Naval War College Review

Volume 72
Number 3 *Summer 2019*

Article 6

2019

Back to Basics—Research Design for the Operational Level of War

Margaret M. Polski

BACK TO BASICS

Research Design for the Operational Level of War

Margaret M. Polski

The current competitive environment creates new imperatives in the Navy to sharpen our research skills and expand the range of methods we use to investigate questions about warfare and the Navy enterprise. The call to innovate has given rise to a number of very thoughtful critiques and suggestions.[1] For example, there is renewed interest within the Navy in using wargaming, game theory, and experimentation to illuminate war-fighting challenges.[2] However, none of these critiques or suggestions takes into consideration the nature of the research challenge or research standards.

Sometimes the use of particular research methods is promoted without understanding why analysts choose to use one method versus another or how to use each one most effectively to achieve particular objectives.[3] As many in the war-fighting research community are quick to point out, well-intended but overzealous analytical enthusiasms often lead to inefficient investment and disappointment, which inevitably generate critiques and calls for reform. In the worst case, a misguided analysis is used to support decisions that have tragic outcomes.

Innovating and expanding the range of methods we use to understand war fighting are a good idea—provided they are informed by a clear understanding of our objectives and research-performance criteria. However, in its review of joint professional military education research institutions, the Government Accountability Office concludes

Margaret M. Polski earned a doctorate in political science at Indiana University under the supervision of Nobel laureate Elinor Ostrom. Her methodological concentrations included game theory and experimental design, and she did postdoctoral work in cognitive neuroscience and computational social science. Prior to joining the Naval War College faculty, Dr. Polski provided technical assistance to address a wide range of national security challenges at home and abroad, supporting defense, diplomacy, development, homeland security, and law-enforcement missions. She is a research fellow in the Krasnow Institute for Advanced Study at George Mason University, serves as a panel reviewer for the National Science Foundation, and coleads a NATO systems analysis study research task group on innovation in analytical wargaming.

Naval War College Review, Summer 2019, Vol. 72, No. 3

that the Department of Defense (DoD) does not have criteria to assess research performance.[4] The results of a 2017 workshop on the Navy's operations research enterprise suggest that the Navy shares this challenge.[5]

The purpose of this article is to provide a starting point for addressing the military research-performance criteria gap by examining the nature of war fighting as a research challenge in the context of professional research standards and exploring how we could better assess, select, and evaluate methods.[6] The first section analyzes the nature of war fighting as a research challenge and the implications for designing and evaluating research. The second section provides an overview of research design and the role of professional research standards in selecting methods and evaluating findings. The third and fourth sections analyze the strengths and limitations of wargaming, experimentation, and game theory against criteria that emerge from the analyses in sections 1 and 2. The final section concludes by offering suggestions for improving our research practices.

THE NATURE OF THE RESEARCH CHALLENGE

With regard to analyzing warfare, the Navy (and the military more generally) has two challenges, which often are conflated. One challenge is to understand the nature of warfare as it evolves and how to fight, which informs the full range of the Title 10 concerns of the Chief of Naval Operations (CNO). The other is to educate current and future warfighters, which primarily informs the CNO's responsibility to organize, man, and train Navy forces.[7]

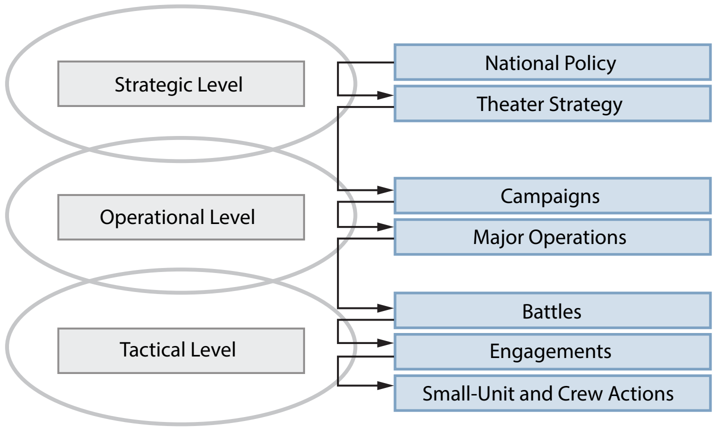
The first challenge involves ensuring that we design and conduct high-quality investigations that advance knowledge and inform everyday decision-making. The second is a pedagogical challenge that involves ensuring that our teaching methods achieve our learning objectives. If we wish to select the best methods for war-fighting research, we must be clear from the outset about the nature of the research challenge, which requires specifying the level of analysis, the nature or context of war fighting, and the questions we wish to investigate.

Specification

Joint doctrine provides a starting point for considering a broad range of military research questions at multiple levels of analysis. Figure 1 illustrates the levels of warfare. *Doctrine for the Armed Forces of the United States* (JP 1) defines *warfare* as "the mechanism, method, or modality of armed conflict against an enemy. It is the 'how' of waging war."[8] While most of the Navy's war-fighting challenges have implications at all levels of analysis, to craft a rigorous and coherent research design one must focus on a particular level of analysis. For the purpose of discussion, this article narrows its analysis to one of the Navy's three primary missions: how we can be ready to fight (and win) in a specific area of responsibility (AOR). This is a research question at the operational level of warfare.

Polski: Back to Basics—Research Design for the Operational Level of War

FIGURE 1
LEVELS OF WARFARE



Source: U.S. Defense Dept., *Doctrine for the Armed Forces of the United States,* fig. 1-2.

One way to specify the operational level of war fighting in an AOR is as a series of strategic interactions. In this specification, the term *strategic interaction* does not pertain to the strategic level of warfare but rather to a type of behavior in which participants interacting in a system attempt to anticipate each other's preferences, decisions, and actions/reactions; formulate perceptions and beliefs about the behavior of other participants, on the basis of available information or past experience; and make decisions about their own actions on the basis of their perceptions and beliefs about the other participants' actions/reactions.

War fighting is a strategic interaction because two or more units are engaged in some form of rivalry, the scope of which may include a range of strategic behavior, from cooperation to competition to full-scale conflict. The focus of a research investigation is on some aspect of the behavior and effects of war fighting: the decisions that warfighters make; the signals they produce; the processes and capabilities they employ; the effects of inputs to decision-making, such as information, beliefs, and incentives to act; the costs, benefits, risks, and effects of the choices that warfighters make; and so on.

