“Games as Structured Comparisons: A discussion of methods”

Elizabeth M. Bartels, Pardee RAND Graduate School

Abstract:

One way of designing games for research is to treat them as structured comparisons that can be set up and analyzed in much the same way as case studies. Comparisons can be made between historical events and a game, or between multiple games, and each may be appropriate depending on the purpose. This flexibility can provide traction on problems for which it is not otherwise possible to gather data. However, because games are inherently artificial, care should be taken when drawing conclusions. Honest discussion of where and how comparisons are insufficient is critical to analyzing games responsibly.

When using national security games as a tool for research, the key question for the investigator is how to make choices about game design elements in order to design a game that has the best chance of answering the project's research questions. Existing texts on game design stress the importance of linking design to purpose, but offer very little advice on how to achieve this goal. The most-cited handbooks on the design of games identify a range of different options for each element of a game, and discuss the advantages and draw-backs of each.[1] These works also stress the importance of linking the choice of design elements to the purpose of the game.

However, when it comes to *how* to make the linkage, these texts are largely silent. Even the best-respected book on game design states: "There is no recipe for translating a game's objectives into its mechanics… ultimately the designer's talent dictates how and how well the translations from objectives to mechanics works"[2] In other words the core task of the designer is surprisingly under-theorized. This is particularly surprising because within policy communities games are frequently used to examine poorly-structured problems on short timeline. As a result, games often demand greater simplification and more judgment on the part of the game designer than is typical in other types of research, where more time is available to delve into the empirical evidence base.

In part, the absence of a strong research design literature is due to a long-running debate that has raged since at least the 1970s in the practitioner community about whether gaming is an artistic or scientific endeavor. Designers from a strong positivist tradition worry that because games are not repeatable, and thus results not replicable, games are not tractable to scientific analysis.[3] Critics of gaming suggest that this makes games unsuitable to inform policy analysis,[4] while advocates of games who share never the less are concerned about the merits of games as experiments argue that the utility of games come from artistic, rather than scientific, mechanisms.[5] This later perspective tends to see game design as an artistic process that benefits from the instincts of experienced designers. In this view, systematized logics for design at best

---

[1] For well-known examples of this approach, see: Shawn Burns, ed. *War Gamer's Handbook: A Guide for Professional War Gamers* Newport, RI: War Gaming Department, U.S. Naval War College; James F. Dunnigan, *Wargames Handbook: How to Play and Design Commercial and Professional Wargames*, 3rd ed. IUniverse, 2000; Matthew B. Caffrey, *On Wargaming: How Wargames Have Shaped History and How They May Shape the Future* Newport, RI: Naval War College Press, Forthcoming. and Peter Perla, *The Art of Wargame: a guide for professionals and hobbyists*, Annapolis MD: Naval Institute Press, 1990.

[2] Perla 1990.

[3] Igor S Mayer, "The Gaming of Policy and Politics of Gaming," *Simulation & Gaming* 40, no. 6, 2009.

[4] For examples of this argument, see: Garry D. Brewer, *Dealing with Complex Social Problems : The Potential of the "Decision Seminar."* Santa Monica, CA: RAND, 1972. and Peter DeLeon, *Scenario Designs : An Overview*, Santa Monica, CA: RAND, 1973.

[5] A commonly cited example of this perspective can be found in: Peter P. Perla and ED McGrady, "Why Wargaming Works," *Naval War College Review* 64, no. 3, 2011.

encourage "cookie cutter" design in which gamers replicate a few designs[6] and at worst attempt to turn games into something they fundamentally are not—tools for testing and prediction.[7] This perspective also tends to stress that the centrality of people, particularly players, to games adds immense value, but prevents them from being truly scientific.

At the same time, these concerns have generally marginalized the use of games within political science to serve as a tool for teaching rather than research. As a result, while there are many articles on how to leverage games in the classroom, there is far less discussion of the use of games as a tool for research. Generally, discussion is limited to policy-oriented researchers with strong ties to the U.S. government,[8] or scholars using historical games as a data source about U.S. policy decision-making processes.[9] Thus, political science has not explored the use of games for research to their full potential.

The reluctance to develop formal theory on the part of policy gamers, and the hesitancy by political science to treat games as a serious tool for research has created a gap, where policy researchers have not leveraged existing best practices for research design to strengthen the design of games used in policy research. In this paper, I argue that there is a class of games that shares a similar logic of inquiry with structured comparison techniques. Games are often asking questions that are fundamentally comparative, such as: "Does conflict over water manifest differently in different environments? "How does the presence or absence of a particular stakeholder shape deliberations and decisions?" "How does using military strategy A vs strategy B shape the conflict?" or "Does theory A or B better align with practitioner understandings of particular phenomenon?" In these cases, I argue that it is appropriate to leverage the logic of structured comparison that has been articulated in the case study design literature to think through key design choices.

