

POLICY FORUM

INTERNATIONAL SECURITY

Next-generation wargames

Technology enables new research designs, and more data

By **Andrew W. Reddie**^{1,2}, **Bethany L. Goldblum**³, **Kiran Lakkaraju**⁴, **Jason Reinhardt**⁴, **Michael Nacht**⁵, **Laura Epifanovskaya**⁴

Over the past century, and particularly since the outset of the Cold War, wargames (interactive simulations used to evaluate aspects of tactics, operations, and strategy) have become an integral means for militaries and policy-makers to evaluate how strategic decisions are made related to nuclear weapons strategy and international security (1). These methods have also been applied beyond the military realm, to examine phenomena as varied as elections, government policy, international trade, and supply-chain mechanics. Today, a renewed focus on wargaming combined with access to sophisticated and inexpensive drag-and-drop digital game development frameworks and new cloud computing architectures have democratized the ability to enable massive multiplayer gaming experiences. With the integration of simulation tools and experimental methods from a variety of social science disciplines, a science-based experimental gaming approach has the potential to transform the insights generated from gaming by creating human-derived, large-*n* datasets for replicable, quantitative analysis. In the following, we outline challenges associated with contemporary simulation and wargaming tools, investigate where scholars have searched for game data, and explore the utility of new experimental gaming and data analysis methods in both policy-making and academic settings.

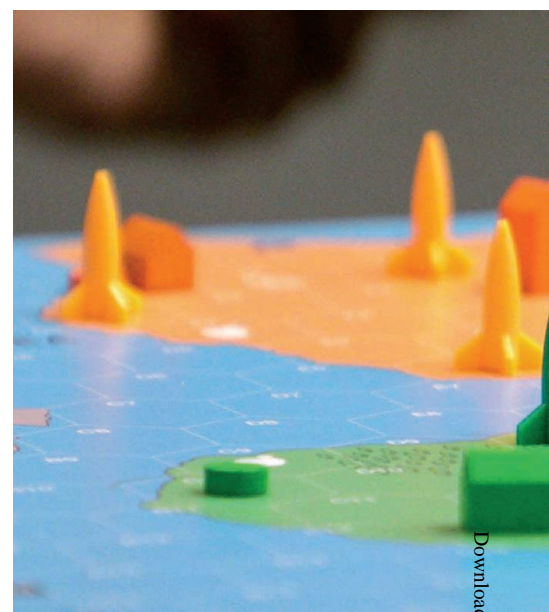
THEORY RICH, DATA POOR

Increasingly, simulations have relied upon mathematical computer-based models to make inferences about real-world behavior regarding conflict and cooperation. How-

ever, in certain situations, observational data are limited or there are practical or ethical quandaries associated with producing them. This lack of data is a particularly salient problem for nuclear deterrence models, given the fortunate lack of observational data regarding nuclear weapon use. In such situations, these simplified, “toy” models used to explain phenomena as complicated as international cooperation or nuclear escalation patterns can fail to take into account the human factors that drive policy-making decisions. For example, model assumptions such as player rationality may not hold in conditions of crisis or when players have little time to make decisions.

To introduce human factors into simulations, policy-makers, economists, and social scientists have relied on exploratory simulation games, structured play environments that can be used to devise representations of real-world decision-making (2). Focus has traditionally been on analog games involving a limited player set and a single scenario. The constraints of the predigital environment restricted the complexity of the gameplay and largely prohibited the collection of sufficient data for generalizable inquiry, leading to wargaming being described as an “art rather than a science” (3). For example, the Sigma II-64 wargame created to strategize U.S. policy in Vietnam required more than 40 analysts and months of planning at the RAND Corporation to develop scenarios that involved upward of 35 players.

