, including how they work together.

War gaming and large-scale computer-based combat/campaign simulation differ little in their inability to predict quantitative outcomes. The scientific value of the pseudoexperiment lies in the objectivity, rigor, and usefulness of the theory the pseudoexperiment represents. This includes the motivations, tastes and beliefs, and expertise of all the participants, including the client.

War gaming has a record of anticipating factors that largely govern outcomes, thus preventing surprise. Because DoD has used combat/campaign simulation for quantitative prediction, its performance at comparing quantitative results of combat models with actual combat has been less accurate and less reliable than that of war gaming that explored the processes and nonquantitative features that would affect a campaign most. Whereas those commanding and conducting operations rarely have the motivation and skills to become deeply involved in combat/campaign modeling, they can make the time and do have the skills to participate in war gaming. Repeated war gaming can provide firsthand experiences

to limit surprise and facilitate recognitive decision making that allows rapid adaptation to emerging situations.

Using governing factors uncovered through war gaming, detailed computer models, campaign analyses, or other techniques to create simple models of the phenomena requires much more analytical skill than adding detailed models of additional processes to existing computer models. Simpler models provide greater understanding with appropriate precision than complicated computer models with large numbers of variables that give an appearance of precision but whose range of uncertainty is difficult to estimate and grows with the uncertainty of each parameter added and the square root of the number of variables.

Returning to the roots of operations research—observing, modeling operations, and collecting data in the field—is an essential aspect of a cycle of research. Work in the field yields data and knowledge that increase understanding of which concepts actually work and which do not, and provides essential data for use in computer and war-gaming simulation.

Although the discussion of questions and possibilities raised by developments in complexity sciences is incomplete, it suggests a need to reexamine combat models and to extend analytical techniques to add the rigor of appropriate techniques to combat simulation.

The Pentagon needs to overhaul its analysis paradigm if it is to meet growing security challenges with limited budgets. Overhauling the Pentagon's analysis paradigm again will require interdisciplinary teams of scientists—from both hard and social sciences, and with an appreciation for the humanities—interacting in analysis campaigns and cycles of research. Client and contractor use and abuse of need-to-know security barriers and proprietary restrictions on studies present formidable obstacles to implementing scientific standards in DoD studies.

NOTES

1. Extended discussions of war-gaming history are provided in Peter P. Perla, *The Art of Wargaming* (Annapolis, MD: Naval Institute Press, 1990), and John T. Hanley Jr., *On Wargaming: A Critique of Strategic Operational Gaming* (Ann Arbor, MI: Univ. Microfilms International, 1991). Martin L. van Creveld, *Training of Officers: From Military Professionalism to Irrelevance* (New York: Free Press, 1990), provides an excellent history of the development of professional military education, and John B. Hattendorf, B. Mitchell Simpson III, and John R. Wadleigh, *Sailors and Scholars: The Centennial History of the U.S. Naval War College* (Newport, RI: Naval War College Press, 1984), provides a comprehensive history of the U.S. Naval War College. Wayne P. Hughes Jr., ed., *Military Modeling for Decision Making,* 3rd ed. (Alexandria, VA: Military Operations Research Society, 1997), provides an extensive overview of military modeling.

2. Perla, *The Art of Wargaming,* chap. 4.

3. Hanley, *On Wargaming,* pp. 105–10.

4. In 1902, Lt. J. V. Chase wrote a paper at the Naval War College using these equations; the Navy did not declassify it until 1972. Wayne P. Hughes Jr., introduction to Bradley A. Fiske, *The Navy as a Fighting Machine,* ed. John B. Hattendorf (Annapolis, MD: Naval Institute Press, 1988).

5. They are now called federally funded research and development centers.

6. Hanley, *On Wargaming,* pp. 156–57.

7. Alain C. Enthoven and K. Wayne Smith, *How Much Is Enough? Shaping the Defense Program 1961–1969* (New York: Harper and Row, 1971).

