

Building Better Games for National Security Policy Analysis

Towards a Social Scientific Approach

Elizabeth M. Bartels

This document was submitted as a dissertation in March 2020 in partial fulfillment of the requirements of the doctoral degree in public policy analysis at the Pardee RAND Graduate School. The faculty committee that supervised and approved the dissertation consisted of Steven Popper (Chair), Stacie Pettyjohn, and Yuna Wong.



PARDEE RAND GRADUATE SCHOOL

For more information on this publication, visit http://www.rand.org/pubs/rgs_dissertations/RGSD437.html

Published 2020 by the RAND Corporation, Santa Monica, Calif.

RAND® is a registered trademark

Limited Print and Electronic Distribution Rights

This document and trademark(s) contained herein are protected by law. This representation of RAND intellectual property is provided for noncommercial use only. Unauthorized posting of this publication online is prohibited. Permission is given to duplicate this document for personal use only, as long as it is unaltered and complete. Permission is required from RAND to reproduce, or reuse in another form, any of its research documents for commercial use. For information on reprint and linking permissions, please visit www.rand.org/pubs/permissions.html.

The RAND Corporation is a research organization that develops solutions to public policy challenges to help make communities throughout the world safer and more secure, healthier and more prosperous. RAND is nonprofit, nonpartisan, and committed to the public interest.

RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

Support RAND
Make a tax-deductible charitable contribution at
www.rand.org/giving/contribute

www.rand.org

Abstract

This monograph proposes an approach to game design grounded in logics of inquiry from the social sciences. National security gaming practitioners and sponsors have long been concerned that the quality of games and sponsors' ability to leverage them effectively to shape decision making is highly uneven. This research leverages literature reviews, semi-structured interviews, and archival research to develop a framework that describes ideal types of games based on the type of information they generate. This framework offers a link between existing treatments of philosophy of science and the types of tradeoffs that a designer is likely to make under each type of game. While such an approach only constitutes necessary, but not sufficient, conditions for games to inform research and policy analysis, this work aims to offer pragmatic advice to designers, sponsors and consumers about how design choices can impact what is learned from a game.

Table of Contents

Abstract	iii
Figures	ix
Tables	xi
Summary	xiii
Acknowledgments	xvii

CHAPTER 1

Introduction: Games for National Security Policy Analysis and How to Improve Them. 1

What Is a National Security Policy Analysis Game?	1
Why Use “National Security Policy Analysis Games”?	2
Elements of a National Security Policy Game	4
Defining Games in Relation to Analytical Tools	6
Application of National Security Policy Games.....	8
Typologies of Applications of National Security Policy Games.....	10
The Limits and Misuses of Policy Gaming	12
Building Better Games for National Security Policy Analysis	16
What Makes a Game Good?.....	17
Two Competing Explanations for the Gap	19
Towards a Scientific Approach: Defining Logics for Game Design	22

CHAPTER 2

Study Approach..... 23

Research Context.....	23
Limits of Existing Information on Specific Games.....	23
Limits of the Expert Community.....	26
Study Approach	27
Phase 1: Understanding Policy Game Design in Scientific Terms	27
Phase 2: Initial Framework Design	27
Phase 3: Validation.....	31
Phase 4: Examples	33

CHAPTER 3

Towards a Social Science of Policy Games 36

Policy Games as Art: A Dominant Paradigm with Problems.....	37
Pitfalls of Artistic Approaches to Policy Analysis Gaming	39
The Artistic Case Against Science	41
Towards a Science for Gaming	42

Jackson’s Typology of Philosophies of Science	44
Philosophies of Science for Gaming	48
Producing Scientific Knowledge with Games	53
The Nature of Problems Suited to Gaming and the Need for a Bayesian Approach to Certainty	54
Game Evidence: Empirical or Formal?	55
Basis of Game Analysis: Differences or Mechanisms?	56
Conclusion	58

CHAPTER 4

Four Archetypes of Games to Support National Security Policy Analysis	60
Overview of the Four Archetypes	62
Differentiating the Types.....	63
Games with Characteristics of Multiple Types	69
Connecting Philosophy of Science for Games to the Archetypes	69
System Exploration	70
Alternative Conditions	71
Innovation.....	71
Evaluation.....	72
Different Games for Different Philosophies.....	73
Design tradeoffs.....	74

