Incorporating Gaming into Research Programs in International Relations: Repetition, Game Series, and Multi-Method Analysis

E. M. Bartels

APSA Annual Conference
September 11, 2020

Abstract:
Existing literature on the use of games to support research on international relations is largely disconnected from the academic literature on research design generally, and multi-method research design in particular. The majority of gaming literature currently comes out of the interdisciplinary practitioner community, who have generally been focused on pragmatic considerations. Popular works on game design often come out of the commercial gaming industry, where research considerations are not a core driver of design choices. Finally, works from international relations tend to focus on games as a teaching tool or on games as they have contributed to specific avenues of research. It is only recently that the use of games as a tool for research is being addressed as a subject of study it its own right within contemporary political science. As I have previously argued, this turn towards integrating games into the frameworks and concepts applied to other tools for social science research is critical to ensuring that the insights drawn from games are sound, as well as for making gaming as a tool more accessible to new researchers. This paper expands on previous work conceptualizing games within social science research design to discuss how games can be integrated into broader studies by exploring three approaches: repeated games, serial games, and games in multi-methods studies.

Existing literature on the use of games to support research on international relations is largely disconnected from the academic literature on research design generally, and multi-method research design in particular. The majority of gaming literature currently comes out of the interdisciplinary practitioner community, who have generally been focused on pragmatic considerations.[1] Popular works on game design often come out of the commercial gaming industry, where research considerations are not a core driver of design choices.[2] Finally, works from international relations tend to focus on games as a teaching tool[3] or on games as they contributed to specific avenues of research.[4] It is only recently that the use of games as a tool for research is being addressed as a subject of study in its own right within contemporary political science.[5] As I have previously argued,[6] this turn towards integrating games into the frameworks and concepts applied to other tools for social science research is critical to ensuring that the insights drawn from games are sound, as well as for making the tool more accessible to new researchers.

This paper expands on previous work conceptualizing games within social science research design to discuss how games can be integrated into broader studies by exploring three approaches: repeated games, serial games, and games in multi-methods studies. First, I discuss why it might be valuable to run a game more than once—and why value of doing so might differ depending on the information to be produced. I then turn to discussing studies that use multiple

---

[1] Key practitioner texts include: Peter P. Perla, *The Art of War Gaming: A Guide for Professionals and Hobbyists*, ed. John Curry, 2nd ed. (History of Wargaming Project, 2011); Matthew B. Caffrey, *On Wargaming: How Wargames Have Shaped History and How They May Shape the Future* (Newport, RI: Naval War College Press, 2019); Graham Longley Brown, *Successful Professional Wargames: A Practitioner's Guide*, ed. John Curry (The History of Wargaming Project, 2019).

[2] For notable examples, see: Katie Salen Tekinbas and Eric Zimmerman, *Rules of Play: Game Design Fundamentals* (Cambridge MA: MIT Press, 2003); George Skaff Elias, Richard Garfield, and K. Robert Gutschera, *Characteristics of Games* (Cambridge MA: MIT Press, 2012); Engelstein. Geoffrey and Issac Shalev, *Building Blocks of Tabletop Game Design: An Encyclopedia of Mechanisms* (Boca Raton, FL: CRC Press, 2019).

[3] Victor Asal, "Playing Games with International Relations," *International Studies Perspectives* 6, no. 3 (2005).

[4] For recent work, see: Reid B.C. Pauly, "Would U.S. Leaders Push the Button? Wargames and the Sources of Nuclear Restraint," *International Security* 43, no. 2 (2018). Jacquelyn G. Schneider, "Cyber Attacks on Critical Infrastructure: Insights from War Gaming," *War on the Rocks*, July 26 2017. Andrew W. Reddie et al., "Next-Generation Wargames: Technology Enables New Research Designs, and More Data," *Science* 362, no. 6421 (2018); Erik Lin-Greenberg, "(War)Game of Drones: Remote Warfighting Technology and Escalation Control Evidence from Wargames," (SSRN, 2019); Benjamin Jensen and Brandon Valeriano, "Cyber Escalation Dynamics: Results from War Game Experiments," in *International Studies Association* (Toronto, Canada2019).

[5] Recent work in this vein includes: Ivanka Barzashka, "Wargaming: How to Turn Vogue into Science," *Bulletin of the Atomic Scientists*2019. Sawyer Judge, "The Wargaming Guild: How the Nature of a Disipline Impacts Its Craft and Whether It Matters" (Georgetown University, 2019). Elizabeth M. Bartels, "Building Better Games for National Security Policy Analysis: Towards a Social Scientific Understanding" (Pardee RAND Graduate School, 2020). and Erik Lin-Greenberg, Reid B.C. Pauly, and Jacquelyn Schneider, "Wargaming for Political Science Research," in *APAS Annual Conference* (Virtual2020).

[6] Bartels, "Building Better Games for National Security Policy Analysis: Towards a Social Scientific Understanding."

tools, whether multiple types of games or other approaches to conducting research. First, I contextualize games within existing multi-method research practices in international relations. I then discuss some ways to think about using a series of games to make progress in a particular area of research. Finally, I turn to how to use games in conjunction with other approaches to research.

Establishing Games within Frameworks for Inquiry

This paper builds on a conceptual framework previously developed in Bartels 2020, which crosswalks archetypic game purposes referenced in the practitioner literature with Jackson 2010's framework for describing different philosophies of inquiry in international relations.[7] The four types of information that games are frequently used to produce are: systems exploration, alternative conditions, innovation, and evaluation. All but evaluation suggests a dominant philosophy of inquiry, based on the logic that underpins information production, show in Table 1. The following discussion summarizes these arguments. By explicitly laying out the types of information games are designed to produce and linking it to the philosophy of inference that underpins each type, the framework helps create a common reference point between games and other approaches to inquiry in international relations, which developed in the following sections.

Table 1: Degree of alignment between the three philosophies and four archetypes

Archetype	Positivism	Critical Realism	Analyticism
System Exploration	*Weak*	Some	**Strong**
Alternative Conditions	**Strong**	Some	*Weak*
Innovation	*Weak*	**Strong**	Some
Evaluation	Some	Some	Some

Source: Elizabeth M. Bartels, "Building Better Games for National Security Policy Analysis: Towards a Social Scientific Understanding." Pardee RAND Graduate School, 2020. p 73

System Exploration

This archetype highlights games that bring together diverse stakeholders to contribute their understanding of the policy system to generate a rich description. The primary goal of such games is to elicit and synthesize designers' and players' mental model of a policy problem and how it may evolve over time. The output from successful system exploration games is a representation or model of the problem that combines insight from players with research performed by the design team to improve the sponsor's understanding of the nature of the

[7] Patrick Thaddeus Jackson, "The Conduct of Inquiry in International Relations: Philosophy of Science and Its Implications for the Study of World Politics," (New York, NY: Routledge, 2011).

problem. This type of game is similar to past descriptions of games that are used to organize the knowledge of a research team,[8] games to suggest what factors and relationships are important,[9] games to reveal poorly understood dynamics,[10] and games to integrate knowledge,[11] identify issues, and build consensus among participants.[12]

System exploration games align quite well with the logic of analyticist research. Analyticism argues that theory is an act of sensemaking that tries to explain what is being observed.[13] Researchers in the frame develop models that are simple, and thus inherently non-representative, of the true complexity of the world, but are useful for the particular purpose at hand. Such models are rejected not for being wrong but for not being useful in explaining the specific case at issue.[14] If science is primarily an act of sense-making, then the use of games to develop a simplified model that represents the designer's and players' efforts to understand the phenomena of interest appears to be quite a useful activity. If the designers and players (or those who encounter the resulting model) find the resulting simplified model useful, either when confronting a similar problem in the real world or in setting further analytical research programs, then the exercise of system exploration gaming is useful to this mode of inquiry. Given the ample evidence we have for players and researchers finding utility in gaming,[15] practitioners of analyticism will have no problem making a case for the pragmatic usefulness of games to explore issues.