8. Paul K. Davis, *Analysis to Inform Defense Planning despite Austerity* (Santa Monica, CA: RAND, 2014), pp. 6–7.

9. Ibid., p. 21.

10. Deputy Secretary of Defense to Secretaries of the Military Departments et al., memorandum, "Wargaming and Innovation," 9 February 2015, available at news.usni.org/.

11. In his *Analysis to Inform Defense Planning despite Austerity,* Davis has proposed analysis campaigns, and Peter Perla coined the phrase *cycle of research* to capture the need for a process for linking war gaming to fleet exercises and other forms of study and analysis.

12. Bernard O. Koopman, "A Study of the Logical Basis of Combat Simulation," *Operations Research* 18, no. 5 (September–October 1970), pp. 856–57.

13. Abraham Kaplan, *The Conduct of Inquiry: Methodology for Behavioral Science* (San Francisco, CA: Chandler, 1964), p. 5.

14. Ibid., pp. 27–28.

15. These descriptions employ critiques of war gaming and some of the tenets used in establishing the Systems Analysis Office. Davis, *Analysis to Inform Defense Planning despite Austerity,* pp. 6–7.

16. Kaplan, *The Conduct of Inquiry,* p. 35.

17. Olaf Helmer and Nicholas Rescher, "On the Epistemology of the Inexact Sciences," *Management Science* 6, no. 1 (1 October 1959), p. 27.

18. Kaplan, *The Conduct of Inquiry,* pp. 127–28.

19. Helmer and Rescher, "On the Epistemology of the Inexact Sciences," p. 25.

20. Ibid., pp. 38–52; Kaplan, *The Conduct of Inquiry,* p. 95.

21. Helmer and Rescher, "On the Epistemology of the Inexact Sciences," p. 29; Kaplan, *The Conduct of Inquiry,* p. 112.

22. Terms used by Helmer and Rescher in "On the Epistemology of the Inexact Sciences" and Kaplan in *The Conduct of Inquiry,* respectively.

23. Helmer and Rescher, "On the Epistemology of the Inexact Sciences," p. 29.

24. Ibid., p. 48. The concept of *experimentation* here is any exploration used to gain understanding of the subject under study. A large literature on experimental gaming addresses the use of gaming in more-formal experiments.

25. Leonard J. Savage, *The Foundations of Statistics,* 2nd ed. (New York: Dover, 1972), p. 9.

26. See Stafford Beer, *Decision and Control: The Meaning of Operational Research and Management Cybernetics* (London: Wiley, 1966), for an excellent discussion of this process.

27. Garry D. Brewer, *Scientific Gaming: The Development and Use of Free-Form Scenarios* (New Haven, CT: Yale School of Organization and Management, 1978).

28. Koopman, "A Study of the Logical Basis of Combat Simulation," pp. 855–56.

29. Helmer and Rescher, "On the Epistemology of the Inexact Sciences," p. 49.

30. "[E]pistemology is concerned with the role of evidence in the attainment of scientific laws and with the scientific procedures implied by that role." Ibid., p. 27.

31. Ibid., p. 43.

32. O. G. Haywood Jr., "Military Decision and Game Theory," *Journal of the Operations Research Society of America* 2, no. 4 (November 1954), pp. 365–85.

33. Koopman, "A Study of the Logical Basis of Combat Simulation," p. 857.

34. Barry D. Watts, "Doctrine, Technology, and War" (paper presented at the Air & Space Doctrinal Symposium, Maxwell Air Force Base, Montgomery, AL, 30 April 1996), available at www.airpower.au.af.mil/.

35. Koopman, "A Study of the Logical Basis of Combat Simulation," p. 858.

36. Classic methods for calculating transition rates for arrays of states and the solution of stochastic equations are not trivial. Ibid., pp. 860–66. They also employ assumptions of continuity that may not be warranted.

37. Ibid., p. 860.

38. Ibid., p. 867.

39. The following is derived from ibid., pp. 871–72.

40. A unit may be an individual element, such as a soldier, ship, or aircraft, or aggregated elements such as formations of men and women, squadrons, or task forces.