The next step is to specify the context of operational-level war-fighting decision behavior. Many war-fighting behaviors emerge from a complex, adaptive system of physical and social systems. Complexity has a number of implications for designing and conducting research on strategic interaction.[9] In operations research, decision-making challenges associated with complexity are called "wicked challenges." *Wicked challenges* cannot be formulated definitively; they have neither a well-defined set of right-or-wrong solutions nor a well-described set of permissible operations; discrepancies can be explained in many different ways; and there is no immediate or ultimate test of a solution. While developing war-fighting capabilities involves bench science, engineering, and social science, war fighting involves behavioral challenges that cannot be solved or addressed by science or probabilistic mathematics alone. This means that it is inherently a fuzzy rather than a formal logic challenge, which requires methods and tools that allow us to deal with vagueness and classification issues in a systematic and tractable way.[10]

If we wish to investigate decision-making questions at the operational level of warfare, we should employ research methods that are useful for analyzing strategic decision behavior in complex adaptive systems. To do this, we need to specify what we mean by complexity, and its implications for research design.

Complex Adaptive Systems

Research questions that involve complicated systems, such as determining the range of a missile or a torpedo, may be complicated—but they are not wicked. They have knowable parameters and a small number of potential outcomes, and with the right set of methods one can generate estimates within a set of confidence intervals. But questions related to decision-making in *complex adaptive systems,* such as deciding whether to launch a missile or a torpedo or where to target weapons, have fuzzy parameters and a potentially large number of alternative courses of action and outcomes. We may be able to observe patterns in the behaviors that emerge from complex adaptive systems, but we cannot eliminate the possibility that these patterns are transitory or extraneous. Hence, we may not be able to generate reliable estimates.

Similarly, we may find correlations in the behavior of components of complex adaptive systems, but it may not be possible to accumulate evidence that demonstrates causal relations. For example, warfighters can and do change the nature of a rivalry by manipulating factors related to time, space, and force. Variables include the types of participants; functions and enablers, such as command and control, intelligence, fires, logistics, maneuver, position, cyber, and other assets or capabilities in their AOR; the scope of permissible action and rules of engagement; the enforcement of rules; and the imposition of costs and benefits. Warfighters make decisions or take actions on the basis of an estimate of who or what their adversaries and allies are at any given point in time and what actions those adversaries and allies have taken or will take. Hence, to understand the behavior of one warfighter, one must understand the behavior of others within the relevant war-fighting environment, which may extend beyond the immediate AOR.

Behavior that emerges from complex adaptive systems can be differentiated from behavior in complicated systems in a number of ways. First, *behavior* is self-organizing and exhibits both deliberate and spontaneous order. *Self-organizing* *behavior* occurs without explicit command or control and depends on locally operating social and technical processes, which may be quite difficult to observe.

Second, the structure of complex adaptive behavior is distributed—or, more specifically, is polycentric—and it is multiscale. We can expect to find more than one center of decision-making and control, which means that a single, hierarchical form of command and control does not govern the system, even though some subsystems may be centralized or decentralized hierarchies. Multiscale behavior cannot be described by a single rule or analyzed at a single level; the structure of behavior exists on multiple scales and multiple levels.

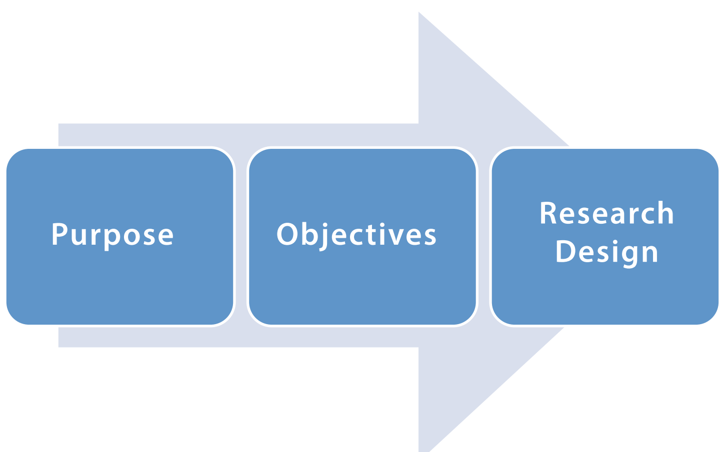
Finally, while complex adaptive systems are sensitive to initial conditions, they often exhibit contradictory behavior. We may find path dependencies and stable equilibrium; however, these systems are ultimately dynamic, and we also can expect to find punctuated equilibrium, "black swans," interdependencies, and nonlinearity.

It is extremely challenging to comprehend complex adaptive systems. If we wish to understand the behavior that emerges from these systems, we must be very careful in how we design and evaluate research.

RESEARCH DESIGN

As figure 2 shows, research design begins by specifying a purpose and objective(s) for the work. Objectives drive research design, which includes reviewing prior research, then selecting approaches, methods, and tools to conduct research. The ultimate test of research findings is the extent to which they meet research objectives, contribute to producing knowledge that is supported by facts and logical reasoning, or clarify the limits of knowledge.

FIGURE 2
THE FOUNDATIONS OF WAR-FIGHTING RESEARCH DESIGN



The objective of operational-level war-fighting research may be to create or contribute to a body of basic research or applied research. For example, a basic-research project on operational-level war fighting could involve trying to develop or test theories about enduring questions, such as why adversaries fight, how certain types of conflict end, or which signaling behaviors are associated with particular types of operations. Findings could contribute to advancing general knowledge about war fighting by providing evidence that either supports or does not support theories about the challenge.

By contrast, an applied-research project could involve identifying the functions and enablers that are required to support an operational concept in a particular AOR under a specific set of conditions. It may or may not test theories about an operational challenge, but it could inform operational planning. The findings could help senior leaders and planners better understand the implications of using the operational concept that is being investigated in a specific context at a particular moment in time.

Once the objectives of research are determined, the principles that guide research design and execution are quite similar. The overarching goal of research is to produce findings that have inferential value for other researchers, educators, and practitioners. However, this can pose a challenge for research design. Professional research communities have expertise in research design. They impose stringent requirements for inferential value because society entrusts them to both build the stock of knowledge and guard the integrity of this stock. In contrast, practitioners might not have expertise in research design and they might accept less-stringent requirements for inferential value because they are not charged with building or guarding the stock of knowledge. Researchers must take these differences into consideration when they design research and share it with others.

War-fighting challenges cross disciplines, and research-design practices can vary across disciplines and research groups. For example, understanding the effects of hitting a target with a missile involves a grasp of knowledge that researchers accumulate in physics, chemistry, material sciences, engineering, social sciences such as psychology and political economy, and so on. However, professional researchers, regardless of their disciplines, are trained to think about their work in the context of normal science and to use some version of the scientific method in the way they design and conduct their research.[11] Table 1 lists the generally accepted steps in research design.