This paper proceeds in four sections. First, I review some of the fundamentals of applying games to national security policy issues, including defining key game design terms which will be used in the rest of the paper. I then describe a class of game purposes that I believe to be tractable to the logic of structured comparison. Third, I discuss the applicability of best practices from the case study literature to game design and analysis. Finally, I discuss the limitations of gaming that should caveat structured comparative analysis.

---

[6] Peter P Perla, "Now Hear This—Improving Wargaming Is Worthwhile—and Smart," *Proceedings Magazine*, January 2016 2016.

[7] Robert A. Levine, Thomas C. Schelling, and William M. Jones, *Crisis Games 27 Years Later : Plus C'est Deja Vu* Santa Monica, CA: RAND, 1991, p 2

[8] Lincoln P. Bloomfield, "Playing games with foreign policy," *United States Naval Institute Proceedings* 129 (7), 2003, pp 68-72.

[9] For example of recent research in this vein, see: Reid B.C. Pauly, "Elite Aversion to the Use of Nuclear Weapons: Evidence from Wargames," *International Studies*, forthcoming.

Fundamentals of National Security Policy Gaming

Fundamentally, games involve humans making decisions in contested environments, then living with the outcomes. They range in form from people sitting around a table reviewing slides in an activity that is on its face value indistinguishable from a meeting, to an individual sitting at a computer, to hundreds of players divided into small groups running back and forth in negotiations, to groups standing around a map with military unit counters and rolling dice. This diversity of form is one of gaming's great strengths, but it has made it difficult to describe appropriate and inappropriate uses and common standards for design that validate the application of these activities to national security policy.

In large part, this diversity is because national security policy games are put to a wide range of purposes.[10] Historically, early highly abstract games like set, chess, and go were used for elite education. Over time, military leaders realized the value of planning for potential future campaigns then testing these plans through games to lead to better operational outcomes. Through the late 19th and early 20th century, both educational and exploratory applications of gaming grew and were used to consider issues ranging from adversary behavior to emerging technology on the battlefield.[11] In the years following World War II, gaming also was embraced by the growing analytical study of national security issues for which games were used as a way to synthesize expert knowledge and stimulate new ideas.[12] Non-military application also grew over the latter half of the 20th century requiring the exploration of political, economic and other non-military tools. All of these applications of games are still practiced today with audiences ranging from students interested in careers in national security to senior leaders.

With few exceptions, games have traditionally been defined either by the audience they draw on (ex. senior leader seminar), the medium of the game (ex. hex and counter or computerized games), or the method of adjudication (ex. matrix games). While these terms can be helpful in setting participant expectations and describing the game after it is run, they are less useful, and sometimes actively problematic, in early stages of game design. For example, setting the audience, medium, and means of adjudication too early can overly constrain designer's choices, preventing them from developing the most appropriate game possible.

More recent work has offered alternative framings. Some work has sought to lay out types based on different uses for games, distinguishing between experiments, instruction, creative

---

[10] This account leaves aside the parallel evolution of games as a recreational activity. For a good discussion of the interrelated development of hobby and "professional" games, see Caffrery, forthcoming.

[11] Caffrery, forthcoming.

[12] For early examples, see: Levine, Schelling and Jones, 1991, Alexander McFarlane Mood, *War Gaming as a Technique of Analysis*, ed. RAND Corporation, Santa Monica, CA: RAND, 1954; Milton G. Weiner, "Trends in Military Gaming," Santa Monica, CA: RAND, 1969.

thinking, and integrating knowledge.[13] For example, one approach types games based on how structured our understanding of the problem is and whether the designers' primary goal is to instruct or learn from the game.[14] Other approaches to categorization focus more on the military problem to be solved by the game such as concept development, capabilities development, science and technology, senior leader engagement, operational decisions, and training and education.[15] However, these typologies have been sources of debate in the field, as different practitioners have not been able to align on which distinctions are useful and where they should be drawn. As a result, there is still little alignment about how best to describe different games to fellow designers, sponsors, and game consumers in consistently informative ways that reflect the design process.

National security policy games vary widely in form to meet this wide range of purposes but generally have four common elements that are manipulated as part of the game design: environment, actors, rules, and adjudication models.[16] While these elements are not fundamentally different from the components of many models of national security problems, the critical role of human decision making in games creates specific opportunities and limitations for the designer.