Alternative Conditions

Alternative conditions games aim to detect patterns of decisionmaking based on similarities and differences in the decisionmaking environment to help advance causal inference. Game designs of this type try to minimize variation in environment, actors, rules, and models across iterations while purposefully changing selected key factors. Successful games of this type produce an understanding of the influence of varying conditions on either the decisionmaking

---

[8] Milton G. Weiner, "An Introduction to War Games," in *Lex Choix Economiques: Decisions Sequentielles Et Simulation*, ed. Pierre Rosenstiehl and Alain Ghouila-Houri (Paris: Dunod, 1960). p 25

[9] Ibid. p 25

[10] Herbert Goldhamer, "The Political Exercise: A Summary of the Social Science Division's Work in Political Gaming, with Special Reference to the Third Exercise July-August 1955," (Santa Monica, CA: RAND Corporation, 1955). p 1-4

[11] Edward Parson, "What Can You Learn from a Game?," in *Wise Choices: Decisions, Games, and Negotiations*, ed. Ralpj L. Keeney Richard J. Zeckhauser, James K. Sebenius (Boston: Harvard Business School Press, 1996).

[12] Perla, *The Art of War Gaming: A Guide for Professionals and Hobbyists*. p 181

[13] Jackson, "The Conduct of Inquiry in International Relations: Philosophy of Science and Its Implications for the Study of World Politics." p 114

[14] Ibid. p 143-144

[15] For example of a famous assertion of the utility of gaming, see: Robert A. Levine, Thomas C. Schelling, and William M. Jones, "Crisis Games 27 Years Later : Plus C'est Deja Vu," (Santa Monica, CA: RAND Corporation, 1991). p 23

process or the eventual decisions. Past typologies describing similar games have referred to games that leverage comparison to better understand the effects of a change in context[16] and well as providing a grounds for theory building about the role of the environment.[17] Descriptions of games as experiments sometimes fit this type (though often also describe applications for evaluation, which I treat separately).[18]

The alternative conditions approach fits extremely well with the positivist research agenda. Positivists are interested in understanding whether general, law-like statements about causality between discrete factors can correctly describe observed patterns.[19] Put differently, this tradition attempts to describe the difference in some outcome Y, based on the presence of different values of some causal factor X. Use of structured cross-game comparison in alternative conditions games make them ideal for studying causality through a difference-based lens. While the inherent variation between players and their interactions prevents perfect experimental control,[20] this can be managed in ways that are consistent with quasi-experimental traditions that are popular in a wide range of positivist research projects.[21] Because the researcher is able to observe the decision process, there is ample opportunity to measure a range of potential causes and their influence on decisionmaking. In other words, alternative conditions games are explicitly framed using the logic of positivist research and thus fit neatly into this philosophical frame.

Innovation

Innovation games seek to develop new decision options that break from the status quo as a form of policy ideation. These games build a model of the world that relaxes constraints in the hopes that doing so might enable new approaches to problem solving. In this way, they share similarities with hypothesis generation and brainstorming activities. The ideal outcome for this type of game is to generate one or more promising ideas for further consideration. Past typologies have discussed these games as developing strategies and plans,[22] producing innovation and strategic inventiveness[23] and to promote creativity and insights.[24]

---

[16] Weiner, "An Introduction to War Games." p 28 and Longley Brown, *Successful Professional Wargames: A Practitioner's Guide*. p 91.

[17] Weiner, "An Introduction to War Games." p 25

[18] Parson, "What Can You Learn from a Game?."

[19] Jackson, "The Conduct of Inquiry in International Relations: Philosophy of Science and Its Implications for the Study of World Politics." p 108

[20] Parson, "What Can You Learn from a Game?." p 238.

[21] Elizabeth M. Bartels to Paxsims, 2015.

[22] Perla, *The Art of War Gaming: A Guide for Professionals and Hobbyists*.p 181

[23] Goldhamer, "The Political Exercise: A Summary of the Social Science Division's Work in Political Gaming, with Special Reference to the Third Exercise July-August 1955." p 1-4

[24] Parson, "What Can You Learn from a Game?."

Perhaps the most natural alignment for innovation games is with the critical realist approach. Critical realist accounts argue that real, but unobservable, phenomena can be studied scientifically through a process of *abduction*. Scientists gather evidence from the surrounding system and make a plausible causal explanation—often in the form of a mechanism—based on all available evidence. As the available evidence changes, the causal theory may evolve; however the theory is still fundamentally unproven by this process—abduction cannot demonstrate truth, only plausibility.[25] Critical realist models of innovation games focus on generating a strategy through abduction—that is, players use the context of the game, including competitive pressures, to generate a strategy. The focus on causal mechanisms pairs nicely with the need for attention to policy process. That is, players cannot simply identify a causal factor to define a strategy, but rather must play out how to enact change over time, through actions and mechanisms that can produce the effect of interest. Critical realists would argue that the resulting theory of success has not been proven to be true, but merely generated as a potential theory for testing as additional evidence is gathered is also highly consistent with the generation of innovative ideas.

Evaluation

The evaluation archetype describes games that aim to judge the potential outcomes of player decisions based on a normative standard—in other words: to evaluate policies, courses of action, or interventions. These games focus great attention on adjudication to generate credible outcomes from player decisions. Because the game must project plausible outcomes in order to enable evaluation of the results of decisions, it must contain a fairly well-developed theory of causality that allows the game staff to project different counterfactual outcomes based on player actions. The desired outcome of these games is an assessment of the potential gains and losses from following a course of action. Other scholars have highlighted similar roles such as: playing out a plan, policy, or weapon to get a sense of its strengths and weaknesses;[26] testing strategies and plans; [27] evaluation; and analysis.[28] Experimental games may also fall in this type, but do not always do so.