The aim of research design is to produce findings that will contribute to a stock of accumulated knowledge. This may involve testing an existing theory about a research challenge or developing a new theory. A peer-review process determines whether research findings will be added to the official stock. While peer reviews are notoriously idiosyncratic, the research communities do have

TABLE 1
STEPS IN PROFESSIONAL RESEARCH DESIGN

Step	Activity	Issues/Questions
1	Specify an interesting purpose and objectives	What do we need to address to advance the stock of knowledge, such that it has broader impact?
2	Survey the scholarly and empirical research literature	What do we know about the issues associated with our purpose and objectives, and how strong is the evidence? That is, what are existing theories (i.e., reasoned and logically consistent speculations about the answers to research questions) about our research challenge? What is the evidence that supports these theories? How extensive is the empirical support? How rigorous is the support?
3	Formulate research questions as hypotheses and develop a detailed plan to investigate or test the degree of support for the hypotheses	Which data can we realistically collect to test theory?[a] How will we collect, structure, record, and archive the data we collect? How will we analyze data? Analytical questions focus on evaluating alternative testing methods to identify the one that best fits our research objectives and the type of data we are able to collect.[b]
4	Conduct the research	How can we use data and conduct analyses to generate inferences that are unbiased by error? How can we increase leverage while minimizing the information used for description or inference?
5	Disseminate research results	Research conducted to advance the stock of public knowledge is submitted for peer review and published as widely as possible in scholarly and other public media. Research conducted to advance a private stock of knowledge also should be peer-reviewed and disseminated as widely as possible within the constraints of the research tasking.

Notes:
a. Data may be quantitative (amounts or quantities, in the form of numbers or proportions) or qualitative (in the form of characteristics or qualities).
b. Mathematical, statistical, and machine-learning methods, which typically are called *quantitative methods,* require quantitative data. By contrast, *qualitative methods,* such as histories, case studies, and interviews, rely on qualitative data.

standard operating practices, which aim to assess and enforce *rigor*. Measures of rigor relate to *validity,* or the extent to which the way research is designed and carried out generates findings that actually measure what the researcher intended to measure; *reliability,* which refers to whether the researcher's measuring procedure, used in the same way, will produce the same measure; and *replicability,* or the extent to which another equally capable researcher could duplicate an analysis using the same data and reach the same conclusions.[12]

Other professional research standards include taking steps to minimize bias and error in research procedures and ensuring, to the extent possible, that research procedures are *parsimonious* (explaining as simply as possible, with as little extraneous detail as possible). Finally, researchers are expected to be humble and to share their knowledge. All knowledge and all inference have limits; the best researchers are collegial skeptics who are able and willing to ask about the relevance of data, the appropriateness of research and analytical procedures, and the possibility of alternative explanations for inferences. They disseminate their research findings widely, in accessible ways, to stimulate and accelerate knowledge production.

METHODS

Research design includes surveying research methods and selecting the one best suited to achieving research objectives. Recently the Navy research community has been reconsidering three methods to investigate questions about decisions and interactions in war fighting: wargaming, experimentation, and game theory. This section provides an overview of these methods, and the following section evaluates them against the criteria discussed in the preceding sections.

Wargaming

The Navy has employed wargaming as a method for investigating war fighting since at least 1887.[13] The Navy's use of wargaming has been shaped indelibly by the work of Captain William McCarty Little, Captain Wilbur R. Van Auken, and Francis McHugh. These early researchers developed wargaming at the Naval War College over a period that spanned the founding of the College, the interwar period, and the Cold War era. McCarty Little is credited with introducing wargaming into the Naval War College in 1887, and he wrote a number of papers on the subject.[14] Van Auken was hired to stand up a research department at the College in 1932 to document and analyze wargame findings.[15] However, it is McHugh who is associated most closely with developing the disciplined and systematic methodological approach to using wargames to analyze war fighting that the College's War Gaming Department (NWC WGD) uses today.[16]

McHugh, incorporating McCarty Little's work, defined *wargaming* as follows: "A war game is a simulation, in accordance with pre-determined rules, data, and procedures, of selected aspects of a conflict situation. It is an artificial—or more strictly, a theoretical—conflict . . . to afford a practice field for the acquirement of skill and experience in the conduct or direction of war, and an experimental and trial ground for the testing of strategic and tactical plans."[17]

Today, the NWC WGD conducts wargames for the senior leadership of the Navy, primarily at the operational level of war in the context of great-power competition. Wargaming is a systematic method for experiencing the effects of war-fighting decisions and analyzing decision behavior. A wargame is a representation of a war-fighting decision-making dilemma—a representation that may or may not conform to what we can or will observe in the naturally occurring world. It is a representation of a real war fight, in the sense that players are actually playing the game, and hence engaging in decision-making related to war fighting. Disciplined wargamers begin with a decision challenge, specify their research objective(s), and then design an experience that will illuminate the decision in such a way that they will achieve these objectives.[18]

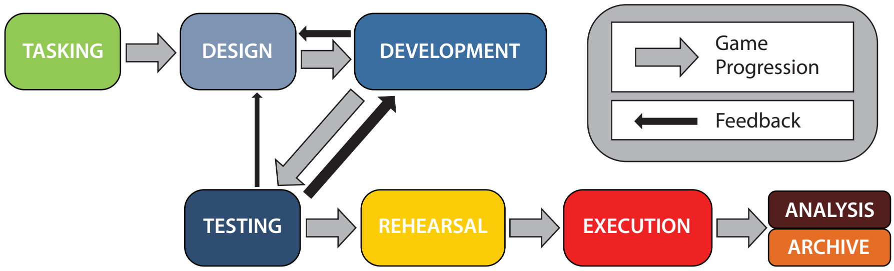
In a typical wargame, players (individuals or groups with experience or responsibilities that are relevant to the decision challenge) are recruited to play the game in teams. Individuals who are not involved directly in playing the game act as facilitators, adjudicators, data collectors, and analysts. Individuals are selected to participate on the basis of their expertise and understanding of the decision context. For example, if a researcher or research sponsor wishes to investigate the strengths and limitations of a concept of operations (CONOPS), the validity and reliability of the findings from game play depend to a great extent on how knowledgeable the participants are about the operating environment and the capabilities that the CONOPS requires them to employ. In other words, experienced warfighters should be asked to fight using the CONOPS and all their acquired knowledge and skill. Typically, none of the participants in a game is selected or assigned randomly.

Players and other participants are provided with *scene setters,* an environment in an operating area, and *scenarios,* which create a decision-making context and provide them with rules of engagement, resources, capabilities, and limitations. Games may be designed to represent conflict between two opposing units (two-sided) or among multiple teams, which may be allied or opposed (multisided).[19] The main point is that the game pits players against a rival or adversary. Participants may use computers in their play; manually manipulate assets and forces on a board; or deliberate, record, and convey their decisions in a seminar or workshop-style format.

Another group of participants controls the play of the game and adjudicates the effects of the decisions that players make. At the end of the game, players discuss the game experience, including the logic they used in playing the game, the challenges they encountered, and so on. Game play—including facilitation and after-action discussion—is observed and recorded by a data collection and analysis team. Following the conclusion of the game, the team organizes game-play data; conducts analysis against a predetermined set of research questions, using qualitative and quantitative methods; reports on findings; and archives all game material.