The first element is an environment or setting which describes a policy problem, conflict, or crisis.[17] This can be as simple as a short, written description or briefing describing a policy problem to be addressed, or as complicated as a model of the physical, social, political, and economic environment that feeds a computerized system for situational awareness. These choices often shape (and are shaped by) the physical environment of the game, ranging from the room the game is played in, to the visualizations and resources available to players.

---

[13] Edward A. Parson, "What Can You Learn from a Game?" in Richard J. Zeckhauser, Ralph L. Keeney, and James Sebenius eds., *Wise Choices: Decisions, Games, and Negotiations*, Boston, MA: Harvard Business School Press, 1996.

[14] Elizabeth M Bartels, "Gaming – Learning at Play," *ORMS Today* 41, no. 4 (2014).

[15] Yuna Huh Wong et al., "Next Generation Wargaming for the U.S. Marine Corps: Recommended Courses of Action," Santa Monica, CA Forthcoming. pp 5-9

[16] Other designers present somewhat different conceptualizations of these categories, such as objective, scenario, database, models, rules, players and analysts (Perla, 1990).) or objectives, scenario, database, models, rules and procedures, infrastructure, participants, analysis, culture and environment, and audiences (Christopher A. Weuve et al., "Wargame Pathologies," (Arlington, VA: CNA, 2004).) I combine many of these categories together because I believe it better allows me to discuss the interactions and interdependent decisions between them. For example, while a games scenario and data bases may be different products, from a design perspective decisions such as the level of analysis or key assumptions should be made consistently across these elements, so I combine them together under the broader category of "environment".

[17] There is a long debate about how much conflict must be present in a game. While a faction believes that war is the only true subject for a game, others focus on conflicts ranging from struggles to contain natural events (such as pandemics) to debates and divisions between different branches of the same bureaucracy. I belong to the later school of thought—games must present a problem to be solved, and are most interesting where there are different goals and objectives in play, but do not require strict competition or warfare between two or more sides to be dynamic and of interest to the national security community.

The second element is actors with different goals, resources, and abilities.[18] Actors can be represented by players in ways ranging from recruiting players to represent their own position to providing players with substantial role descriptions detailing objectives, red lines, and capabilities. In many games, some relevant but secondary actors are controlled by the game's controllers, or are incorporated into the adjudication model.

The third element is actions available to each actor—that is the different ways they can influence other actors or the environment in order to achieve their objectives. These generally consist of both the types of decisions the actor can make, and the process by which they make them. In some cases, such as a computerized model interface, these will be highly prescripted by the designer. In other, less structured games, they may be defined primarily by the flow of conversation between the facilitator and participant as an organic part of the discussion.

Finally, the fourth element is a model of the consequences of different player decisions for actors and the environment. The ability to play out the results of decisions in the future is key to the narrative aspect of successful games. This can be implemented based on expert judgment or formalized in set probabilistic or deterministic adjudication models.

By manipulating how these four elements are represented in the game, designers have a great deal of flexibility in what games look like, and what objectives they can achieve. Designers often face tradeoffs, both within the design of each element of the game, and between different elements. Good design makes these tradeoffs in thoughtful ways that maximize the usefulness of the game for its purpose and transparently account for potential limitations. Done well, the elements interact synergistically to create a compelling intellectual, physical and emotional space for problem solving.

However, other analysists argue that games share many characteristics of other data sources used in social science, and thus are amenable to scientific logics of design.[19] If we can specify sound, useful models in a range of other traditions, why should games not be held to the same type of standards? Put differently, the design process can be governed by a systematic method, and thus be rigorous, even if their limitations prevent us from making scientific claims of causation. In discussions between advocates of both positions, the concept that different game purposes require different design standards has been suggested as a possible middle ground

[18] There is a long-standing debate among gamers about whether games must have two opposing sides played by human players to count as a game. I take the position that one-sided games can, in fact, meet this criteria so long as the team must wrestle with competing objectives. Within a one-sided game, these tensions could come from a range of sources such as differing bureaucratic perspectives within a single government, or from a problem where the actor has multiple objectives they seek to achieve that are in tension with one another.

[19] See for example Yuna Wong, "Preparting for Contemporary Analytic Challenges," *Phalanx* 47, no. 4, 2014. and Elizabeth M. Bartels, Margaret McCown, and Timothy Wilkie, "Designing Peace and Conflict Exercises: Level of Analysis, Scenario, and Role Specification," *Simulation & Gaming* 44, no. 1, 2013.

between the positions,[20] but this concept has yet to be fleshed out in the practitioner literature.[21] This paper is an initial attempt to define one such type of game, and define (or more accurately, import) standards for the design and analysis of this type of games.