41. Koopman, "A Study of the Logical Basis of Combat Simulation," pp. 871–72.

42. William Poundstone, *Labyrinths of Reason: Paradox, Puzzles and the Frailty of Knowledge* (New York: Random House, 1988), pp. 183–88, relates an analysis by Larry J. Stockmeyer and Albert R. Meyer about such a computer.

43. Monte Carlo simulations employ many runs, randomly selecting from the probability distributions coded into stochastic processes, to produce distributions of outcomes. See Hughes, *Military Modeling,* for a comprehensive set of military modeling techniques for different applications; and Koopman, "A Study of the Logical Basis of Combat Simulation," pp. 867–71, for Monte Carlo Lanchester analysis.

44. For examples from several giants in military operations research, see Glenn A. Kent, "On Analysis," *Air University Review* 18, no. 4 (May–June 1967), pp. 50–55; Koopman, "A Study of the Logical Basis of Combat Simulation," pp. 872–79; Bernard O. Koopman, *Search and Screening: General Principles with Historical Applications* (Elmsford, NY: Pergamon, 1980); Wayne P. Hughes Jr., "A Salvo Model of Warships in Missile Combat Used to Evaluate Their Staying Power," *Naval Research Logistics* 42, no. 2 (March 1995), pp. 267–89; Wayne P. Hughes Jr., *Fleet Tactics and Coastal Combat,* 2nd ed. (Annapolis, MD: Naval Institute Press, 1999); and Davis, *Analysis to Inform Defense Planning despite Austerity,* pp. 50–55, 147–53.

45. For a brief discussion of the evolution of computers in solving games of tic-tac-toe,

chess, and Go, see "AlphaGo: Using Machine Learning to Master the Ancient Game of Go," 27 January 2016, *Official Google Blog,* googleblog.blogspot.com/.

46. See later discussion of RSAS.

47. Operations Research Society of America, "Guidelines for the Practice of Operations Research," *Operations Research* 19, no. 5 (September 1971), pp. 1123–1258.

48. Achieving a reasonable culminating point in a war game of a few days or a week requires a theory of where this point occurs and careful design of the move and adjudication structures. Robert C. Rubel [Capt., USN (Ret.)], e-mail to author, 31 March 2016.

49. Hanley, *On Wargaming,* pp. 228–72.

50. Davis, *Analysis to Inform Defense Planning despite Austerity,* p. 32.

51. Ibid., p. 66.

52. Wayne P. Hughes Jr., "Prediction" (address to the Military Application section of INFORMS, Phoenix, AZ, 14–17 October 2012).

53. Helmer and Rescher, "On the Epistemology of the Inexact Sciences," p. 49. Captain Rubel, a former director of the Naval War College War Gaming Department and dean of the Center for Naval Warfare Studies, noted that this is true only if the umpires do not overlook an omission of the players', which may occur if following the rules would require a reset of the game. Robert C. Rubel [Capt., USN (Ret.)], e-mail to author, 18 April 2016.

54. Robert Levine, Thomas Schelling, and William Jones, *Crisis Games 27 Years Later: Plus Ćest Déjà Vu* (Santa Monica, CA: RAND, 1991), p. 27.

55. Thomas A. Brown, *Potential Applications of Manual Games* (Santa Monica, CA: RAND, 1984), p. 2.

56. Operations Research Society of America, "Guidelines for the Practice of Operations Research," p. 1144.

57. Comptroller General of the United States, *Advantages and Limitations of Computer Simulation in Decisionmaking* (Washington, DC: General Accounting Office, 1973).

58. Irving L. Janis and Leon Mann, *Decision Making: A Psychological Analysis of Conflict, Choice, and Commitment* (New York: Free Press, 1977), chap. 3.

59. Irving L. Janis, *Crucial Decisions: Leadership in Policymaking and Crisis Management* (New York: Free Press, 1989), chap. 7.