The NWC WGD has a professional wargame research-design process that includes the steps that are typical in professional research design, which are modified to accommodate the Navy's requirements for wargaming.[20] The WGD's current wargame-research process, which is depicted in figure 3, includes specifying the challenge, purpose, and objectives of wargame research; conducting a literature review and articulating research questions; developing a research design that includes a data-collection and -analysis plan; designing and testing a game to achieve research objectives and address research questions; developing and testing the game; executing the game; conducting analysis; writing a report, which is peer-reviewed and disseminated to an approved audience; and archiving game artifacts. This disciplined process makes it possible to replicate the game, repeat the game, or iterate on some aspect of the game.[21]

FIGURE 3
NAVAL WAR COLLEGE WARGAME RESEARCH-DESIGN PROCESS



Source: Burns, *War Gamers' Handbook*.

The fact that the NWC WGD has a professional wargame research-design process does not guarantee the quality or inferential value of its research products. As in any research enterprise, the research process may not be implemented perfectly at all times, researchers in the department may not always have the skills or tools they need at the time they need them to conduct rigorous research for particular challenges, and those who direct research tasking may wish to sacrifice analytical rigor to achieve other objectives.

NWC WGD wargaming research is conducted at classified levels and added to stocks of knowledge that the Navy and DoD maintain. The department's wargames are peer-reviewed by other wargamers and decision analysts who are engaged in developing and evaluating war-fighting concepts and plans. Wargaming is used in civilian research and learning and reported in publicly accessible scholarly research on conflict and decision behavior. There are many trade press publications on wargaming; however, few address it in the context of professional research criteria.

Experimentation

Most of us intuitively search for associations and causal relationships to explain and improve on our experience, and often we engage in a process of trial and error to arrive at a useful solution to a challenge. However, trial and error is a costly approach. The Navy has employed experimentation to identify a range of useful solutions to war-fighting challenges at the operational level of warfare.[22] Formal experiments test the influence of one or a small number of causes of observed effects.[23] Researchers articulate a specific theory of cause and effect, take an action consistent with their theory (a *treatment*), attempt to control for extraneous influences that could limit or bias the hypothetical causal relationship, and systematically observe and record the effects of the manipulation.

Some analysts associate experimentation with the physical sciences and a tightly controlled laboratory environment. However, the experimental method is used in every discipline, and a number of different types of experiments have been developed over time to address variation in research challenges and analytical settings.[24] Experiments and experimental frameworks are a well-described component of military operations research. Richard A. Kass, the Technical Cooperation Program (TTCP), and the Navy Warfare Development Center (NWDC) have published detailed, practical guides on war-fighting experimentation.[25] All three of these guides categorize analytical wargaming as a type of experimental method. Experiments also are used in the social and computational sciences, which are the scholarly disciplines that address basic and applied decision-making challenges. Moreover, both wargaming and game-theoretic approaches can be used in combination with experimentation.[26]

All *experiments* consist of four elements: (1) units or persons, (2) treatments, (3) observations or outcomes, and (4) settings. Experiments are designed to understand a theory or hypothesis about a causal relationship among these elements. They answer the question: If we do *x* to *y* in a particular setting *s,* will *z* occur? Experimental design involves specifying and operationalizing proxies for *x, y, s,* and *z* and developing a set of protocols for organizing and running the experiment that will reduce bias and minimize threats to validity. *Experimental protocols* address the assignment of units or persons to experimental conditions, measurement of observations or outcomes, comparison groups, and treatment.

Kass, TTCP, and NWDC identify four requirements for a war-fighting experiment to have inferential value.[27] First, the experiment must provide the ability to use the hypothetical concept or capability. Second, the experimental environment must be structured in such a way that the experimentalist can observe an effect from using the hypothetical element. Third, the experimental environment and procedures must permit the experimentalist to isolate the reason for the observed effect. And finally, the experimental findings must relate to a real war-fighting challenge.

Experimentation is an appealing approach to investigating war-fighting challenges because it has the potential to provide evidence about causal variables, which can be used to build knowledge about how to fight. The weakness of the approach is the extent to which causal inferences can be generalized beyond the experimental conditions to explain a broader class of similar war-fighting challenges. Generalization challenges in war-fighting experimentation are related to validity and they are intrinsic to the practical limits on designing and implementing war-fighting experiments.

For example, most war-fighting challenges are necessarily local and particular. They occur in a restricted range of settings, which may or may not be replicable or repeatable, and with a particular version of one type of war-fighting "treatment" rather than all possible versions. Usually they have several different effectiveness measures—each with theoretical assumptions that are different from those associated with other measures—but not a complete set of all possible measures. Moreover, warfare often is conducted with a convenient sample of warfighters rather than one that reflects a well-described population, and it is conducted at a particular moment in time that soon becomes history. Each of these aspects of war fighting poses a challenge for using the findings from experimentation to answer general classes of war-fighting questions or challenges.

All research methods in the decision sciences and operations research have generalization issues. Yet we are not likely to abandon our quest to understand how things work in complex field settings or how we can improve the inferential value of our research findings. The mitigation strategies we use to address threats to validity depend on the type of validity challenge we confront. Kass, TTCP, and NWDC enumerate twenty-one threats to validity in war-fighting experiments and provide a catalog of techniques and procedures for improving validity.

Game Theory

Game theory is a mathematical approach to developing and testing theories of decision challenges that involve conflicts of interest. While it has been used to analyze strategic interactions that may occur in a war-fighting context, it is not per se a method for investigating war fighting or any other strategic interaction at an operational level. Martin J. Osborne and Ariel Rubinstein provide the following definition of game theory, games, and game-theoretic solutions. "Game theory is a bag of analytical tools designed to help us understand the phenomena that we observe when decision-makers interact. . . . A game is a description of strategic interaction that includes the constraints on the actions that the players can take and the players' interests, but does not specify the actions that the players do take. A solution is a systematic description of the outcomes that may emerge in a family of games."[28]

One way to think about a game-theoretic analysis is as a systematic thought experiment that follows formal mathematical rules. It is an abstract representation of a particular interaction or class of interactions, which may or may not conform to real life. Decision makers do not play a game-theoretic analysis; a game theorist designs, "plays," analyzes, and reports on the game. The game environment is an idealized choice context, "players" are mathematical operations, and the game does not provide the opportunity to engage in or experience the effects of decision-making.