Games as Structured Comparisons

One type of game that seems evident, when surveying gaming practice is games that seek to set up a comparison between runs of the game or between the game and real-world events. While a systematic review of policy games is not feasible as the publicly available records are highly incomplete,[22] reviewing available records provides numerous examples of games that depend on comparison to supply evidence in support of analytical claims. Particularly common are games where two or more versions are run allowing comparison between games. For example, a 2010 game run by the U.S. National Defense University looked at how a crisis sparked by water scarcity in three different geographic contexts might provoke similar or different policy responses from U.S. decision-makers.[23] Another example of game-to-game comparisons is the U.S. Marine Corps' 2013 Expeditionary Warrior, which featured three teams of military officers developing strategies to fight the same military conflict with the same capabilities, but different organizational structures.[24] A third game from the same period with a similar setup was the Naval War College's Final Destination 2, in which two groups of players of similar players were given the same scenario, but one team received more access to planning information then the other.[25] More rarely, games have been compared to real world events. For example, in the 1960s RAND ran a series of games that asked players to make choices about future strategic posture. A decade later, the results of these games could be compared to the historical decisions of actual US military planners.[26] Without making any claims that these games are representative, they provide examples of some of the ways that comparative design has been practices by different designers tackling different national security problems.

---

[20] Elizabeth M Bartels to Paxsims, 2015.

[21] In part, this is simply because many game designers work as practitioners in industry or government, whether methodological publication is not incentivized. This paper is part of a larger project that seeks to develop a more complete set of types of games by codifying practitioner's activities in a more formal, and thus accessible format.

[22] Sources of omission include formal restrictions, such as the protection of government (i.e. classified or otherwise sensitive data) or corporate data (i.e. intellectual property or trade secrets), as well as informal barriers like the limited professional rewards for publication by practitioners.

[23] Bartels, McCown, and Wilkie, pp. 40-41

[24] United States Marine Corps, *Expeditionary Warrior 2013: Future Maritime Operations for the 21st Century Operating Environment Final Report*, June 6, 2013. Available as of March 31, 2018: http://www.mcwl.marines.mil/Portals/34/Documents/EW13%20Final%20Report_FINAL.pdf.

[25] Hank J. Brightman and Melissa K. Dewey, "Trends in Modern War Gaming: The Art of Conversation," *Naval War College Review*, Winter 2014, p 25-26.

[26] T.A. Brown and E. W. Paxson, *A Retrospective Look at Some Strategy and Force Evaluation Games*, R-1619-PR, Santa Monica, CA: RAND Corporation, September 1975

In the existing gaming literature, games with this type of structure are often discussed either as experiments or as creative thinking exercises to generate new ideas,[27] however, neither frame seems to capture the logic underpinning these games well. In the case of games as experiments, many experienced gamers worry that treating games as experiments is fundamentally problematic, as games cannot be repeated due to variation in players, and cannot fully replicate the complexities of the real world, limiting the validity of applying game results to real world phenomenon.[28] As a result, gamers argue that an experimental logic is incompatible with the gaming as a method. On the other hand, attempts to articulate a logic to guide gaming as a creative approach are profoundly underdeveloped, and fail to capture the comparative logic that we see underpinning gaming practice. In short, there seems to be something of a goldilocks problem: the logic of experiments is too ridged to fit gaming well, while framing design around creative thinking undersells the structure that is evident in how games are designed and analyzed in practice.

Instead, the more apt analogy for the design of these games seems to be approaches to structured comparison from the case study literature.[29] Rather than resting on replication, case studies seek to trace patterns in rich data over time, generating a deeper understanding of a small number of examples. However, they do so in a focused, structured way that provides a rigor that might not be present without a clear set of norms. By asking consistent questions of multiple games, or of games and real-world cases, we can leverage both the evolution of the phenomenon of interest over time (that is, within-case variation) and create clear comparisons between multiple cases (cross-case or between variation). These patterns of similarity and different then form the bases for an explanation of the relationship or mechanism that is generating the pattern as a means of theory generations or testing.[30]

Games that have a comparative structure seem to follow much of this underlaying logic, even if it may not be explicitly articulated in game documentation. In the cases of games that are designed to be compared to one another, designers generally opt to systematically vary a key variable of interest, while trying to minimize other differences in starting conditions. The process and outcomes of decisions are then compared between the different cases to identify similarities and differences, allowing researchers to posit an explanation for the differences. In the case of comparisons between games and real life, variation between the game and real-world events may be similarly intentional (for example, one could imagine a counterfactual game that intentionally

---

[27] For a frequently cited example, see Parson, 1996

[28] For examples of this perspective, see Bartels 2015 and Parson 1996.