Among existing simulation game approaches, there are few experimental studies. Instead, both policy-makers and scholars have tended to focus upon process-oriented investigations of behavior inside of the game environment (for example, analyzing the dialogue between players to achieve exploratory insights) rather than utilizing each game as a unit of analysis for causal inference. These exploratory games may take a variety of forms—whether assessing what questions business executives might ask following a substantial drop in oil prices or examining military planners’ decision-making processes related to the use of cyber weapons (4, 5). These discussions can be particularly valuable when high-level policy-makers involved in real-

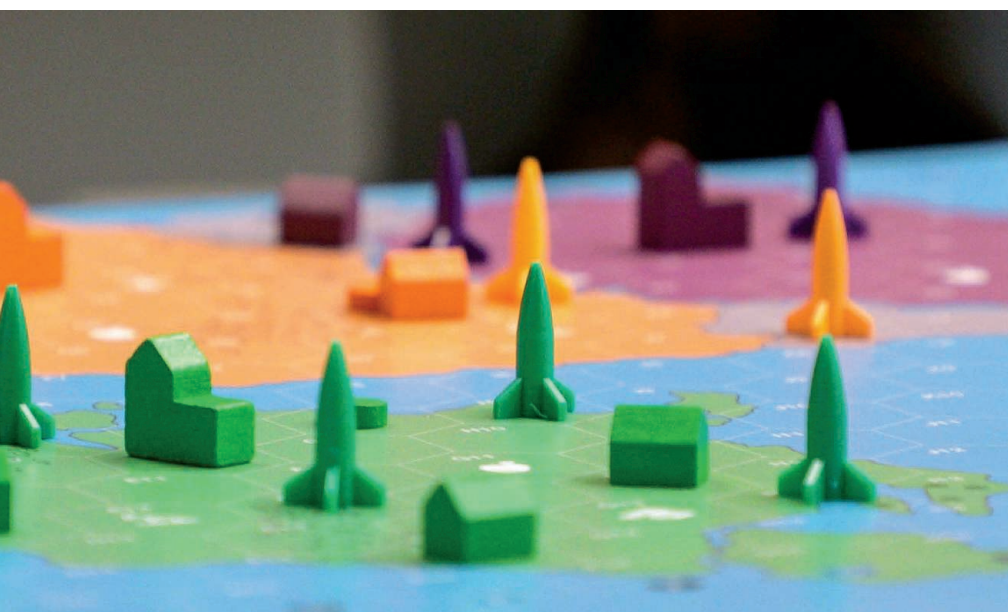


world decision-making are engaged and provide their expertise and insights. The U.S. Naval War College and U.S. Strategic Command, for example, conduct the Deterrence and Escalation Game and Review series, a two-sided game to explore escalation dynamics during crises. These types of games also offer an opportunity to explore counterfactuals while allowing game designers to track the discussions.

Alternatively, structured exercises akin to board games offer designers increased control over game dynamics and the potential to increase the number of play-throughs for postexperiment analysis. For example, Karl Mueller at RAND has led an effort to create a tabletop exercise to explicitly address the challenge posed by a resurgent Russia in the Baltic region following the former’s invasion of Ukraine in 2014 (6). The goal of this game, carried out with players from the U.S. Air Force and U.S. Army, is to inform the appropriate force composition necessary to defend North Atlantic Treaty Organization (NATO) members and deter adversaries in the region, given adversary capabilities. These insights into strategic decision-making have subsequently driven debate concerning the appropriate qualitative and quantitative force postures in the region.

Even with this type of game using stylized rules, players are subject to laboratory effects. For example, players sitting across from one another may hold back from aggressive maneuvers given the reputational costs associated with taking such an action amid their peers. Inferences related to structured exercises have also been called into question given the small number of players involved and the limited number of turns that may fail to capture real-world dynamics.