CHAPTER 5

Designing Games for System Exploration.....	76
Overview of Example System Exploration Games	77
Exploring Across Escalating Scenarios: U.S. Air Force/ RAND Project Sierra Middle East Games— Jordan Series	77
Building to a Structured Seminar Game for System Exploration: U.S. Army/ RAND Gray Zone Wargame	79
Design Tradeoffs Related to the Game Environment.....	82
Challenges of Selecting the Game Environment.....	82
Challenges of Scaling the Game Environment.....	84
Challenges of Building a Credible Environment.....	85
Design Tradeoffs Related to the Game Actors	86
Scoping What Actors are In Play Shapes System Exploration Games	86
Depicting Actors Not Assigned to Players.....	87
Player Selection is Critical in System Exploration Games	88
Player Engagement is Essential to Credible Results	89
Common Limitations of Players in System Exploration Games.....	89
Design Tradeoffs Related to the Game Rules	91
Use the Best Available Evidence to Build Initial Rules.....	91
Adjudication Should Focus on Transparency.....	92
Managing Time	94

Challenges of Hidden or Incomplete Information.....	95
Conclusions	96

CHAPTER 6

Designing Games for Alternative Conditions	97
Overview of Example Alternative Condition Games.....	99
US Air Force/RAND Strategy and Force Evaluation (SAFE) Games	100
RAND Force Structure Decision Analysis Game	102
Design Tradeoffs Related to the Game Environment.....	104
Challenges of Explicit World Building	104
Challenges of Second Order Effects and Change over Time	105
Design Tradeoffs Related to the Game Actors.....	106
Challenges of Players	106
Challenges of Representing Actors	108
Design Tradeoffs Related to the Game Rules	108
Challenges of Formalized Rules.....	109
Challenges to Comparison from Player Decisionmaking	109
Challenges to Comparison from Adjudication.....	112
Conclusions	113

CHAPTER 7

Designing Games for Innovation.....	114
Overview of Example Innovation Games	115
Innovation in Warfighting: OSD(P)/ CNA Persistent Hobgoblin Series	116
Innovation in Processes and Procedures: U.S. AFRICOM/ RAND OCEANS 17 Table Top Exercise	117
Design Tradeoffs Related to the Game Environment.....	119
Design Tradeoffs Related to the Game Actors.....	121
Challenges of Blue	121
Challenges of Red	122
Design Tradeoffs Related to the Game Rules	123
Conclusions	124

CHAPTER 8

Designing Games for Evaluation.....	125
Overview of Example Evaluation Games	126
Experimental Design: DARPA/CNA ScudHunt	127
Comparative Case: AFRICOM/RAND Security Force Assistance Game.....	129
Design Tradeoffs Related to the Game Environment.....	131
Design Tradeoffs Related to the Game Actors.....	133
Design Tradeoffs Related to the Game Rules	135

Conclusions	138
CHAPTER 9	
Trends in RAND Corporation National Security Policy Analysis Gaming: 1948 to 2019	139
Gaming for Nuclear Comprehension at RAND: 1948-1958	140
Early Force Structure, Posture, and Planning Games.....	142
Managing Crises and Fighting General Wars in Games	145
Gaming the Prospects of Limited War: Project Sierra	149
Gaming and Gaining a Deeper Understanding: 1959-1970	151
Continuation of the Tradition of Political-Military Crisis Games for Systems Exploration.....	151
Exploring Alternative Conditions' Impact on Force Structure	154
Towards Evaluation and Away from National Security Policy Analysis Games	158
Games in Eclipse at RAND: 1970-1990.....	159
Repeating the Cycle of Boom and Bust while Expanding Scope: 1990-2014	163
Gaming to Understand the Post-Soviet Policy System	163
Supporting Defense Gaming: A New Model	165
The Current Era of RAND Gaming: 2014-2019	166
Thoughts on the Future.....	169
CHAPTER 10	
Conclusions, Policy Recommendations, and Next Steps.....	170
Conclusions	170
Policy Recommendations	176
Recommendations for the Sponsors of Games.....	176
Recommendations for the Designers of Games	178
Recommendations for the Consumers of Games	180
Next Steps: Options for Testing the Framework	181
APPENDIX	
A. Sample Template for Documenting Game Designs	184
Game Purpose and Objective	184
Game Design Tradeoffs.....	184
Environment	184
Actors	184
Rules	185
Insights and Finds.....	185
Recommendations	185
Bibliography	186