[29] Except when noted, I relay on Alexander George and Andrew Bennet, *Case Studies and Theory Development in the Social Sciences*, Boston, MA: MIT Press, 2005 and Matthew Lange, *Comparative-Historical Methods*, Los Angeles, CA: SAGE, 2013 for their treatment of case study methods.

[30] Timothy J. McKeown, "Case Studies and the Limits of the Quantitative Worldview," in Henry E. Brady and David Collier eds. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, Oxford, UK: Rowman & Littlefield Publishers, 2004.

changes a key variable from an historical case) or exploited variation between games that predicted a future and the actual subsequent events.[31] Regardless of whether the variation is intentional or not, the fundamental approach to comparing similarities and differences in the process and outcome of decisions based on the starting conditions remains consistent.

Beyond the similar logics of inquiry, treating games as comparative case studies makes sense because of the characteristics of the data that can be generated by games. Games have long been praised for their ability to tackle complex problems, featuring the interactions of diverse decision-makers on complicated environments, where many aspects of the relationship are under-theorized.[32] Less often discussed is a corollary of this—most games produce rich data by generating records of discussions and written interactions of the different players, as well as the actual decisions made by players. While data capture protocols vary widely, games can be instrumented to produce a wide range of different data. For example, as single game could produce individual level data (including player demographics or surveys on attitudes and opinions before, during, and after play), records of group discussions (typical notes or transcripts of discussion), and observations by researchers (including records of affect, interpretations of experts with different specialties). Data capture occurs throughout a game, which typically lasts hours or even days, providing time for events in the game world, and player relationships and perceptions to shift and change. Critically, the plan for data capture should be developed early in the process in a way that is tailored to the specific research question, and be refined as research progresses in order to take advantage of unexpected observations. Depending on the specific research question, data can also be captured before and after the game about individual and organizational responses to the game. In short, the same approaches to collecting detailed information in a systematic way that are advocated in case research also apply to games that seek to enable comparison.

In a similar way, the analysis of the information collected during games can also mirror techniques from case study analysis. First, designers report what happened in each game, in effect offering a causal narrative about how the starting conditions of the game shaped player discussion and eventual decisions. Second, the causal narratives can be compared to highlight similarities and differences across games, or between the game and real-world events. For example, games that features intentionally varied starting conditions would want to offer a narrative of how (if at all) differences in the outcome of the game can be explained by the differences in starting conditions. In some cases, designers may have a hypothesis about what types of variation they expect to see, in which case the evidence of the game may support or undermine the hypothesis. Alternatively (and perhaps more commonly) designers may not have a strong intuition about what effect variation may have, and may be looking at the pattern of outcomes in order to develop a new hypothesis. Either way, these claims will only be as strong

[31] Brown and Paxson, p v.

[32] For a particularly cogent description of these characteristics, see Parson 1996.

as the pattern the analyst is able to develop through the narrative of each game or real-world case.

Of course, it is critical to recognize that in a key way, games differ from traditional comparative case studies: Games are synthetic environments. This creates several key concerns, that will be discussed in detail in later sections. First, the generalizability of results may be more concerning, as one has to worry not only about moving to a different real-world context (as is true for most case studies) but also from the artificially simplified world of the game to reality. Thus, game designers need to pay considerable attention to discussion of the representativeness in their design for results to be credibly transferred. Second, the synthetic environment of the game requires a set of designers to make choices, which introduces different opportunities for biases to impact finding. This is particularly true of games where the designer is also the lead analyst. Third, games are not only constructed by the designer, but also by participants, who's beliefs and decisions shape the decision-making process and outcomes in key ways. Thus, the analyst not only has to consider the designer, but also the players, when investigating the representativeness of the game.

It is also important to note that not all games do, or should, use this logic of design. Games are run for a wide variety of purposes, including model development, evaluation, and education, not all of which will be well served by structured comparison. As a result, I do not advocate for applying this logic to all games, but rather suggest that when structured comparison is underpinning the game's logic, it would be helpful to use this framing more explicitly. In the next section, I discuss specific best practices that can be applied in these cases.