The research-design process for a game theorist is focused principally on specifying an interesting decision challenge and assumptions about decision-making, finding or developing an appropriate solution concept, and constructing a mathematical proof of the concept. The esoteric nature of the method dictates the research objective; it is an analytical exercise that is abstracted from a physical and social operating environment. While game theorists often motivate or impel their models with a stylized description of a specific operational challenge, they aim deliberately for a sparse context so they can identify strategies that mathematically dominate other strategies for the interaction of interest. However, the product of the research—namely, the game-theoretic model—may be, and often is, tested empirically in richer environments using other research methods, such as experimentation and agent-based simulations. Empirical testing can offer the opportunity for greater specificity, as well as provide experience with the decision challenge.[29]

Game-theoretic assumptions about decision-making address the following aspects of the incentives and motivations of hypothetical decision makers: decisions and the outcomes of decisions (e.g., the payoffs or the costs/benefits associated with decisions); decision makers' knowledge about alternative choices; their preferences for one decision over another, and the consistency of these preferences; their knowledge about the decisions that other decision makers will make, both preemptively and in response to their own choices; their beliefs or expectations about the likelihood of obtaining the payoffs associated with choices; the rules they use to make a decision; and the bases on which they update their knowledge about the state of play.

Game-theoretic analyses are based on a "rational" model of decision-making, in the sense that the analysis of decision makers' incentives and motivations is based on the fundamental assumption that they are aware of their alternatives, form expectations about unknowns, have clear preferences, and make choices using some type of optimization process. Games can be designed to make the decision challenge more interesting by introducing uncertainty and incomplete or imperfect information. For example, a model may assume that players are uncertain about the objective parameters of the environment, imperfectly informed about the events that unfold in a game, uncertain about the actions of other players, or uncertain about the reasoning of other players. However, rationality assumptions are maintained in the face of uncertainty and are resolved by assuming that players determine the value of a choice on the basis of an estimated value of a utility function with respect to a probability measure.

COMPARING METHODS

Wargaming, experimentation, and game-theoretic modeling—all are used to analyze strategic decisions and interactions. However, they are very different methods and have different strengths and weaknesses for analyzing war-fighting decision challenges. Analysts using these methods to investigate war-fighting decisions approach research design in ways that have profoundly different implications for achieving the Navy's research and education objectives. Table 2, which

TABLE 2
COMPARISON BASED ON MEASURES OF RIGOR

Method	Internal Validity	External Validity	Reliability	Replicability
Wargaming	High	Medium	High	High
Experimentation	High	Low	High	High
Game theory	High	Low	High	High

provides a side-by-side comparison of wargaming, experimentation, and game theory against the standard measures of rigor discussed previously, shows that no method provides perfect rigor, and it highlights the trade-offs associated with using one method versus another.[30] Game-theoretic modeling and experimentation have high internal validity, reliability, and replicability, but low external validity. This makes these methods good candidates for conducting rigorous basic research on the effects of war-fighting decisions under tightly constrained conditions, but comparatively poor candidates for addressing research questions associated with less-constrained conditions, such as those associated with complex adaptive systems or the intensity and fog of war. In contrast, wargaming has high internal validity, reliability, and replicability, and its findings are more likely to be useful for addressing relatively unconstrained war-fighting decision questions.

The estimates of rigor associated with each method listed in table 2 are based on the potential of the method to generate findings in a rigorous way if research is designed and executed properly. However, the potential of a method to generate rigorous findings does not guarantee the quality or usefulness of the research; even the best researchers are not always able to meet standards for rigor, because they may not have the data, tools, facilities, or funds they need at the time they need them, or they may choose to sacrifice some analytical rigor to achieve other objectives.

Another way to look at methodological selection is through the lens of the purpose and objectives of research. Research purposes and objectives may include any of the following:

- Advancing general knowledge about decision-making in war fighting
- Recognizing decision points in war fighting, so as to fight more effectively
- Providing warfighters with the opportunity to rehearse decision-making in a fight (e.g., to prepare themselves to fight, without actually fighting)
- Educating current and future warfighters by providing experience in the art and science of decision-making in war fighting

When we consider the nature of many war-fighting research questions and potential purposes and objectives, it suggests that we need a method that will accommodate both fuzzy and formal logic. As with any requirements analysis, if we are going to select a method properly to fit the nature of the war-fighting challenge and meet our purposes and objectives, we need a commonly understood criterion. In research requirements analysis, this criterion is *inferential value,* or the extent to which one can infer implications from research. However, inferential value is really a shorthand way of referring to two standard methodological selection criteria: the extent to which the selected method will generate analytically rigorous findings (i.e., findings that are valid, reliable, and replicable) and the feasibility of operationalizing and executing the method.

Table 3, which compares our selected decision-analytical methods on the basis of four distinct purposes and objectives, shows that, once again, none of the methods meets all our potential requirements, although wargaming and experimentation meet most of them. Wargaming provides warfighters with an opportunity to think deeply about the use of an operating concept, an operating plan, or a course of action, or to experience decision-making and the effects of interaction with a determined adversary, whereas experimentation provides researchers a means to analyze alternative concepts and capabilities empirically. However, wargaming and experimentation are not the best candidates for research that aims to advance general knowledge of warfare. Yet while game-theoretic modeling is a strong method for obtaining rigorous insights into general classes of tightly constrained decision questions, it has limited use for addressing the kinds of decisions and decision environments that are typical in joint warfare and combined arms. As Thomas C. Schelling observed when the military reinvigorated mathematical modeling and simulation to address war-fighting decisions in the Cold War era, when humans are in the loop it is impossible to sustain rigid parameters; we can and do change the rules of the game as we play and, more importantly, as we fight.[31]

TABLE 3
COMPARISON BASED ON PURPOSES AND OBJECTIVES

Objectives	Wargaming	Experimentation	Game Theory
Advance general knowledge of warfare	Weak	Weak	Strong
Recognize decision points	Strong	Strong	Weak
Rehearse decisions	Strong	Strong	Weak
Experience decisions	Strong	Strong	Weak

Researchers understand that no method ever will allow them to achieve perfect inference or analytical rigor—these are ideal standards, not minimum thresholds. They also understand that their most interesting challenges, such as fighting peer and near-peer adversaries in a global political and economic system with four physical domains, are messy and difficult to investigate. This implies that we must reconcile ourselves to trading some degree of inferential value and rigor to make progress. For example, it may be difficult to find data that will measure what we would like to measure, data may be difficult to clean for analysis because it is incomplete or corrupted, it may be difficult to mitigate bias in data collection and analysis, we may not understand or be able to agree on how to measure or interpret data, and we may not have the analytical capabilities or tools we would like to have.