¹Department of Political Science, University of California, Berkeley, Berkeley, CA, USA. ²Center for Global Security Research, Lawrence Livermore National Laboratory, Livermore, CA, USA. ³Department of Nuclear Engineering, University of California, Berkeley, Berkeley, CA, USA. ⁴Sandia National Laboratories, Albuquerque, NM, USA. ⁵Goldman School of Public Policy, University of California, Berkeley, Berkeley, CA, USA. Email: areddie@berkeley.edu



The external validity of postgame analysis is also made difficult by the limited number of play-throughs of each game and the challenges associated with collecting data in a manner that allows for replication of the experiment. These difficulties combine to reduce the likelihood of statistical analysis and limit the production of generalizable insights. As a result, board wargame outcomes, like exploratory games, tend to be characterized by insights derived from game play rather than generalizable conclusions based on objective postgame analyses.

These exploratory approaches have also led to a focus on simulated games for training or educational purposes rather than experimentation. For example, the Apex Gold scenario-based discussion program designed by the National Nuclear Security Administration for senior-level policy-makers tests how players work together and respond to a hypothetical nuclear terrorism threat. A series of questions and polls also drive the discussion between participants to emphasize the challenges associated with addressing nuclear security.

In sum, computer-based simulations, exploratory games, and structured exercises have led to a theory-rich, but data-poor, environment for scholarly inquiry.

THE SEARCH FOR EXISTING GAME DATA

There are a variety of ways to address this paucity of data by using experimental methods. A number of scholars are attempting to use archival material to reexamine past games for generalizable insights. For example, Reid Pauly uses material from the Massachusetts Institute of Technology (MIT) and the U.S. Department of Defense to collect notes and game outcomes from

wargames designed by Lincoln Bloomfield and Thomas Schelling in the 1960s (7). Observing similar behavior across the collection of games, Pauly suggests that policy-makers exhibit a generalizable pattern of behavior resulting in nuclear restraint during crises and in spite of provocation. Jacquelyn Schneider's recent project similarly examines longitudinal data created over 7 years of Naval War College wargames involving cyber weapons and finds that, contrary to statements from policy-makers, cyber capabilities do not appear to contribute to crisis instability (8).

Both projects mine collections of outcomes obtained from traditional wargaming approaches to provide quantitative insights not inferable from a single play-through. These approaches, however, present research design challenges, including being time intensive and having no potential for automation in terms of data collection or for adaptation to address alternative research questions that go beyond the original scenario for which the games were designed. There are also concerns about comparing apples to oranges as scenarios shift in terms of their framing, the identity and institutional affiliation of the players, and changes to the geopolitical context.

As well as using archival material from traditional wargames, scholars have looked to commercial games to provide natural experiments during the course of gameplay that are analogous to real-world settings—in spite of not explicitly being designed for research purposes (9). A famous example of commercial data providing a simulation of reality for scholars comes from the World of Warcraft, in which the first gamewide epidemic in a massive, multiplayer, online

Scholars at play study the impact of weapons capabilities on conflict escalation using the Project on Nuclear Gaming's board game platform, SIGNAL.

role-playing game led to 4 million player characters being affected by something not unlike a virus. Epidemiologists subsequently used the virus that spread from player to player throughout the game's "world" to model transmission rates and the chain of infection and compared these findings to real-world pandemics (10).

Over the past decade, numerous scientific disciplines have started to explore the use of commercial games for experimental inquiry. For example, virtual worlds have become a laboratory for ethnographic research concerning social behavior on platforms such as Second Life, where players have the potential to create entirely new identities and social relationships. Economists have been examining virtual currencies and financial systems in online games and their implications related to cryptocurrencies, decentralized finance, and blockchain banking. For war planners, battles such as the "Bloodbath of B-R5RB" on the Eve Online platform involving about 7500 human players, 20 million virtual soldiers, and 600 warships provide a virtual example to study the origins, conduct, and outcome of large-scale warfare in the absence of real-world corollaries (11).