Best Practices for the Design and Analysis of Structured Comparison Games

Given that some games are tractable to the logic of structured comparison, it is natural to look to the rich case study literature for best practices that can be imported to help improve the design and analysis of games for research. In other words, given the goal of developing insights either by comparing games, or comparing games and real-world events, what choices improve the credibility of findings? George and Bennet describe five key tasks in the design of case study research which provide a helpful template: 1) specify the problem and research question, 2) specification of variables, 3) case selection, 4) describe the variance in variables, and 5) formulate data requirements.[33] While all are critical to successful research, the first and last are somewhat less specific to the comparative project, and thus I do not discuss them in any depth. In the case study context, specification of variables involves defining the key dependent, intervening, independent, and constant variables to be studied. This is done by defining an initial hypothesis about the potential relationship of how the independent variable might affect the dependent variable, through the mechanism of intervening variables. These will be the key

[33] George and Bennet, Chapter 4.

sources of variation in the study, as generally, we want to vary the independent variables to observe how it relates to differences in the dependent variable.

Because games are centered on decision-making, the dependent variable will generally relate to deliberative process or decision-making. In part, this is the happy intersection of the types of question asked by political science, which focus on the use of power by actors and institutions, and the structure of games. Because so much of the starting conditions and adjudication process is defined by the designer of the game, when conducting research, we focus on the choices made by players as the primary empirical observation. As a result, where the researcher is designing and running a game, it generally makes sense to use the starting conditions to instantiate variation in the independent variable of interest, then observe the decisions of player as the outcome.

The greater degree of researcher control over starting conditions, and relative inability to guarantee variation in player decision also tends to make "most similar" comparative designs most attractive. In a most similar design, cases are selected to be as similar as possible except for variation in the independent variable, which is then used to explain any differences in the outcome variable.[34] Since a game designer will have a relatively strong ability to ensure variation in starting conditions, but will not be able to ensure a diversity of outcomes, a most similar approach increases the chances that the games will produce data that aligns with the research strategy. The down-side is that the strength of findings will depend to a large degree on the extent to which the analyst can argue that variations in the games beyond the independent variable do not offer an alternative explanation for any variation that occurs in the dependent variable.

In contrast, when a researcher is working with historical game data, they have the ability to ensure variation in decisions, and thus it becomes more feasible to use a decision as the independent variable, opening up more research strategies. For example, games could be selected that had similar outcomes despite quite different contexts following a "most different" logic. However, even in these cases, generally the outcome of interest will still be a decision—for example looking at escalation by or deterrence of opponents as a result of the use of particular types of military force. As a result of these structural factors, while there may be opportunities to develop alternative designs, as a general rule comparative research involving games will still focus on a "most similar" design with a deliberative process or decisions as the key outcome, and will use the starting conditions or another aspect of the game that can be varied by the designer as the independent variable of interest.

Given this formulation, we can then consider what types of independent variable might be examined through games. As described earlier in the paper, in games we often think about the primary design elements as consisting of the environment, actors, actions, and adjudication model. The first three are often defined in the starting conditions of the game, and thus are the

[34] George and Bennet, p 81

easiest to manipulate in a most similar design when we are trying to understand decision-making. In the case of the environment, this might be done by providing scenarios that describe differences in the context of the crisis. Depending on the research question, these might be quite narrow (for example, changing the number of initial casualties) or broad (as in the NDU water game, where conflicts with similar narratives were postulated in quite different political, social, economic, and geographic contexts). Similarly, the actors can be varied either by changing the identity, objectives, and resources available in the starting conditions, or through careful selection of the human players representing each role. Varying the actions available to players can allow for investigations of the impact of processes (such as communication protocols or deliberative protocols). Of these four elements, the adjudication model is somewhat less likely to be used as an independent variable in analysis, since generally competing models of the outcomes of decisions are more credibly evaluated. However, it still may be appropriate to consider, depending on the specific research question.

Unlike in traditional case research, a game-based study has the advantage of being about to generate hypothetical cases to meet the requirements of the research design. For example, a game designer can posit a war between particular actors, employing particular capabilities, even when no such empirical case exists. This is of course, one of the major advantages of games—they can provide a base of evidence about events that have not occurred. However, to do so credibly, it is generally necessary that the narrative of the game seem plausible to participants, and not all potential cases will meet this standard. The process of determining what synthetic cases both meet the criteria of the research design and are plausible can be helpful form of analysis in its own right.

However, this degree of control can be somewhat illusionary, as requirements imposed by both perceptions of plausibility, the logistical realities of running an event with human players make design choices about how to specify variance in the independent variable, and consistency in variables that could provide an alternative explanation for change in the dependent variable. In the case of the independent variable, this is mostly a matter of setting up the cases to be different enough that the provide a clear comparison, without pushing the bonds of plausibility. However, the considerations for how much control is necessary to allow for comparison is somewhat more varied by the element of the game design.