Evaluation games have an imperfect fit with all three approaches. Evaluation games share a common causal setup with most positivist evaluation. The catch is that rather than observing direct effects of the causal relationship as in alternative conditions games, game outcomes depend to a substantial degree on the use of a model to generate outcomes. Because adjudication models must, by definition, bake in a model of causality, the game cannot be used as evidence of

---

[25] Jackson, "The Conduct of Inquiry in International Relations: Philosophy of Science and Its Implications for the Study of World Politics." pp 82-83

[26] Weiner, "An Introduction to War Games." p 28

[27] Perla, *The Art of War Gaming: A Guide for Professionals and Hobbyists*. p 181

[28] Longley Brown, *Successful Professional Wargames: A Practitioner's Guide*. pp 89-91

the truthfulness of that causal model since it is endogenously connected to the results. Concretely, if a weapons system is assigned great destructive power in the adjudication model, findings of the weapon's destructive power are not an empirical result, they are a model artifact that contributes to positivist research only to the extent that the model has been generated using other approaches. As a result, while careful research is possible, positivists are likely to be skeptical of games for evaluation until evidence of either credible adjudication or lack of dependence on endogenous models is demonstrated.

Critical realist approaches to evaluation games are also possible but face some sharp limitations. Critical realism's focus on causal mechanisms puts greater weight on the evaluation of process than do positivist approaches that focus on measuring effect size through differencing. However, critical realists would be quite hesitant to make strong claims on the back of games alone—games can present evidence that is consistent with the posited causal process, but strong evidence likely requires other research approaches to generate. Furthermore, results may only be generalized to a very narrow set of cases that share similar context. Because games involve many artificial elements, it may be more difficult to define what set of cases the theory might reasonably extend to.

Analyticism also coheres with the goals of process evaluation to some degree, but the claims that result from such analysis are somewhat different than for the other two approaches. In analyticist approaches, the ideal type model is judged by usefulness—so an analyticist evaluation game might be best thought of as a test for the usefulness of some model of policy in a particular situation. The catch is that the situation is fictitious, and analyticism does not support efforts to generalize. Thus, the output of an analyticist evaluation game is the determination that a model is useful for the specific context of the game. It may prove to also be useful in other contexts, but may not, and the researcher must accept, and defend, the risk that game results will not prove to have real world utility before undertaking such an effort.

The Value of Repeating Games

What additional information might we obtain by repeating a game, and what does that imply about how we are learning from it? First, it is worth being clear about what is meant by a repeated game. I use the term to refer to using the same game design to generate the same types of information for the same analytic purpose. There will always be changes between runs of a game—most obviously players will be different, either because new players are recruited, or their decisions in the new run will have been changed by the previous play of the game. Additional changes to game materials are likely to occur as players provide feedback about what works and whether there are unintuitive elements or factual errors. However major design elements, and most importantly, the information the game is designed to generate, should be consistent.

In the practitioner community, the number of repetitions is determined primarily by resources or other logistical considerations rather than by the research design of the study. There is a general perception that running games multiple times is preferable[29] but with few clear articulations of what may be gained through repetition, how much analytic value it provides, and whether the gains are the same between different types of games. Thus, the general approach is to run as many games as is feasible given time and other resources—generally only one, sometimes a couple of repetitions, rarely more than 2-3. This section attempts to lay out a logic for what might be gained by repetition in the case of each of the four game archetypes under the dominant philosophical foundation described above to enable more thoughtful consideration of when and how repetition is analytically useful.

In considering a more nuanced argument in favor of repetition, I have found it useful not only to consider the philosophical foundations of each game type, but also to draw an analogy to other types of analysis to consider how games are similar (and different) from the logic of other potential approaches to conducting research. In part, relating games to existing approaches provides a common reference point to other, potentially more familiar, approaches to research which have been better documented. This logic also works in reverse—without a deliberate comparison, consumers may incorrectly apply the standards of other methods to games. For example, a major criticism leveled against games (particularly by practitioners of operations research and statistics) is that because games are generally run only a small number of times, the results of the game are not statistically significant and thus do not contribute valid knowledge. This criticism misses that the information created by games is generally not interested in establishing the central tendency of mathematical interactions, and thus a statistical logic is not applicable to most research designs calling for repeated games. By exploring why we might want to repeat games, we can help clarify the actual benefits, and how they are different for different types of games. However, it is important to be clear that in drawing these comparisons, the unique nature of games remains. As noted earlier, games inherently involve both synthetic elements and uncontrolled interactions between players. The analogies made in this section are just that—highlighting similarities in some respects but not necessarily positing exact correspondence. Appropriate consideration and caveats of the overall analysis are still necessary to produce credible information from games.

Systems Exploration

System exploration games, generally operating in an analyticist philosophical framework, are repeated to add more viewpoints, and thus build a more comprehensive understanding of the problem. In my experience, this process is analogous to conducting expert interviews as part of early stage research. For example, expert interviews are common tools for exploratory research, as well as more advanced work to synthesize expert technical and process knowledge or generate

[29] Ibid. p 114

"a theoretically rich conceptualization of (often implicit, yet reconstructible) knowledge, conceptions of the world and routines, which the experts and elites develop in their activities and which are constitutive for the functioning of social systems."[30] Put differently, systems exploration games and expert interviews often share the goal of building a model of a particular phenomenon that is informed by the explicit and tacit knowledge of specialists in the domain.

Turning this analogy to the question of repetition suggests two frameworks for valuing repetition. The first extends the logic of survey design and treats interview selection as a potential sampling issue, with large random samples held up as the ideal because of their constancy with generalizability for positivist inference.[31] This analogy focuses attention on the range of participants in the games and how well that lines up with real world stakeholders. For example, a game might be run several times at different stakeholder organizations to make sure the final model incorporated views of all major decisionmakers. However, attempts to use multiple games to ensure coverage of the population of interest generate several pragmatic concerns. First, expert and elite communities are often difficult to establish a robust sampling frame, or to make contact with a random, rather than a convenience sample[32]—a problem made worst by the necessity of coordinating multiple subjects to be available for a single, long period of time for game play. As a result, games will generally be more dependent on convenience sampling strategies. Unlike interviews, games involve groups of players making collaborative and competitive decisions, so the representation of different views in each game needs to be considered. While segmenting games by organization may be practically necessary (for example, if stakeholders are geographically dispersed, it may be more cost effective to take the game to each group separately) it's rarely ideal from a design perspective. Additionally, as noted above, systems explorations projects are unlikely to use a positivist frame, suggesting that this logic is likely to be less pressing.

The second, and perhaps more traditional approach is that of "saturation"—that is "you keep asking as long as you are getting different answers, and that is a reminder that with our little samples we can't establish frequencies but we should be able to find the RANGE of responses."[33] As you interview more people, you gain an understanding of what parts of their understanding are shared as well as some insights about what might be driving differences. At some point, the marginal returns of conducting more interviews is minimal because you've

---

[30] Alexander Bogner, Beate Littig, and Wolfgang Menz, "Generating Qualitative Data with Experts and Elites," in *The Sage Handbook of Qualitative Data Collecation*, ed. Uwe Flick (London, UK: SAGE Publications Ltd, 2018). P 10

[31] Julia F. Lynch, "Aligning Sampling Strategies with Analytic Goals," in *Interview Research in Political Science*, ed. Layna Mosley (Ithaca, NY: Cornell University Press, 2013).