However, these gaming environments are largely outside of researchers' control, which may limit the theories that can be tested. In an attempt to control an experimental environment using commercial software, social scientists have also created stand-alone "mods"—user-generated environments—in games such as World of Warcraft, Star Wars Galaxies, and Starcraft 2 to consider theories of human behavior, cooperation, and conflict from political science, economics, and sociology (12). One of the first of these efforts, NetLab, led to the creation of a variety of "collaboratories" in the early 2000s that sought to use the internet to field social and behavioral experiments (13). "Tribes," for example, modeled on intertribal rivalry in real-world Sudan measures the (artificially created) inter- and intragroup dynamics. Although these mods provide researchers with increased control over the structured play environment, they remain subject to the virtual world, characters, and player pools associated with the original game publisher.

BUILDING EXPERIMENTAL CONDITIONS

Now, scholars are increasingly able to build experimental settings from scratch with the goal of conducting replicable, quantitative analyses that focus on a particular real-

world policy decision. Building games from the ground up addresses the problem that researchers do not have control of the game setting and, subsequently, the treatments provided to game participants. Although this option was, until recently, prohibitively expensive, the production of online experiment-based games tailored to address particular questions of interest is becoming progressively accessible to researchers as low-cost gaming architectures and visual scripting systems proliferate.

With that said, game design and development presents its own challenges. At the outset, researchers need to decide what type of game to build—they range from simple to complex, turn-based to simultaneous, and vary in terms of the number of players, treatment variables, aesthetics, mechanics, and the interface that players interact with. Each of these features can affect the parsimony, internal validity, and external validity associated with the game. For example, maps in real-time strategy games can be designed to provide perfect information to all players or, alternatively, the moves of adversaries can be hidden under a “fog of war.” The latter is more realistic but also more complicated to design and build into a gaming framework.

Although more research is needed to understand the optimal design and limitations of online experimental games, they do allow for replicable, structured rulesets, iterative turn-based play, and an increased number of play-throughs that overcome a number of challenges related to traditional wargaming methods. The online environment also provides access to a diverse set of research participants that traditional wargames do not engage with—Steam’s gaming platform, for example, regularly hosts more than 10 million unique concurrent online gamers, whereas Amazon’s Mechanical Turk provides an alternative source of experimental participants in online settings.

Our Project on Nuclear Gaming (PONG)—a collaboration between the University of California, Berkeley; Lawrence Livermore National Laboratory; and Sandia National Laboratories—and its SIGNAL game serves as a first execution of this large-scale, experimental gaming approach in an examination of nuclear deterrence and conflict escalation dynamics. SIGNAL provides a flexible gaming environment that can be used to explore a variety of research questions and mimic aspects of warfare. The platform also allows for large-*n*, quantitative analysis of game outcomes in a multi-player context, tracks demographic data, and automatically collects player and game data in real time.

As new experimental gaming tools mature, they are well placed to take advantage of advances in data science and machine learning. For example, game data can be used as an input in machine learning algorithms to expand the amount of data available to create models of optimal behavior given particular experimental conditions (14).

These data might be leveraged to create autonomous agents that are representative of various player strategies in the training data as well as serve as venues for human-machine and machine-machine gameplay. Comparative analysis of these player models and their parameters may reveal insights about player “types” that can further augment our understanding of conflict strategies and crisis communications. Alternatively, inverse reinforcement learning methods can be applied to gameplay data to assess player perceptions of rewards, constraints, and optimal implementation strategies within a simulated environment.

METHODOLOGICAL CHALLENGES

Although the ability to build customizable online games that can be tailored to specific research questions offers a promising path toward experimental gaming, there are a number of methodological challenges that scholars must consider. First, game designers must address how the online game setting might lead to its own unique laboratory effects related to player behavior within the game environment. Might it, for example, make players more aggressive than they otherwise might be? Relatedly, the impact of using nonexperts rather than experts in the player pool is in need of further examination. Analyses are needed to compare expert to nonexpert players in the context of decision-making pertaining to national security settings.