Actors are a particular concern because the role of human players makes it difficult to create controls between multiple groups. Common solutions are either to have the same group of players play multiple games, or to attempt to recruit multiple comparable groups of participants. In the first case, it is then necessary to account for learning between the rounds of play. For example, if the same group of players is presented with the same decisions context twice, we would expect that their decision would in part be influenced by the outcomes of their decisions in the first round of play. There is also the challenge of maintaining engagement and interest in multiple rounds of truly identical play, which often requires that variations be introduced. The second option is to try to recruit multiple groups of players. However, for the comparison to be

credible, the designer must be able to define what characteristics of the players are salient, and defend why the two groups are similar on these dimensions. This can be quite challenging, given the range of experience that may shape decision making, and the limited knowledge about player background that game designers may have available. Regardless of which option is selected, it is extremely unlikely that difference in player experience, attitudes, and beliefs will not play a role in shaping decision making. As a result, it is generally best to treat these factors as an alternative explanation for differences in decisions, and explicitly discuss why the patterns of discussion support that the independent variable, rather than participants, is driving differences in decisions.

Comparative control over the environment is somewhat more under the control of the designer, but requires carefully thinking though second order effects of variance to create coherent scenarios. Unless a game is looking at a quite small variation in independent variables, it is quite likely that variation in the environment, actors, or available actions will create other differences. For example, in games comparing two different sets of military capabilities, there may also be changes in the relative economic performance as a result of civilian uses of technology. The key here is to consider both what variation is demanded by plausibility, and what variation threatens to complicate the comparative narrative. As with actors, if variation between environments that is required for coherence could provide an alternative explanation, this should be addressed explicitly in analysis.

How much of a concern variation in available actions and the adjudication model might be will vary considerably based on how formalized the available actions and means of determining outcomes are. To take one extreme, if players can take any action that occurs to them, there may be a great deal of variation in the actions that are perceived to be available between different groups of players, and resolution of outcomes may differ considerably. If on the other hand, available actions and adjudication are highly formalized it is much easier to ensure consistency between rounds of play. However, the choice of how structured a game system to use is often dictated by the existing knowledge base about the problem. For example, there is a great deal of available information about the capabilities of existing weapons systems, and very little about how well potential future weapons systems might work. In this later class of cases, a more flexible option for what players can do, and what the impact of those actions are on other actors and the environment, may need to be left less specified, and any resulting variation accounted for in subsequent analysis.

What Can You Learn from a Structured Comparison Game?

Having now discussion the basic setup of a structured comparison game, I now consider the strengths and weakness of the approach. While many of the same caveats that have long been articulated in the case study literature apply to comparative games, the synthetic nature of games adds particular concerns about representativeness (and thus concept validity and generalizability) and additional concerns about bias. While these threats to validity are real, games also have specific advantages over traditional cases studies. Thus, I argue that while care is needed to

recognize potential limitations of the approach, games can provide a useful addition to a researcher's tool kit, and political science would benefit from broader use.

The most obvious limitation of games is that they take place in a synthetic environment. All games make choices about what aspects of the environment to abstract, and while these worlds will still be quite rich by the standards of many quantitative models, they are still restricted compared to the full range of data available in most cases. This poses a threat to one of the key benefits of traditional case studies: concept validity.[35] If some of the specificity of a phenomenon is lost in the design process, it is possible that key concepts will be simplified and stretched in unhelpful ways. While this is undoubtably a concern, games with experienced players have something of a built-in corrective, since if key concepts lack face validity with players, they will generally comment, or even stop game play to correct the concern. As a result, while designers should be sensitive to the risk of over-simplification and concept stretching, it is not a serious a concern as it may at first appear.

More serious is the effect of simplification on our ability to generalize findings from a game to a broader universe of empirical cases. The question of generalizability is often a prickly one for case study methods. Formal discussions of case study approaches carefully caveat that the deep focus on context strictly limits the extent to which findings can be transferred to other cases.[36] Games add to this the additional hurdle of translating findings from an artificial environment to the real world. Thus, modesty in any claim about the generalizability of results should be practiced. When claims are translated to broader real-world contexts, analysts of games should be prepared to discuss ways in which the representation of the game world may have deviated, and their potential effects on findings. For example, a frequent concern is that game players in games for research are generally not current officials, and may not represent their decision-making process well. Attempts to characterize the degree of difference are helpful here, as recent former officials might make for a more credible representation that college students. Beyond simple consideration of representativeness, findings may be more credible if the author can develop hypotheses about the potential direction of bias created by non-representative elements of the game. For example, an argument may be made that the synthetic environment may make individuals less risk adverse because they recognize costs are artificial, enabling more aggressive behavior. This type of non-representative behavior will pose more problems for findings dependent on aggressive play then findings that leaders look for off ramps. In short, attempts to anticipant and address potential short comings clearly will do much to mitigate these risks.