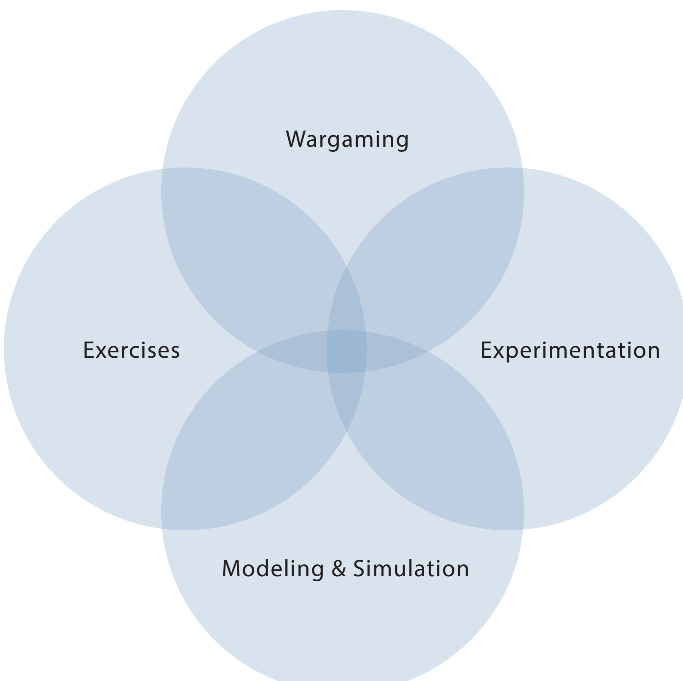
Wargaming, experimentation, and game theory have different but potentially complementary strengths for investigating questions about war fighting at the operational level. Wargaming is a research process that provides the opportunity to experience and think through the implications of operational decisions and to identify gaps and shortfalls in potential war-fighting operations. Similarly, experimentation is a disciplined and systematic investigatory process for isolating and identifying associations and causal relationships among a range of variables. When used to investigate war fighting, it may produce a range of useful findings with a relatively high degree of inferential value. Game theory can serve both wargaming and experimentation by providing a framework for specifying the structure of decision makers' incentives and motivations in a potential war-fighting operation, which is useful in designing, implementing, and interpreting research findings.

However, when used on a stand-alone basis, wargaming, experimentation, and game theory all have serious limitations as research methods for improving the useful stock of knowledge about war fighting at the operational level. Each method may produce biased results that can affect the validity and reliability of findings. The technique most often recommended to strengthen the validity of a research design is to use formal probability sampling of instances of units, treatments, observations, or settings. However, implementing this approach requires both the existence of clearly delineated populations of each of these variables and the ability to sample with known probability from within each of these populations. Even if this were possible in critical war-fighting research areas, formal sampling methods usually offer only a limited solution to generalization issues.

TTCP suggests an alternative approach to improving the inferential value of war-fighting research. It recommends using multiple research methods, integrating methods into a coherent and concerted research program, and using an iterative process of analysis.[32] The Navy has an ongoing experiment that applies this approach to exploring critical war-fighting decision issues.[33] Beginning in late 2014, the Navy began experimenting with integrating a number of different research methods to explore the challenges associated with potential war-fighting scenarios in the Pacific Fleet AOR. Figure 4 illustrates the Navy's perspective on integrated research. The aim of the project (still ongoing, on a classified basis) is to develop concrete recommendations that can inform decisions about strategy, concepts of operations, mission analyses, operation planning, campaign analyses, and resource allocation. The sponsor for the project is the fleet commander. The analytical team includes the CNO's strategy and planning staff (N3/N5), his campaign analysis staff (N81), the fleet's war-fighting assessments and readiness staff (N9), and the NWC WGD.

Naval analysts are using the insights they have obtained from this research program to inform ongoing modeling, simulation, development of CONOPSs, mission analyses, and strategic thinking; to refine analytical agendas; and to conduct further research. The team that originally organized and directed this program of research characterized it as an innovation in decision analysis that has required cultural and process changes. They believe they have demonstrated the

FIGURE 4
AN INTEGRATED PROGRAM FOR OPERATIONAL WAR-FIGHTING RESEARCH



Source: Rear Adm. Patrick Piercey, USN, and David Yoshihara.

value of linking and more tightly integrating operations analysis and research and of basing operational and command decisions on findings that rely on multiple research methods.

The team has offered the following recommendations, which have implications for designing and conducting research, education, and training.[34]

1. *Begin with the end(s) in mind.* Use a structured, systems approach to identify opportunities to achieve actionable results, specify the context for decision-making, generate objectives, and outline requirements. If integrative analysis is something new in an organization, create a Skunk Works, provide top cover, and communicate strategically.
2. *Keep it simple.* Research is time-consuming and costly; resist the tendency to overspecify requirements, to focus on tools and technologies rather than decisions and insights, and to include too many people. Do not be afraid to slow things down to avoid "rushing off to failure."
3. *Facilitate collaboration*. Build the infrastructure for collaboration into the research design and integration process, and ensure that there is openness and transparency among all associated staff members.
4. *Understand what is required and delivered.* The commander and analytical staff need to have a good understanding of research methodologies and of what is reasonably achievable. Manage expectations about the inferential value of the research and how component projects will align within the larger research program.

Professional research communities are rife with—and often thrive on—controversy and professional jealousies. However, if senior Navy leaders wish to optimize returns on investment in operations research, it would behoove them to focus on the inferential value of research findings, be wary of provoking unproductive turf wars, and eschew searches for silver bullets and all-encompassing methods. Researching complex system-of-system challenges such as war fighting requires multiple methods, integrated research programs, and strict accountability to research criteria related to inference, intellectual merit, and broader impacts.

There is no single operations research method that is demonstrably better than any other method for advancing knowledge of operational-level warfare. Every method has strengths and limitations in producing findings that can help the Navy make decisions that will protect and successfully prosecute U.S. interests over the short, medium, and long terms.

To grapple more effectively with complexity and rapidly changing threat environments, we need wargames and we need more, not less, senior leader engagement in these games. But these games must be designed and conducted with an experimental mindset, integrated within a broader research program, and informed by professional research practice, and researchers must be willing to take on board advances in science, mathematics, and technologies. We can use game theory to inform and refine operations research practices, but we cannot use it to analyze contemporary or future policy, strategy, operational design, or tactical challenges.

The history of the U.S. Navy is replete with examples of innovation in warfare at the operational level. If senior leaders wish to innovate, they must focus their attention and efforts on ensuring that decision-making in the Navy enterprise is based on professionally vetted research, then hold the research enterprise to these standards.