[32] Kenneth Goldstein, "Getting in the Door: Sampling and Completing Elite Interviews," *PS: Political Science & Politics* 35, no. 4 (2002).

[33] Harry Wolcott, quoted in Sarah Elsie Baker and Rosalind Edwards, "Introduction," in *How Many Qualitative Interviews Is Enough? Exert Voices and Early Career Reflections on Sampling and Cases in Qualitative Research*, ed. Sarah Elsie Baker and Rosalind Edwards (National Centre of Research Methods, 2012). p 4

already captured the vast majority of how experts see the problem—that is, you have achieved saturation. So too with system exploration games—additional runs of the game allow you to speak to new experts and watch different combinations of experts interact with one another, until the model built in the game is fairly stable.

This "saturation" approach offers useful guidance, but also some clear pitfalls for a researcher. A commonly cited issue with this approach is that it requires "the researcher to combine sampling, data collection, and data analysis, rather than treating them as separate stages in a linear process."[34] One implication of this is that it is difficult, if not impossible for a researcher to anticipate how many interviews will be needed for saturation to occur in advance. Given the cost and time involved in each run of the game, this level of uncertainty poses a concern for a researcher planning out a campaign of games-based research. However, existing work stresses that the degree to which full saturation is necessary will depend on the research question and community norms.[35]

Applying this same logic to games, depending on the nature of the research topic and the initial game's make-up of players, relatively few or even no repetitions of the game may be valuable. If the purpose of the game is to scope follow on research, and the 80% solution is enough to inform next steps, then expensive additional iterations to get to 100% saturation may well not be a cost-effective use of a limited research budget. For more poorly understood or complex problems, repetition may be an important way of building understanding. However, after several repetitions, designers often find it productive to change the focus of the game significantly enough that they are no longer using the same design (an approach discussed in more detail later in this chapter). As a result, it is unusual to see the same system exploration game played more then a handful of times for research and analysis purposes.

Alternative Conditions

Alternative condition games use repetition in order to explore more variation without losing control over comparisons. Two possible analogies from the traditional tool kit of international relations are available: comparative case studies and experiments. The first focuses on exploring the range of variation. For example, if there are two factors, each with two variations of interest but you can only run two games, you have to accept that the factors will co-vary and be difficult to disentangle. In contrast, if you run four games, it is possible to compare games which are only one factor different to see if decisionmaking changes. The other uses many runs of the game to understand the central tendency of the population—that is it is fundamentally statistical in its approach. The case study analogy tends to be used by researchers interested in causal mechanism focused research in which rich data is critical, whereas the experimental metaphor tends to attract

[34]Alan Bryman quoted in ibid. p 5

[35] Ibid. pp 5-6

researchers interested in different-based measures. As a general rule, the former is more common in policy research, and the later in more academic studies of international relations.

In traditional comparative case study analysis cases are often selected to maximize variation on the key variable of interest while ensuring other variables are as comparable as possible.[36] This suggest that the number of games will be tied directly to the number of variables of interest and their relationships, with one game for each potential combination of variable states. This approach is particularly consistent with research programs that seek to describe causal mechanisms across a small number of variables, creating, in effect a small number of deep cases that can be compared for nuanced differences. However, as with case studies, it is critical to be cautious about over-generalizing the difference observed in these games. For one thing, game designers never have complete control over variation between games because of the key role of players. Thus, like case studies, analysis of repeated alternative conditions games should focus attention on exploring the evidence to support alternative hypotheses before making any type of strong claim from the game series. Because the games are artificial, and thus there is no empirical record of total cases to observe, it is not possible to use approaches like qualitative comparison analysis[37] in which all possible cases of a class are examined to draw stronger conclusions. Thus, while increasing the number of games can strengthen the evidence for a causal relationship by exploring potential alternative explanations, there will always be limits on what repetition can contribute to the robustness of findings. Put differently, repeated games can allow for more nuanced discussion of how mechanisms work under different conditions, but do less to address generalizability, since the full universe of cases is usually not knowable.

An alternative comparison that has become increasingly common in political science research looks to experimental design as a model.[38] This approach shares the basic structure of intentional variation between runs of the game, and often adopts the experimental vocabulary of a “treatment” and “control” condition to describe this variance. In contrast to the case study approach that studies a small number of iterations through deep comparison, experimental approaches seek larger samples of dozens to hundreds of games.[39] Here, repetition is intended to provide an understanding of the central tendency of the population, and so requirements for the number of repetitions are driven by statistical rules. However, often practical limitations, not

---

[36] Alexander George and Andrew Bennet, *Case Studies and Theory Development in the Social Sciences* (Boston, MA: MIT Press, 2005).

[37] Charles C. Ragin, *The Comparitive Method: Moving Beyond Qualitative and Quantitiative Strategies* (Oakland, CA: University of California Press, 2014).

[38] Susan D. Hyde, "Experiments in International Relations: Lab, Survey, and Field," *Annual Review of Political Science* 18 (2015).

[39] Reddie et al., "Next-Generation Wargames: Technology Enables New Research Designs, and More Data."; Jensen and Valeriano, "Cyber Escalation Dynamics: Results from War Game Experiments."

only in terms of resources but also in terms of access to the population of interest complicate the practicality of this approach.[40]

Innovation

Because innovation is inherently speculative and proscriptive, its goals tend to deviate from mainstream international relations to the domain of public policy, and thus there is less of a natural point of comparison with standard political science approaches to research. However, innovation games have a very simple reason motivating multiple runs: novel, good ideas are hard to come by; running the game more times provides more chances for a rare event to occur. Solutions that are likely to occur to stakeholders are also likely (though not always!) to have been discarded for good reason. To come up with a new answer requires looking at the problem differently, being willing to adjust the bureaucratic rules, or envisioning a new tool that has not been seen before. Any and all of these are difficult and may not happen in a particular game, no matter how well designed. Thus, running the game multiple times (particularly incorporating feedback about issues and concerns from past games) may increase the chances of a lightning strike of a good idea. Of course, anyone who has worked with a rare event model will tell you that it is all too possible to invest in multiple games, and never have the needed lucky break—more runs of the game may increase the chances of innovation, but they do not guarantee it.