Figures

Figure S.1 Basic Types of Information Generated by Games	xiv
Figure 3.1: Jackson's typology of philosophies of science.....	45
Figure 4.1: Game archetypes organized by defining characteristic	65
Figure 5.1. Process of Project Sierra Games.....	79
Figure 5.2: Sample Action Cards from the Gray Zone Game	81
Figure 5.3: The Gray Zone Game Board	81
Figure 5.4: Factors in air strike adjudication	94
Figure 6.1: Strategies represented in plays of the SAFE games	101
Figure 6.2: Process of SAFE game play	102
Figure 6.3: Comparative format for the Force Structure Analysis game.....	104
Figure 6.4: Sample of Menu of Investment options in SAFE.....	111
Figure 7.1: Game board for OCEANS 17, Showing the Threat Actor in Red and Available U.S. Air Bases in Violet.....	118
Figure 8.1: SCUD Hunt Game Board.....	128
Figure 8.2. Sample cards to show investment in potential partner forces	130
Figure 9.1: Planning and Aggregated War Game Process.....	143
Figure 9.2: Play of STRAW with RAND Board of Trustees in 1953	144
Figure 9.3. Details of SWAP game boards for procurement and warfighting.....	145
Figure 9.4: Status Board for Tactical Air War Game	147
Figure 9.5: Project Back Stop Study Process.....	153
Figure 9.6: Adjudication of SAFE Game play.....	156
Figure 9.7: Sample NATO Planning Map for TOTEM Ground Combat Simulation	160
Figure 9.8. RSAS structure	162
Figure 9.9: General Scheme of a Day After Game	165
Figure 10.1: Archetype of Information Produces by Games and the Philosophies that Underpin Them	174

Tables

Table 2.1: Affiliation of interview and selected workshop subjects.....	33
Table 3.1: Implications of philosophies on games-based research.....	52
Table 4.1: Secondary distinguishing characteristics of each game archetype	66
Table 4.2: Degree of alignment between the three philosophies and four archetypes	73

Summary

A game involves human players representing actors, who make decisions in a competitive environment based on a set of implicit or explicit rules, and grapple with the potential consequences of their actions. As an approach to policy analysis, gaming has a long history in U.S. national security circles and has shaped important policy discussions on topics ranging from early Cold War nuclear policy to emerging challenges around unconventional warfare today. In recent years, sponsor interest in games has resurged, fueled in part by renewed interest in historic games that shaped national security policymaking at key decision points, as well as new challenges demanding novel insights to address them. Despite long-standing investment in using gaming as an important tool for policy analysis, game sponsors, designers, and consumers often note that games are of uneven quality and do not achieve their intended objectives.

Positing a Social Scientific Approach to Game Design

This monograph proposes a new framework that seeks to describe a social scientific approach to game design. This framework is predicated on the argument that by using existing logics for scientific inquiry, it will be easier to design choices to the desired analytic ends. While a truly great game requires creativity and artistry on the part of designers, this work argues that ensuring the basis of games are logical is a necessary step to generating results that further research and analysis. This argument advances in three parts.

The first element is a description of several major philosophies found in social science and how they might be applied to gaming. This section presents evidence that expert designers already use these logics, but describes the philosophies using a more generalized vocabulary that seeks to make the commonalities between games and other tools for scientific research more evident. This section also argues that while there is more than one approach to science that can generate credible evidence, the standards that each has are different. As a result, work produces under one approach should not be judged by the standards of the other approaches.