60. Beer, *Decision and Control,* chap. 2.

61. Daniel Kahneman, *Thinking, Fast and Slow* (New York: Farrar, Straus, Giroux, 2011).

62. Philip E. Tetlock, *Expert Political Judgment: How Good Is It? How Can We Know?* (Princeton, NJ: Princeton Univ. Press, 2005), p. 20.

63. Jacob Marschak and Roy Radner, *Economic Theory of Teams* (New Haven, CT: Yale Univ. Press, 1972), provides a framework and analysis for team decision making based on tastes and beliefs.

64. Quoted in Stephen Downes-Martin, "Adjudication: The *Diabolus in Machina* of War Gaming," *Naval War College Review* 66, no. 3 (Summer 2013), p. 74.

65. Kaplan, *The Conduct of Inquiry,* p. 128.

66. Wayne P. Hughes Jr., e-mail to author, 1 March 2016. Also see Hughes, "Prediction," for a discussion of rapid analysis before actual combat.

67. Kaplan, *The Conduct of Inquiry,* p. 283.

68. Phillip M. Morse and George E. Kimball, *Methods of Operations Research* (New York: Wiley, 1951).

69. Comptroller General of the United States, *Advantages and Limitations of Computer Simulation in Decisionmaking.*

70. This publication has had different names over the years.

71. The author participated in the design, conduct, and reconstruction of LANTSUBASWEX 2-88 that discovered that an analyst had embedded a thumb rule (quasi law) into the Submarine Fleet Mission Program Library search calculations that provided results that were conservative by an order of magnitude.

72. For example, the coordination-in-direct-support and over-the-horizon targeting programs conducted by Naval Electronic Systems Command, PME-108, in the late 1970s.

73. Hughes, "Prediction."

74. Ibid.

75. Thomas Mahnken, "The Gulf War in Retrospect," *FP,* 20 January 2011, foreignpolicy.com/.

76. Michael A. Ottenberg, e-mail to author, 11 March 2016.

77. Koopman, "A Study of the Logical Basis of Combat Simulation," p. 880.

78. Model verification and validation is an arcane subject covered only lightly in this article.

79. Brown, *Potential Applications of Manual Games,* p. 3.

80. Koopman, "A Study of the Logical Basis of Combat Simulation," p. 867. Adam Elkus, "Strategic Theory and the Logic of Computational Modeling," *Infinity Journal* 5, no. 1 (2015), also emphasizes that the main value of such simulation derives from other-than-quantitative prediction, quoting Joshua M. Epstein, "Why Model?," *Journal of Artificial Societies and Social Simulation* 11, no. 4 (2008), p. 12, "[T]here are many other reasons to model other than to predict. Models can explain phenomena of interest, guide data collection, suggest useful analogies, cast light on core uncertainties, expose hidden assumptions, bound outcomes to plausible ranges, illuminate core aspects of interest, challenge conventional wisdom, and generally reveal what is simple to be complex," and Petter Holme and Fredrik Liljeros, "Mechanistic Models in Computational Social Science," *Frontiers in Physics,* 17 September 2015, journal.frontiersin.org/, "By constructing a computational artifact that renders a theory precise in its qualitative assumptions and performing experiments with it, researchers attain the opportunity to discover flaws and hidden assumptions that might not otherwise be clear from the verbal theory alone. This may be an interesting theoretical result in and of itself or a spur to further research. Likewise, building a computational artifact and then performing experiments with it may suggest interesting new hypotheses for future research. Even familiar situations may look very different when their core assumptions are altered and the results simulated."

81. Comptroller General of the United States, *Advantages and Limitations of Computer Simulation in Decisionmaking;* decades of observing operations research practice.

82. Kaplan, *The Conduct of Inquiry,* p. 284.

83. Herman Kahn and Irwin Mann, *War Gaming* (Santa Monica, CA: RAND, 1957), pp. 11–12.

84. Hanley, *On Wargaming,* chap. 4.

85. The author was program director of the CNO SSG from 1985 to 1998 and helped to

conceive and orchestrate this game. At the time, U.S. policy was that Iran was the adversary and we were friendly to Iraq.