Evaluation

As the earlier discussion of the relationship between evaluation games and philosophies of inquiry suggest, there are several possible analogies to think through when considering the value of multiple evaluation games. The most common approach within the practitioner literature is to treat evaluation games like classic operations models and simulation—but the analogy holds to formal models in academic international relations. Here, repetition is used to better understand the central tendencies of outcomes in the model, much as in the experimental analogy described for alternative conditions games. However, in contrast to a computerized model, where actor behavior is usually held consistently (that is, the plan in enacted the same way in each run of the simulation) the benefit of a game is that human players respond to the results they receive to refine and adjust their strategy. On one hand, this likely makes for a more credible representation of actual behavior, since real decisionmakers will have similar opportunities to adapt. On the other hand, this can create challenges for an analyst hoping to discover and present central tendencies using statistical frameworks—put simply the number of iterations needed is simply too large to be practical if players are given free choice. Furthermore, the same practical limitations that hamper multiple iterations, particularly access to a sufficiently-sized player base that still is drawn from the population of interest can be difficult. As a result, insights from across evaluation games are likely to be either fairly broad or fairly tentative.

---

[40] Lin-Greenberg, Pauly, and Schneider, "Wargaming for Political Science Research."

The Value of Games in Multi-Method Research

Both practitioner policy analysts and political science share a belief that research that leverages multiple tools and approaches generates more credible results. However, to date the practitioner community has under-articulated both the value of such research plans and how they can best be constructed. Considering games in light of the more developed political science literature on multi-method research offers richer advice for how games can be leveraged in such studies.

Generally, when defense policy practitioners discuss using gaming in conjunction with other tools, they refer to the "cycle of research,"[41] illustrated in Figure 1, which argues that games should be linked to analysis and exercises. Though Perla makes it quite clear that the cycle of research can move in different sequences depending on the research questions at hand, discussion is dominated by a specific sequence exemplified by the historical example of naval innovation in the interwar years.[42] This canonical process argues that gaming should be used at the start of research for exploration and idea generation, followed by operations research analysis, most often campaign analysis using computerized modeling and simulation, to refine the approach with higher precision analysis, followed by field exercises to test how the ideas work in practice. The dominance of this model has not only calcified this particular sequence of tools, it also sets expectations that efforts to incorporate multiple tools require multi-year projects that involve multiple analytic offices—a scale simply not achievable for most analytic efforts. Thus, while the "cycle of research" is held up as an ideal, existing texts do not support researchers in thinking critically about what tools will combine to produce more credible findings, at a realistic scale.[43]

This gap is particularly ironic, because developing and analyzing a game often requires that a researcher leverage multiple means of gathering and analyzing evidence. All games, regardless of type, require research in the pre-play stages in order to develop a model of the policy problem for players to interact with. Interviews, comparative case studies, and formal modeling are all common tools used to develop the game. Data generated by the game must also be analyzed, which can take a wide range of forms including but not limited to qualitative coding, process tracing, text analysis, social network analysis, and regressions, depending on the information the game generates. How such approaches are selected and connected to one another is rarely discussed, and in practice is often a matter of the skills and data collection tools known by the project team. Again, a general, systematic framework is lacking.

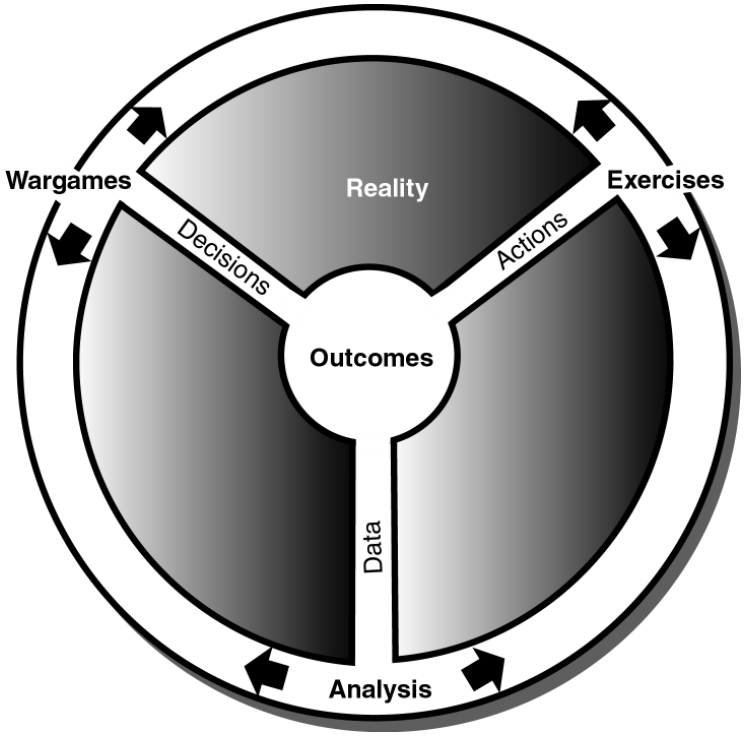
---

[41] Perla, *The Art of War Gaming: A Guide for Professionals and Hobbyists*; Peter Perla et al., "Rolling the Iron Dice: From Analytical Wargaming to the Cycle of Research," *War on the Rocks*, October 21, 2019 2019; Phillip E Pournelle, "Can the Cycle of Research Save American Military Strategy?," ibid.

[42] For a high level example of how this model is evoked, see: Robert Work and Paul Selva, "Revitalizing Wargaming Is Necessary to Be Prepared for Future Wars," ibid., December 8 2015.

[43] Jon Compton, "The Obstacles on the Road to Better Analytical Wargaming," ibid.2019.

Figure 1: Perla's Cycle of Research



Source: Peter P. Perla, *The Art of War Gaming: A Guide for Professionals and Hobbyists*, ed. John Curry, 2nd ed. (History of Wargaming Project, 2011).

In contrast, political scientists have spent considerably effort conceptualizing the value and practice of multi-method projects. Critically, recent work shifts attention from the number or kind of tools used to the question that is being asked at each stage in research. Gary Goertz offers a particular clear division of different approaches into three broad categories, illustrated in Figure 2: tools to articulate causal mechanisms, tools to explore a specific empirical incident in depth, and tools to explore cause and effect across many cases to understand general behavior. The first group of approaches focus on formal modeling—that is conceptual or mathematical arguments that lay out a concept of how a cause achieves a particular effect. The second is exemplified by case study research that deeply considers a single empirical case to advance an argument about what happened and why. Third are tools that look across multiple cases to establish trends about general behavior. In Goertz's view, work that tries to balance these three perspectives in invested in a multi-method approach.[44]

---

[44] Gary Goertz, *Multimethod Research, Causal Mechanisms, and Case Studies: An Integrated Approach* (Princeton, NJ: Princeton University Press, 2017). pp 1-5