Experimental gaming techniques that allow for large-*n* analysis using nonexpert players should also be compared against insights generated from traditional wargames involving experienced decision-makers. Addressing these challenges requires cross-method and cross-subject comparison. And although all gaming frameworks are predominantly focused on internal rather than external validity, more work is needed to further link game findings from online environments to real-world dynamics using observational data, where available.

As experimental gaming methods mature, its toolkit can be tailored to address extant questions being asked by the policy and defense communities—testing conventional wisdom and providing another data source for decision-makers. It may also offer a capability to analyze increasingly complex security dynamics that involve new

types of actors and new domains of warfare, including space, cyber, and gray-zone operations. Already, government agencies, particularly those related to national security, are increasing their use of simulations in policy planning. Efforts are also underway to build data repositories such as the nascent Wargaming Repository created by the U.S. Office of the Secretary of Defense to pool insights, generate conclusions that might otherwise be hidden, and further refine existing wargaming methods—whether analog or digital.

Although an ability to reliably predict the actions of adversaries and the outcomes of conflict remains unlikely, new types of experimental tools have the potential to shed light on dynamics that have thus far existed only in theory. ■

REFERENCES AND NOTES

1. R. Work, “Wargaming and Innovation,” Memorandum, Deputy Secretary of Defense (9 February 2015).
2. P. Sabin, *Simulating War: Studying Conflict Through Simulation Games* (Bloomsbury Academic, 2014).
3. P. Perla, *The Art of Wargaming: A Guide for Professionals and Hobbyists* (Naval Institute Press, 1990).
4. N. H. Barma, B. Durbin, E. Lorber, R. E. Whitlark, *Int. Stud. Perspect.* **17**, 117 (2015).
5. J. Schneider, “Cyber and crisis escalation: Insights from wargaming” (U.S. Naval War College, 2017); <https://pacs.einaudi.cornell.edu/sites/pacs/files/Schneider.Cyber%20and%20Crisis%20Escalation%20Insights%20from%20Wargaming%20Schneider%20for%20Cornell.10-12-17.pdf>.
6. K. Mueller, “Paper wargames and policy making: Filling the Baltic gap,” *Battles Magazine*, No. 11 (May 2016); https://www.rand.org/content/dam/rand/pubs/external_publications/EP60000/EP66660/RAND_EP66660.pdf.
7. R. B. C. Pauly, *Int. Secur.* **43**, 151 (2018).
8. J. Schneider, “The information revolution and international stability: A multi-article exploration of computing, cyber, and incentives for conflict,” thesis, The George Washington University, 2017.
9. W. S. Bainbridge, *The Warcraft Civilization: Social Science in a Virtual World* (MIT Press, 2012).
10. E. T. Lofgren, N. H. Fefferman, *Lancet Infect. Dis.* **7**, 625 (2007).
11. D. Fandino, in *Virtual Dark Tourism*, K. McDaniel, Ed. (Palgrave Macmillan, 2018).
12. A. M. Kaplan, M. Haenlein, *Bus. Horiz.* **52**, 563 (2009).
13. S. Teasley, S. Wolinsky, *Science* **292**, 2254 (2001).
14. G. N. Yannakakis, J. Togelius, *Artificial Intelligence and Games* (Springer, 2018).

ACKNOWLEDGMENTS

This work was supported by a grant from the Carnegie Corporation of New York through their International Peace and Security Program. The authors recognize support from the Nuclear Science and Security Consortium at the University of California, Berkeley, under contract DE-NA0003180; the Center for Global Security Research at Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344; and Sandia National Laboratories, a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy’s National Nuclear Security Administration under contract DE-NA0003525. The views and opinions of the authors expressed herein do not necessarily state or reflect those of the United States government; Lawrence Livermore National Security, LLC; or Sandia National Laboratories. The authors would also like to acknowledge members of the Project on Nuclear Gaming past and present and thank two anonymous reviewers for their helpful comments.

10.1126/science.aav2135