86. Richard D. Duke and Jac L. A. Guerts, *Policy Games for Strategic Management: Pathways to the Unknown* (West Lafayette, IN: Purdue Univ. Press, 2004), p. 23.

87. Levine, Schelling, and Jones, *Crisis Games 27 Years Later,* pp. 23–26.

88. Andrew Marshall, e-mail to author, 3 March 2016. I thank Mr. Marshall for encouraging me to highlight these points.

89. Rudolf Hofmann, *War Games* (Washington, DC: U.S. Army Dept., 1952), provides a comprehensive account of German army gaming before and during World War II.

90. Brown, *Potential Applications of Manual Games,* p. 7.

91. Paul Bracken, "Unintended Consequences of Strategic Gaming," *Simulation and Games* 8 (September 1977), pp. 283–318, discusses the unintended consequences of gaming, which also apply to campaign simulation.

92. Brown, *Potential Applications of Manual Games,* p. 4.

93. Levine, Schelling, and Jones, *Crisis Games 27 Years Later,* pp. 31–33. Also see Robert C. Rubel, "The Epistemology of War Gaming," *Naval War College Review* 59, no. 2 (Spring 2006), pp. 108–28, for his discussion of whispers in war games.

94. U.S. Defense Dept., *Joint Operational Planning,* JP 5-0 (Washington, DC: Joint Staff, 2006), p. III-31.

95. See Karol G. Ross et al., "The Recognition-Primed Decision Model," *Military Review* (July–August 2004), pp. 6–10, on recognition-primed decision making, and Levine, Schelling, and Jones, *Crisis Games 27 Years Later,* p. 27, for the Schelling quote.

96. Paul K. Davis and James A. Winnefeld, *The RAND Strategy Assessment Center: An Overview and Interim Conclusions about Utility and Development Options* (Santa Monica, CA: RAND, 1983), p. 15.

97. Ibid., p. vi.

98. "Our intention is that analysts and other substantive experts work directly with the software tools to build, test, use, and adjust the models." Paul K. Davis, Steven C. Bankes,

and James P. Kahan, *A New Methodology for Modeling National Command Level Decision-making in War Games and Simulations* (Santa Monica, CA: RAND, 1986), p. 60.

99. Davis and Winnefeld, *The RAND Strategy Assessment Center,* p. vii.

100. Hanley, *On Wargaming,* pp. 421–53.

101. Davis, *Analysis to Inform Defense Planning despite Austerity,* p. 60.

102. George F. Stone III and Gregory A. McIntyre, "The Joint Warfare System (JWARS): A Modeling and Analysis System for the Defense Department" (presentation at the INFORMS Winter Simulation Conference, Arlington, VA, 2001), pp. 691–96.

103. Koopman, "A Study of the Logical Basis of Combat Simulation," p. 857.

104. The author participated on the senior committee proposing scenarios and overseeing the set of activities for a year during 2006–2007.

105. See Robert M. Gates, *Duty: Memoirs of a Secretary at War* (New York: Knopf, 2014), p. 117.

106. Wayne P. Hughes, "An Operations Analyst's View of the New Modeling Technologies" (paper presented at the Symposium on Command and Control and Control Research and Decision Aids, Monterey, CA, 22 June 1994), p. 3.

107. In 2014, the dollar thresholds for a major defense acquisition program were $480 million in research and development or $2.79 billion in constant 2014 dollars.

108. For a critique of capability-based planning, see Thomas E. Ricks, "The Pentagoner: The Long, Slow Death of Capability-Based Planning," *Foreign Policy,* 5 January 2015, foreign policy.com/. A detailed study is provided in John T. Hanley Jr. et al., *Improving Integration of Department of Defense Processes for Capability Development Planning* (Alexandria, VA: Institute for Defense Analyses, 2006).