Figure 2: Goertz's Research Triad

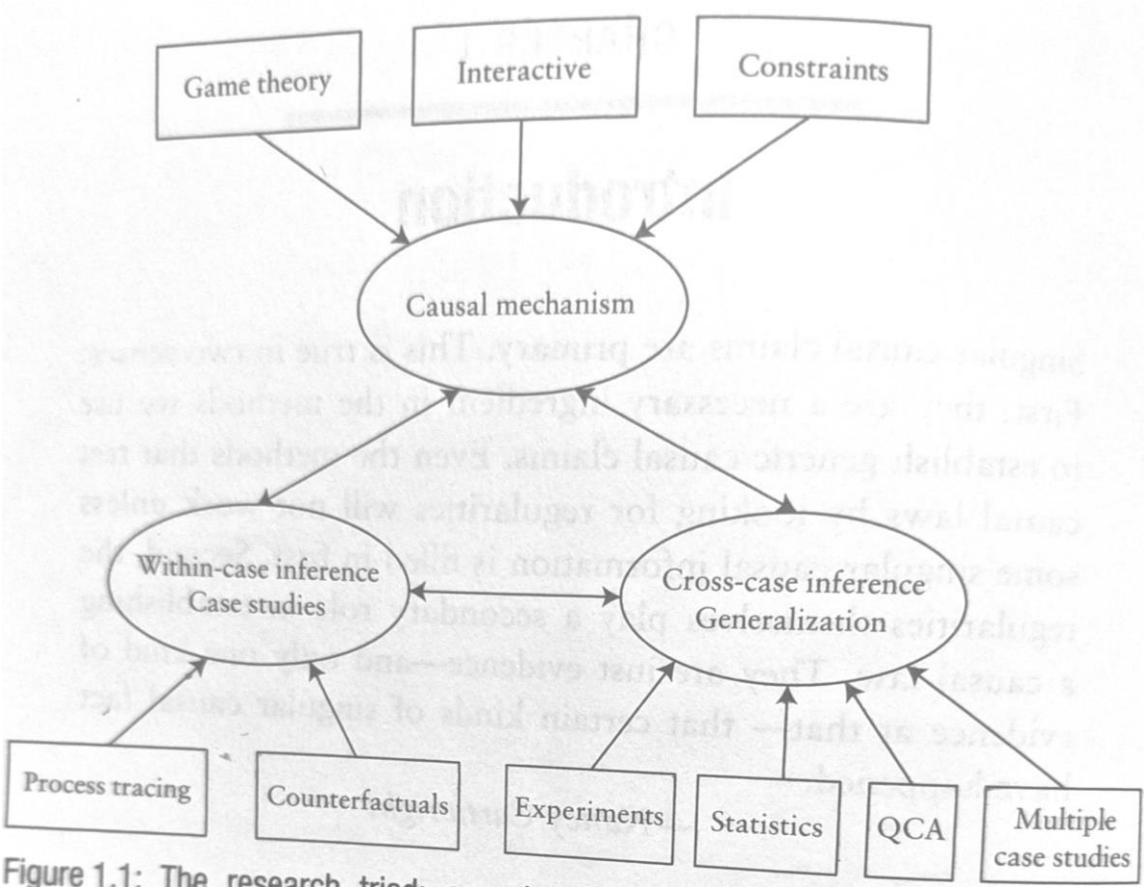


Figure 1.1: The research triad

Source: Gary Goertz, *Multimethod Research, Causal Mechanisms, and Case Studies: An Integrated Approach* (Princeton, NJ: Princeton University Press, 2017 p 2

This approach is helpful in that it moves the debate away from specific historical examples or existing skill sets towards fundamental questions about what types of work the study aims to do. Studies that only need to propose a new model, explain a single case, or describe cross cutting trends need not venture into multi-method work. Projects that seek to answer questions that require elements of all of these will require a broader approach.

Within social science work, there are a few typical ways of leveraging the triad. The first is the use of broad cross-case analysis to identify general trends, followed by deeper analysis of a specific case identified from the broader analysis that is used to generate a causal mechanism model. Another common approach is the generation of a causal model from a single case, which is then expanded into cross-case studies to ensure that the causal mechanism is, in fact, widespread. Both general pathways can be populated by different collections of tools, but all are designed to deal with causal arguments that are more complex then X treatment generates Y effect. This can include causal narrative in which a series of factors must be present one after another and in which different factors interact with one another in key ways. Given the

complexity and contingencies we tend to associate with group decisionmaking about complex policy decisions, such multi-factor causal arguments are more likely than not to be occurring in policy contexts.[45]

So where do games fit into this triad? Comparing the tools and approaches analogized to gaming earlier in this article to Goetz's triad at even a superficial level suggest that there is more than one possible purpose for gaming, which means that not all games will serve the same function in the triad. The follow sections discuss the ability of games to function in each of the three roles of multi-method research. In cases where more then one type of game is used, we tend to describe the study as using a series of games, whereas when games are paired with other approaches, then the label multi-method is more likely to be used. However, fundamentally the process of selecting tools that are well suited to each task of the triad is the same regardless of what portion of the selected approaches turn out to be games—as a result, no distinction is made in the discussion below.

One characteristic of games that will become particularly salient in this discussion is whether games are empirical evidence or a type of model. Fundamentally, games sit in an unusual position in this debate, with characteristics of both empirical and formal studies. On one hand the game structure as created by the designer is a formal model, feature both ersatz history (that is, fictitious scenarios) and people (in the form of role-playing players).[46] On the other hand, games feature players communicating and making decisions which are observed and analyzed by the researcher[47]—while these observations occur in an artificial environment, the same is true of many laboratory experiments which are treated as empirical without comment. Thus, at some level the determination about whether game data is formal or empirical depends on the specific data being collected—observations of player actions are more likely to be treated as empirical, whereas data about the behavior of the game system as a whole, particularly adjudicated outcomes, is more likely formal. However, the philosophy of inquiry held by the researcher is also likely to shape how they position game findings, with analyticist being more likely to treat games as formal models and positivists and critical realists more likely to argue for an empirical view.[48]

Games as Causal Mechanism

The first leg of the triad focuses on laying out a model of a specific causal mechanism. Fundamentally, the task can be thought of as drawing figures showing the links between

---

[45] Ibid. pp 6-13

[46] Levine, Schelling, and Jones, "Crisis Games 27 Years Later : Plus C'est Deja Vu." pp 3-12

[47] For a carefully argued example of this approach, see: Peter Perla, Michael Markowitz, and Christopher Weuve, "Game-Based Experimentation for Research in Command and Control and Shared Situational Awareness," (Alexandria, VA: CNA, 2005).

[48] Bartels, "Building Better Games for National Security Policy Analysis: Towards a Social Scientific Understanding." pp 55-56

independent variables, mechanisms, and dependent variables.[49] One particularly common approach is to describe mechanisms as pathways from cause to effect,[50] recognizing that most social processes are sufficiently complex that they will generally require more than one step[51] and may consist of interacting necessary and sufficient elements.[52]

When used in this leg of the triad, games can be thought of as an approach to build and document a mechanistic model. This aligns particularly well with system exploration games run under an analyticist framework, since this mode of games seeks to produce models that describe the elements and relationships between different parts of a policy problem. The resulting model meets the bar for usefulness so long as it can clearly articulate mechanisms, but the game does not provide an empirical check on findings or any generalizability without being linked to work in the other two corners of the triad.