NOTES

1. The call to innovate in the Department of Defense began with Deputy Secretary Robert O. Work's memorandum dated February 9, 2015, available at news.usni.org/. In the Navy, Secretary Ray Mabus issued a call to innovate with his memorandum dated May 5, 2015, available at www.doncio.navy.mil/. For reflections on the implications for research at the Naval War College, see Thomas J. Culora, "A War-Gaming Renaissance," U.S. Naval Institute *Proceedings* 142/5/1,359 (May 2016). Similarly, see John Hanley's call for change in analysis and research in John T. Hanley Jr., "Changing DoD's Analysis Paradigm: The Science of War Gaming and Combat/Campaign Simulation," *Naval War College Review* 70, no. 1 (Winter 2017), pp. 64–103. Chief of Naval Operations Admiral John Richardson has emphasized the role of analysis and research. John Richardson [Adm. USN], "A Design for Maintaining Maritime Superiority Version 2.0," December 2018, available at www.navy.mil/.
2. Hanley proposes that we use game theory to enhance operational wargaming practice. John T. Hanley Jr., "Planning for the Kamikazes: Toward a Theory and Practice of Repeated Operational Games," *Naval War College Review* 70, no. 2 (Spring 2017), pp. 29–48.
3. With respect to wargaming, see, for example, Hank J. Brightman and Melissa K. Dewey, "Trends in Modern War Gaming: The Art of Conversation," *Naval War College Review* 67, no. 1 (Winter 2014), pp. 17–30, and Culora, "A War-Gaming Renaissance." With respect to game theory, see Thomas C. Schelling, *The Strategy of Conflict* (Cambridge, MA: Harvard Univ. Press, 1960). For a more recent critique, see, for example, Ariel Rubinstein, "Comments on Economic Models, Economics and Economists: Remarks on *Economics Rules* by Dani Rodrik," *Journal of Economic Literature* 55, no. 1 (March 2017), pp. 162–72.
4. Government Accountability Office, *Joint Professional Military Education: Opportunities Exist for Greater Oversight and Coordination of Associated Research Institutions,* GAO-14-216 (Washington, DC: March 2014).
5. See Jon Scott Logel and Margaret M. Polski, *2017 Navy Operations Analysis and Research Workshop Final Report* (Newport, RI: U.S. Naval War College, December 29, 2017).
6. Substantial portions of this article are based on Margaret M. Polski, "A Warfighter's Guide to Analysis" (working paper WGD20181, War Gaming Department, U.S. Naval War College, Newport, RI, April 2018), and Margaret M. Polski and Jon Scott Logel, "Doing Analysis" (working paper WGD20191, War Gaming Department, U.S. Naval War College, Newport, RI, January 2019).

7. The Chief of Naval Operations and the other service chiefs are charged under Title 10 of the U.S. Code with providing policy advice; organizing, manning, training, and equipping the force; providing guidance for the development of doctrine, concepts, courses of action, and standard operating policies and procedures; budgeting, research and development, and capital investment; and building strategic partnerships.

8. U.S. Defense Dept., *Doctrine for the Armed Forces of the United States,* JP 1 (Washington, DC: Joint Chiefs of Staff, March 25, 2013, incorporating change 1 of July 12, 2017), p. I.A(3).

9. The following sections are drawn from Margaret M. Polski, "Extending the Institutional Analysis and Development Framework to Policy Analysis and Design," chap. 2 in *Institutional Diversity in Self-governing Societies,* ed. Filippo Sabetti and Dario Castiglione (Lanham, MD: Lexington Books, 2017), pp. 25–47.

10. *Fuzzy logic* is a means of classifying objects or concepts that are inherently vague or difficult to categorize as either true or false. The modern fuzzy logic concept and fuzzy logic set theory are attributed to Lotfi A. Zahdeh. For an accessible account of fuzzy logic, see Daniel McNeill and Paul Freiberger, *Fuzzy Logic: The Revolutionary Computer Technology That Is Changing Our World* (New York: Simon & Schuster, 1993).

11. Classic texts on the philosophy of science include Karl R. Popper, *The Logic of Scientific Discovery,* Routledge Classics (Abingdon, U.K.: Routledge, 2002), and Thomas S. Kuhn, *The Structure of Scientific Revolutions* (Chicago: Univ. of Chicago Press, 1962). For an interesting analysis from a sociological perspective, see Randall Collins, *The Sociology of Philosophies: A Global Theory of Intellectual Change* (Cambridge, MA: Harvard Univ. Press, 1998).

12. These definitions of *validity, reliability,* and *replicability* are drawn from Gary King, Robert O. Keohane, and Sidney Verba, *Designing Social Inquiry: Scientific Inference in Qualitative Research* (Princeton, NJ: Princeton Univ. Press, 1994). Donald Campbell and Julian Stanley, who are concerned with experimental design, distinguish *internal validity* from *external validity*. Findings have internal validity if we can state that the experimental treatment made a difference in a specific experimental instance. External validity refers to the extent to which findings can be generalized across other populations, settings, treatment variables, and measurement variables. Donald T. Campbell and Julian C. Stanley, *Experimental and Quasi-Experimental Designs for Research* (Chicago: Rand McNally, 1963). For a general primer on research design and planning, see Paul D. Leedy and Jeanne Ellis Ormrod, *Practical Research: Planning and Design,* 9th ed. (Harlow, Essex, U.K.: Pearson, 2010).

13. The proper spelling of *wargaming* is controversial. In McCarty Little's and McHugh's published writings, the term was rendered as *war gaming* (i.e., open). Today, however, other than in the name of the War Gaming Department, the College uses the *wargame/-er/-ing* spelling (i.e., closed up).

14. McCarty Little was appointed in 1887 as a member of the faculty and developed two-sided wargaming at the College. Brightman and Dewey, "Trends in Modern War Gaming." McHugh also credits McCarty Little with introducing wargaming at the College. See Francis J. McHugh, *Fundamentals of War Gaming,* 3rd ed. (Newport, RI: U.S. Naval War College, 1966). McCarty Little's publications on wargaming include *Rules for the Conduct of War Games* (Newport, RI: Naval War College, 1901, 1905), "The Strategic Naval War Game or Chart Maneuver," U.S. Naval Institute *Proceedings* 38/4/144 (1912), and *The Chart Maneuver* (Newport, RI: Naval War College, 1920).

15. Jon Scott Logel reports on Van Auken's appointment as director of the Research Department at the College and his team's work analyzing wargames in the interwar period. See Jon Scott Logel, "Captain Van Auken and the Research Department of the Naval War College: Considerations of Analytical War Gaming in the Decade before Midway" (paper prepared for the 2017 McMullen Naval History Symposium, U.S. Naval Academy, Annapolis, MD, September 14–15, 2017). Also see John M. Lillard, *Playing War: Wargaming and U.S. Navy Preparations for World War II* (Lincoln: Univ. of Nebraska Press, 2016), p. 100.

16. McHugh's influence is reflected in the current edition of the College's reference work on the subject. Shawn Burns, ed., *War Gamers' Handbook: A Guide for Professional War Gamers* (Newport, RI: U.S. Naval War College, [2015]), available at usnwc.edu/. For a professional wargamer's perspective on McHugh's influence, see David DellaVolpe's foreword.