109. Dov S. Zakheim and Jeffrey M. Ranney, "Matching Defense Strategies to Resources: Challenges for the Clinton Administration," *International Security* 18, no. 1 (Summer 1993), pp. 51–78. Zakheim and Ranney created a computer program that challenged players to create future force structures using historical DoD cost growth and budgets.

Using this program, the CNO SSG in 1992 told the CNO that the U.S. Navy of 2012 would be closer to 250 ships than the approximately 550 it had at the time.

110. Usually, the rule of 72 is employed to determine how long it will take to double an investment. One divides 72 by the rate of return. However, in a similar fashion, dividing 72 by the percentage difference between cost growth and budget growth gives the number of years until forces that can be funded are half of what they were.

111. The author served in OSD Acquisition, Technology, and Logistics as deputy for acquisition concepts during 2007–2008 to institute some of these reforms.

112. "I call it the *law of the instrument,* and it may be formulated as follows: Give a small boy a hammer, and he will find that everything he encounters needs pounding. It comes as no particular surprise to discover that a scientist formulates problems in a way which requires for their solution just those techniques in which he[/she] himself[/herself] is especially skilled." Kaplan, *The Conduct of Inquiry,* p. 28.

113. Davis, *Analysis to Inform Defense Planning despite Austerity,* p. 23.

114. Hughes e-mail, 1 March 2016.

115. Peter Schwartz, *The Art of the Long View: Planning for the Future in an Uncertain World* (New York: Doubleday, 1991).

116. Davis, *Analysis to Inform Defense Planning despite Austerity,* p. 24.

117. For example, the U.S. Army's Center for Army Analysis developed a Wargame Analysis Model (C-WAM) to help its analysts understand the theater backdrop, engaged forces, and CONOPS of a new theater campaign before coding into their high-resolution campaign model, JICM. Working with combatant commands and their component commands over several years both enhanced C-WAM's representation of joint capabilities and resulted in the commands using C-WAM to game their operations plans. Briefing, Military Operations Research Society Wargaming Community of Practice, 20 April 2016. (This briefing is not available yet, but it should be within a year or so via the Navy war-gaming virtual community of practice.)

118. The author wrote portions of the CIDS fleet exercise analysis guide and designed, participated in, and analyzed individual exercises and analyses using the results from all exercises. He assisted in employing the data from these exercises in the Warfare Environment Simulator.

119. John T. Hanley Jr., "Creating the 1980s Maritime Strategy and Implications for Today," *Naval War College Review* 67, no. 2 (Spring 2014), pp. 11–29.

120. Mathematical programming covers linear, nonlinear, integer, and dynamic programming and similar optimization techniques.

121. Hughes, "Prediction."

122. James G. Roche and Barry D. Watts, "Choosing Analytic Measures," *Journal of Strategic Studies* 14, no. 2 (June 1991), pp. 165–209.

123. Although this list addresses military officers, analogous statements apply to strategy/policy games involving civilian authorities and academics intending to affect higher-level decision makers. Elizabeth Bartels, e-mail to author, 1 April 2016.

124. Hughes e-mail, 1 March 2016.

125. Glenn A. Kent, *Thinking about America's Defense: An Analytical Memoir* (Santa Monica, CA: RAND, 2008).

126. A large sequence of Poisson distributions produces a normal Gaussian distribution.

127. Albert-László Barabási, "The Origin of Bursts and Heavy Tails in Human Dynamics," *Nature* 435 (May 2005), pp. 207–11.

128. For example, see Andy Ilachinski, "An Artificial Approach to War" (briefing slides, Center for Naval Analyses, Tactical Analysis Team / Operations Evaluation Group), available at www2.dcs.hull.ac.uk/.

129. Edmund L. DuBois, Wayne P. Hughes Jr., and Lawrence J. Low, *A Concise Theory of Combat* (Monterey, CA: Naval Postgraduate School Institute for Joint Warfare Analysis, 1997), provides a concise theory of combat to provide a framework for analysis.

130. Hughes e-mail, 1 March 2016.

131. Ibid.

132. Kaplan, *The Conduct of Inquiry,* p. 29.