In this approach, games are most often used early in the study design, then followed by other approaches to provide evidence about whether behavior seen in the game is representative of real-world behavior, or if the artificialities of the game have suggested a causal mechanism that seems sound in theory but is not evidenced in practice. Games make tradeoffs that simplify many aspects of the world in order to illuminate others. Players are often not the same as decision-makers, aspects of the environment are not included in the game's scenario, and layers of bureaucracy are stripped away. In other words, the game cannot simulate everything about the real world and these simplifications could change the findings. As a result, it is critical to ask how simplifications affected the key results of the game, and how the mechanism might be complicated once applied in real-world contexts.

For a researcher who believes that games cannot contribute empirical evidence, the most frequent strategy will be to turn to non-gaming tools. However, since games are frequently used when compelling empirical observation from past or present events is not easily available, gather evidence to populate the other two legs of the triad may not be trivial. One way to conceptualize the value of games in such combinations is that using a game to develop robust causal mechanisms can help make the most of the limited available data. For example, games have been used to fill in gaps in historical accounts of battles[53]—taking the empirical case as a starting point, the game is developed to posit mechanisms that logically connect the limited data that is available. Alternatively, causal mechanism games may be used first to develop a robust theory that can help identify reasonable proxies for unobserved (or unobservable) phenomenon. No

---

[49] Goertz, *Multimethod Research, Causal Mechanisms, and Case Studies: An Integrated Approach*. p 33

[50] John Gerring, "The Mechanism Worldview: Thinking inside the Box," *British Journal of Political Science* 388 (2008).p 178

[51] Goertz, *Multimethod Research, Causal Mechanisms, and Case Studies: An Integrated Approach*. p 36

[52] Ibid. p 42

[53] Philip A. G. Sabin, *Lost Battels: Reconstructing the Great Clashes of the Ancient World* (London, UK: Bloomsbury Academic, 2015).

empirical evidence will be available about how a war will be fought 10 years in the future, but clarifying key capabilities needed to see an outcome may help focus attention on studying specific capabilities and doctrine whose current development will shape future conflict. Put simply, the game's exploration of causal mechanisms becomes an aid for how to make the most of limited empirical data in the other two corners of the triad.

Even in studies that opt to use games as empirical evidence in one of the other legs of the triad, often a causal mechanism-focused run of the game is conducted as a "playtest." This stage of research is generally discussed as an element of games design best practices,[54] rather than being treated as an key element of research in its own right, but the process of playing the game with the goal of improving the game design is, in fact, an important effort to examine and improve the formal model of the game. This work ensures that necessary mechanisms are included in game play, potential alternative causal narratives are considered, and that extraneous elements are not included in the interest of parsimony. This work helps to ensure that the empirical data that is collected in later runs of the game is as salient as possible.

Games for Within-Case Inference

The second leg of the triad focuses on the use of a single case to provide an empirical evidence of a causal mechanism. Generally, selection of a single case focuses on a case that is believed to be a good example of the causal mechanism of interest,[55] which can allow for close study of how the posited mechanisms play out in practice. In older works on case study selection this is thought of as a "most likely" case for the phenomena of interest to occur.[56] In other words, the within case study selects one of the possible cases that would be studied in the cross case leg of the triad and uses it to refine the formal model developed in the third leg.

The use of games in this corner of the triad is effectively a type of artificial case study. Given the analytic preference and relative accessibility of traditional case studies, games are likely to only be used as cases when real-world observations are not available or are impractical. In effect, the researcher is using the game as an artificial case they are not able to observe in the real world. Such an approach requires that a researcher treat the game results as an empirical form of evidence, which tends to be more common under positivist or critical realist approaches. In terms of purpose for the game, systems exploration, innovation, and evaluation games might all apply, depending on the where the causal mechanism of interest sits. Likely the most frequent are innovation games, in which the innovative approach of the players is the causal mechanism of choice. System explorations games are somewhat less common because they are more likely to be aligned with approaches that view key results as formal, rather than empirical, however if a researcher can make a credible argument for empiricism, they may be used.

---

[54] Perla, *The Art of War Gaming: A Guide for Professionals and Hobbyists*. pp 218-219

[55] Goertz, *Multimethod Research, Causal Mechanisms, and Case Studies: An Integrated Approach*. pp 63-66

[56] George and Bennet, *Case Studies and Theory Development in the Social Sciences*.

Here, the within case game generally falls relatively early in study. In some cases, it may be the first tool used as a means of providing a foundation for later theory development. For example, a new research team might use a game to explore a novel problem, and then use that shared experience to develop a model using traditional formal approaches. Alternatively, theorizing may come first (either using a game as discussed above or another approach) which is then refined by collecting rich data about how key mechanisms play out in a game. Finally, while it is possible for cross-case comparison work to occur first, that is likely to occur only in cases in which there is sufficient data to allow a population of cases to be defined and typed, but not sufficient to enable historical case study analysis, making this approach relatively uncommon.

Games for Cross-Case Inference

The final leg of the triad explores how the causal mechanisms generalize across a range of cases. Fundamental to this task is the definition of the scope of cases where we would expect the causal mechanism to apply. This allows confirmation (or lack thereof) of the generalizability of the causal mechanism and requires consideration of cases that could falsify or complicate the posited mechanism by exploring confounding factors or adjusting scope.[57]

As with within-case inference, games for cross case inference operate as synthetic cases to produce empirical evidence but are expanded to multiple games to allow comparison. Alternative conditions games, by definition, fall into this approach, as can evaluations games if they are set up to explore sufficient variation. As discussed above, two general approaches to cross case comparison are seen with games—the first which focuses on a relatively small number of cases in sufficient depth to study variation in causal mechanisms, and a second which seeks to use a larger number of games to observe average differences in behavior over the population. Which approach is appropriate will likely depend on the nature of the question under study. More complex causal mechanisms will tend to require richer data to meaningfully explore.

While successful studies can certainly start with cross-case companions, this may be less likely when it comes to games. Running any sizable number of games is time consuming and expensive—doing so as a means of initially scoping the problem will not often be practical. In part, this is because the synthetic nature of games means that the designer is free to manufacture a very wide range of potential cases indeed—without the focus provided by previous research into causal mechanisms such efforts are likely to be scattershot. Far more common is the use of cross case comparison games after mechanisms have been formalized, at minimum through the use of a playtest game, and thus the selection of game cases can be minimized to the extent possible.

---

[57] Goertz, *Multimethod Research, Causal Mechanisms, and Case Studies: An Integrated Approach*. pp 60-61, pp 66-71

Conclusions

Policy researchers have long touted the benefits of multiple games and games combined with other methods, but historically have not articulated how repetition, series, and multi-methods approaches contribute to stronger findings beyond the invocation of historical examples. In contrast, discussion of research methods in international relations has devoted considerable time to research design but has not historically included games as part of this discussion. This paper brings the two communities into dialog, by exploring the range of games commonly discussed in policy analysts using the vocabulary and concepts from academic international relations. This crosswalk reveals that games can be analogized to multiple traditional research tools, depending on how they are used. While this prevents one-size fits all guidance about how games should be incorporated into broader studies, it illustrates the diversity of ways games can be used alongside more traditional methods. My hope is that this effort contributes to ongoing attempts to better leverage games by allowing scholars to construct more robust studies and communicate those efforts in ways that are more legible to the broader IR community.