Finally, there may be a concern that the power of the game designer and participants to shape causality within the game provides opportunities for bias and non-representative elements to enter into the game and influence results. As with most academic work, the most important

---

[35] George and Bennet, p 19

[36] Lange, p 118

hedge against this risk is the integrity of participants and researchers in confronting their own biases. As with representativeness, there is also opportunities for designers and participants to identify, and perhaps balance, biases.

Despite these limitations, the obvious advantage of games over traditional forms of case study analysis is the ability to consider events for which there is limited real world evidence at the time the game is conducted. For example, games can be constructed to compare two hypothetical events, or as a means of constructing a hypothetical counterfactual with internal logic and consistency to real world events. This means that games can form a body of evidence that would not otherwise exist. While the restrictions noted above are real concerns, weak evidence is still preferable to no evidence.

References

Elizabeth M Bartels to Paxsims, 2015.

--- "Gaming – Learning at Play," *ORMS Today* 41, no. 4, 2014.

Elizabeth M. Bartels, Margaret McCown, and Timothy Wilkie, "Designing Peace and Conflict Exercises: Level of Analysis, Scenario, and Role Specification," *Simulation & Gaming* 44, no. 1, 2013.

Lincoln P. Bloomfield, "Playing games with foreign policy," *United States Naval Institute Proceedings* 129 (7), 2003, pp 68-72.

Garry D. Brewer, *Dealing with Complex Social Problems : The Potential of the "Decision Seminar."* Santa Monica, CA: RAND, 1972.

Hank J. Brightman and Melissa K. Dewey, "Trends in Modern War Gaming: The Art of Conversation," *Naval War College Review*, Winter 2014, p 25-26.

T.A. Brown and E. W. Paxson, *A Retrospective Look at Some Strategy and Force Evaluation Games*, R-1619-PR, Santa Monica, CA: RAND Corporation, September 1975

Shawn Burns, ed. *War Gamer's Handbook: A Guide for Professional War Gamers* Newport, RI: War Gaming Department, U.S. Naval War College.

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

Peter DeLeon, *Scenario Designs : An Overview*, Santa Monica, CA: RAND, 1973.

James F. Dunnigan, *Wargames Handbook: How to Play and Design Commercial and Professional Wargames*, 3rd ed. IUniverse, 2000

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

Matthew Lange, *Comparative-Historical Methods*, Los Angeles, CA: SAGE, 2013 for their treatment of case study methods.

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

Igor S Mayer, "The Gaming of Policy and Politics of Gaming," *Simulation & Gaming* 40, no. 6, 2009.

Timothy J. McKeown, "Case Studies and the Limits of the Quantitative Worldview," in Henry E. Brady and David Collier eds. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, Oxford, UK: Rowman & Littlefield Publishers, 2004.

Alexander McFarlane Mood, *War Gaming as a Technique of Analysis*, ed. RAND Corporation, Santa Monica, CA: RAND, 1954

Edward A. Parson, "What Can You Learn from a Game?" in Richard J. Zeckhauser, Ralph L. Keeney, and James Sebenius eds., *Wise Choices: Decisions, Games, and Negotiations*, Boston, MA: Harvard Business School Press, 1996.

Reid B.C. Pauly, "Elite Aversion to the Use of Nuclear Weapons: Evidence from Wargames," *International Studies*, forthcoming

Peter Perla, *The Art of Wargame: a guide for professionals and hobbyists*, Annapolis MD: Naval Institute Press, 1990.

---"Now Hear This—Improving Wargaming Is Worthwhile—and Smart," *Proceedings Magazine*, January 2016.

Peter P. Perla and ED McGrady, "Why Wargaming Works," *Naval War College Review* 64, no. 3, 2011.

United States Marine Corps, *Expeditionary Warrior 2013: Future Maritime Operations for the 21st Century Operating Environment Final Report*, June 6, 2013. Available as of March 31, 2018: http://www.mcwl.marines.mil/Portals/34/Documents/EW13%20Final%20Report_FINAL.pdf

Milton G. Weiner, "Trends in Military Gaming," Santa Monica, CA: RAND, 1969.

Christopher A. Weuve et al., "Wargame Pathologies," Arlington, VA: CNA, 2004.

Yuna Huh Wong, "Preparting for Contemporary Analytic Challenges," *Phalanx* 47, no. 4, 2014.

Yuna Huh Wong, Sebastian Bae, Elizabeth M. Bartels, and Benjiman Smith, "Next Generation Wargaming for the U.S. Marine Corps: Recommended Courses of Action," Santa Monica, CA Forthcoming.