17. McHugh, *Fundamentals of War Gaming*.

18. For detailed descriptions of wargaming methodology, see ibid., and Burns, *War Gamers' Handbook*.

19. Wargamers have different ways of categorizing the structure of their wargames. Common categories are one-sided, one-and-one-half-sided, two-sided, multisided, and red teaming.

20. See Burns, *War Gamers' Handbook,* for a description of the process.

21. By *replicate* I mean that another equally capable research team could analyze the data from game play and obtain the same findings. By *repeat* I mean that the same game design could be played again by another research team or the same research team with the same players or another group of players. By *iterate* I mean that a new game could be designed and played that builds on previous findings to extend the investigation of the war-fighting challenge.

22. The Navy Warfare Development Command (NWDC) manages and executes the Fleet Experimentation program on behalf of Commander, U.S. Fleet Forces Command and Commander, U.S. Pacific Fleet. For information, see "Navy Warfare Development Command," *America's Navy,* www.navy.mil/.

23. The following section draws on M. M. Polski, "Technical Brief on Quasi-Experimental Design" (paper prepared for the WG 30 Special Session, "Are War Games Quasi-Experiments?," of the Military Operations Research Society's 83rd Symposium, Alexandria, VA, June 23, 2015).

24. William Shadish, Thomas Cook, and Donald Campbell identify five different types of experiments: an experiment, a randomized experiment, a quasi experiment, a natural experiment, and a correlational study. They also provide examples of the use of the experimental method across disciplines. See William R. Shadish, Thomas D. Cook, and Donald T. Campbell, *Experimental and Quasi-Experimental Designs for Generalized Causal Inference* (Boston: Houghton Mifflin, 2001).

25. See Richard A. Kass, *The Logic of Warfighting Experiments,* Command and Control Research Program Publication Series (Washington, DC: Office of the Assistant Secretary of Defense, 2006); the Technical Cooperation Program [hereafter TTCP], *Guide for Understanding and Implementing Defense Experimentation* [hereafter *GUIDEx*], ver. 1.1 (Ottowa: Canadian Forces Experimentation Centre, February 2006); and NWDC, *Experiment Planning Guide* [hereafter *EPG*], 2017 ed.

26. See, for example, Elinor Ostrom, Roy Gardner, and James Walker, *Rules, Games, and Common-Pool Resources* (Ann Arbor: Univ. of Michigan Press, 1994); John H. Kagel and Alvin E. Roth, eds., *The Handbook of Experimental Economics* (Princeton, NJ: Princeton Univ. Press, 1995); Nigel Gilbert and Klaus G. Troitzsch, *Simulation for the Social Scientist,* 2nd ed. (New York: Open Univ. Press, 2010); John T. Cacioppo, Penny S. Visser, and Cynthia L. Pickett, eds., *Social Neuroscience: People Thinking about Thinking People* (Cambridge, MA: MIT, 2006); and Dietrich Dörner, *The Logic of Failure: Recognizing and Avoiding Error in Complex Situations* (New York: Basic Books, 1997). Many empirical social scientists conceptualize decision challenges using game theory, then test these game-theoretic models using experimental methods.

27. Kass, *The Logic of Warfighting Experiments;* TTCP, *GUIDEx;* NWDC, *EPG.*

28. Martin J. Osborne and Ariel Rubinstein, *A Course in Game Theory* (Cambridge, MA: MIT Press, 1994); John von Neumann, "Zur Theorie der Gesellschaftsspiele," *Mathematische Annalen* 100 (1928), pp. 295–320; John von Neumann, "Über ein ökonomisches Gleichungssystem und eine Verallgemeinerung des Brouwerschen Fixpunktsatzes," *Ergebnisse eines mathematischen Kolloquiums* 8 (1937), pp. 73–83; John von Neumann and Oskar Morgenstern, *Theory of Games and Economic Behavior* (Princeton, NJ: Princeton Univ. Press, 1944). In addition to von Neumann and Morgenstern, classic texts

include R. Duncan Luce and Howard Raiffa, *Games and Decisions: Introduction and Critical Survey* (Hoboken, NJ: Wiley, 1957); Martin Shubik, *Game Theory in the Social Sciences: Concepts and Solutions* (Cambridge, MA: MIT Press, 1982); Hervé Moulin, *Game Theory for the Social Sciences* (New York: New York Univ. Press, 1986); Drew Fudenberg and Jean Tirole, *Game Theory* (Cambridge, MA: MIT Press, 1991); Roger B. Myerson, *Game Theory: Analysis of Conflict* (Cambridge, MA: Harvard Univ. Press, 1991); and Kenneth G. Binmore, *Fun and Games: A Text on Game Theory* (Lexington, MA: D. C. Heath, 1992).

29. Game-theoretic analyses are used extensively in institutional political economy, experimental economics, and computational social science. See, for example, Ostrom, Gardner, and Walker, *Rules, Games, and Common-Pool Resources;* Kagel and Roth, *The Handbook of Experimental Economics;* Robert H. Bates et al., *Analytic Narratives* (Princeton, NJ: Princeton Univ. Press, 1998); and Gilbert and Troitzsch, *Simulation for the Social Scientist*.

30. For an alternative but complementary approach to analyzing and comparing methods, see Paul K. Davis, *Analysis to Inform Defense Planning despite Austerity* (Santa Monica, CA: RAND, 2014), and Paul K. Davis and Amy Henninger, *Analysis, Analysis Practices, and Implications for Modeling and Simulation* (Santa Monica, CA: RAND, 2007).

31. For an early critique of game-theoretic analyses, see Schelling, *The Strategy of Conflict,* chap. 4. Also see Thomas C. Schelling, "The Role of War Games and Exercises," in *Managing Nuclear Operations,* ed. Ashton B. Carter, John D. Steinbruner, and Charles A. Zraket (Washington, DC: Brookings Institution, 1987), and Thomas C. Schelling, "Red vs. Blue" (paper presented at a wargaming conference organized by the National Defense Univ. at U.S. Marine Headquarters, Quantico, VA, August 2015). Schelling is often—and wrongly—described as a game theorist. In a conversation with him in September 2014, I asked about this identification. He told me that he had never worked with game theory, but that he thought some of its organizing ideas were useful for thinking systematically about strategic interaction.

32. TTCP, *GUIDEx* contains fourteen principles for designing defense research; five of them (nos. 4–7 and 10) are related to integrating and leveraging multiple operations research methods.

33. This section is based on Margaret M. Polski, "The Navy's Experience with Integrative Decision Analysis" (unclassified case study prepared for the Military Operations Research Society wargaming community of practice) and classified cross-game studies the author conducted in July 2017, December 2017, August 2018, and November 2018 for the CNO.

34. Polski, "The Navy's Experience wth Integrative Decision Analysis."