References

Asal, Victor. "Playing Games with International Relations." *International Studies Perspectives* 6, no. 3 (2005): 359-73.

Baker, Sarah Elsie, and Rosalind Edwards. "Introduction." In *How Many Qualitative Interviews Is Enough? Exert Voices and Early Career Reflections on Sampling and Cases in Qualitative Research*, edited by Sarah Elsie Baker and Rosalind Edwards: National Centre of Research Methods, 2012.

Bartels, Elizabeth M. "Building Better Games for National Security Policy Analysis: Towards a Social Scientific Understanding." Pardee RAND Graduate School, 2020.

———. "Mors 83th Panel Aar: Typologies of Game Standards." In *Paxsims*, edited by Rex Brynen, 2015.

Barzashka, Ivanka. "Wargaming: How to Turn Vogue into Science." *Bulletin of the Atomic Scientists*, 2019.

Bogner, Alexander, Beate Littig, and Wolfgang Menz. "Generating Qualitative Data with Experts and Elites." In *The Sage Handbook of Qualitative Data Collecation*, edited by Uwe Flick, 652-65. London, UK: SAGE Publications Ltd, 2018.

Caffrey, Matthew B. *On Wargaming: How Wargames Have Shaped History and How They May Shape the Future.* Newport, RI: Naval War College Press, 2019.

Compton, Jon. "The Obstacles on the Road to Better Analytical Wargaming." *War on the Rocks*, 2019.

Elias, George Skaff, Richard Garfield, and K. Robert Gutschera. *Characteristics of Games.* Cambridge MA: MIT Press, 2012.

Geoffrey, Engelstein., and Issac Shalev. *Building Blocks of Tabletop Game Design: An Encyclopedia of Mechanisms.* Boca Raton, FL: CRC Press, 2019.

George, Alexander, and Andrew Bennet. *Case Studies and Theory Development in the Social Sciences.* Boston, MA: MIT Press, 2005.

Gerring, John. "The Mechanism Worldview: Thinking inside the Box." *British Journal of Political Science* 388 (2008): 161-79.

Goertz, Gary. *Multimethod Research, Causal Mechanisms, and Case Studies: An Integrated Approach.* Princeton, NJ: Princeton University Press, 2017.

Goldhamer, Herbert. "The Political Exercise: A Summary of the Social Science Division's Work in Political Gaming, with Special Reference to the Third Exercise July-August 1955." Santa Monica, CA: RAND Corporation, 1955.

Goldstein, Kenneth. "Getting in the Door: Sampling and Completing Elite Interviews." *PS: Political Science & Politics* 35, no. 4 (2002): 669-72.

Hyde, Susan D. "Experiments in International Relations: Lab, Survey, and Field." *Annual Review of Political Science* 18 (2015): 403-24.

Jackson, Patrick Thaddeus. "The Conduct of Inquiry in International Relations: Philosophy of Science and Its Implications for the Study of World Politics." New York, NY: Routledge, 2011.

Jensen, Benjamin, and Brandon Valeriano. "Cyber Escalation Dynamics: Results from War Game Experiments." In *International Studies Association*. Toronto, Canada, 2019.

Judge, Sawyer. "The Wargaming Guild: How the Nature of a Disipline Impacts Its Craft and Whether It Matters." Georgetown University, 2019.

Levine, Robert A., Thomas C. Schelling, and William M. Jones. "Crisis Games 27 Years Later : Plus C'est Deja Vu." Santa Monica, CA: RAND Corporation, 1991.

Lin-Greenberg, Erik. "(War)Game of Drones: Remote Warfighting Technology and Escalation Control Evidence from Wargames." SSRN, 2019.
Lin-Greenberg, Erik, Reid B.C. Pauly, and Jacquelyn Schneider. "Wargaming for Political Science Research." In *APAS Annual Conference*. Virtual, 2020.
Longley Brown, Graham. *Successful Professional Wargames: A Practitioner's Guide.* Edited by John Curry. The History of Wargaming Project, 2019.
Lynch, Julia F. "Aligning Sampling Strategies with Analytic Goals." In *Interview Research in Political Science*, edited by Layna Mosley, 31-44. Ithaca, NY: Cornell University Press, 2013.
Parson, Edward. "What Can You Learn from a Game?". In *Wise Choices: Decisions, Games, and Negotiations*, edited by Ralpj L. Keeney Richard J. Zeckhauser, James K. Sebenius, 233-52. Boston: Harvard Business School Press, 1996.
Pauly, Reid B.C. "Would U.S. Leaders Push the Button? Wargames and the Sources of Nuclear Restraint." *International Security* 43, no. 2 (2018): 151-92.
Perla, Peter , Michael Markowitz, and Christopher Weuve. "Game-Based Experimentation for Research in Command and Control and Shared Situational Awareness." Alexandria, VA: CNA, 2005.
Perla, Peter, Web Ewell, Christopher Ma, Justin Peachey, Jeremy Sepinsky, and Basil Tripsas. "Rolling the Iron Dice: From Analytical Wargaming to the Cycle of Research." *War on the Rocks*, October 21, 2019 2019.
Perla, Peter P. *The Art of War Gaming: A Guide for Professionals and Hobbyists.* Edited by John Curry. 2nd ed.: History of Wargaming Project, 2011.
Pournelle, Phillip E. "Can the Cycle of Research Save American Military Strategy?" *War on the Rocks*, 2019.
Ragin, Charles C. *The Comparitive Method: Moving Beyond Qualitative and Quantitiative Strategies.* Oakland, CA: University of California Press, 2014.
Reddie, Andrew W., Bethany L. Goldblum, Kiran Lakkaraju, Jason Reinhardt, Michael Nacht, and Laura Eipifanovskaya. "Next-Generation Wargames: Technology Enables New Research Designs, and More Data." *Science* 362, no. 6421 (2018): 1362-64.
Sabin, Philip A. G. *Lost Battels: Reconstructing the Great Clashes of the Ancient World.* London, UK: Bloomsbury Academic, 2015.
Schneider, Jacquelyn G. "Cyber Attacks on Critical Infrastructure: Insights from War Gaming." *War on the Rocks*, July 26 2017.
Tekinbas, Katie Salen, and Eric Zimmerman. *Rules of Play: Game Design Fundamentals.* Cambridge MA: MIT Press, 2003.
Weiner, Milton G. "An Introduction to War Games." Chap. 11 In *Lex Choix Economiques: Decisions Sequentielles Et Simulation*, edited by Pierre Rosenstiehl and Alain Ghouila-Houri. Paris: Dunod, 1960.
Work, Robert, and Paul Selva. "Revitalizing Wargaming Is Necessary to Be Prepared for Future Wars." *War on the Rocks*, December 8 2015.