Second, this monograph develops a typology of four purposes of games, each of which aligns with one or more philosophy:





- **Systems exploration** games develop an understanding of a particular policy problem from a range of perspectives. They elicit and synthesize the mental models of expert players. For example, such a game might be designed to develop a simple but useful model of what factors might lead to rapid increases in Iranian nuclear capabilities.
- **Alternative conditions** games seek to understand how a key factor shapes decisionmaking processes and choices. They do this by comparing player decisions under different conditions. A series of alternative conditions games might examine the

extent to which varying levels of domestic unrest contribute to Iran accelerating or decelerating the development of nuclear capabilities.

- **Innovation** games seek to generate new solutions to policy problems. They highlight where new decisions could be made to change how a system works and motivate players to propose new ideas. Games to spark innovation tend to focus on competition within a less constrained environment than current decisionmakers face. An innovation game might seek to generate new strategies for arms control that could blunt the advancement of operational Iranian nuclear capabilities.
- **Evaluation** games are intended to assess policies and strategies—while they cannot offer proof that a strategy will succeed, they can suggest potential pitfalls, or offer modest evidence in favor of a promising potential solution. They do this by building a credible representation of the outcomes of player decisions to enable judgement. For example, a game could compare the Iranian reaction to several different proposed treaties in order to better anticipate potential sticking points and unintended second order effects.

Figure 1 illustrates how the relationship between these types are defined by two major decisions about the purpose of the game. The first is whether the priority is to develop an understanding of the policy problem, or to develop strategies to address the problem. This will inform the focus of game analysis. The second is characterizing whether the project is early stage research to inform the research team and sponsor, or whether it is part of a more mature effort to influence outside stakeholders. This informs how credible the information generated by the game needs to be—a game to shape follow-on study priorities will not need to stand up to the same level of scrutiny as a game designed to inform a major decision.

Figure S.1 Basic Types of Information Generated by Games

	Develop an understanding of the problem	Develop strategies to address the problem
Early research to inform research team and sponsor	 <p>System Exploration</p>	 <p>Innovation</p>
Mature research to influence external stakeholders in a decision process	 <p>Alternative Conditions</p>	 <p>Evaluation</p>

The third element of this work extends the discussion of the four archetypes to consider the design tradeoff associated with each type, illustrated with design choices from two games. These chapters argue that while game designers will always need to work within constraints, design choices that undermine the core logic of the philosophy of science and game archetype pose a threat to the ability of the game to produce credible information that meets the game's purpose.

Policy Recommendations and Next Steps

Taken together, these components offer a perspective on how to sponsor, design, and consume games.

Sponsors should consider providing *additional, specific guidance* to game designers *about the purpose of the games and what kinds of evidence will be persuasive to them*. At the same time, they should be sure to *limit the purpose and scope* of each game so that a coherent design can be generated. Finally, they should work with their designers to *oversee a logical link* between design choices and anticipated findings.

Designers should consider how to *advise sponsors* more directly and effectively on the limits of the kinds of evidence and insights games can provide. As they design games, they should *make explicit tradeoffs*, given those limits, to maximize the usefulness of the game and insights generated, and *document the links* between the design and the findings generated.

Consumers of game insights should be sure to *evaluate the insights generated by games based on their intended purposes and approach*, rather than applying other standards (for example, rejecting insights from a system exploration game intended to explore a problem because they did not rigorously generate and evaluate solutions for that problem). As gaming continues to expand again as a set of tools used in policy analysis, they should use care when applying the findings from a game to different purposes, in order to maximize useful insights generated.

This monograph presents a social scientific way of thinking about game design and offers a number of practices that follow on from this approach. However, available data was insufficient to conduct a rigorous test to verify the extent of the frameworks descriptive power, nor offer empirical support for the approaches ability to improve the utility of games for policy analysis. While the framework cannot be comprehensively tested at this time, feedback from designers, sponsors, and consumers who opt to employ the framework can serve as an interim form of assessment.

Acknowledgments

While my name is alone on the cover of this monograph, it is the product of many people's intellect, hard work, and patience accumulated over the 10 years that I have been working as a national security gamer. Any effort to thank all those who deserve credit is bound to fall short, and so I apologize in advance for those names I omit in error.

First to my committee. One could not ask for a kinder, more thought-provoking chair than Steven Popper—thank you for straightening up both my ideas and my prose at every turn. Stacie Pettyjohn was a tireless advocate for this project, pushing me to connect the work both to broader RAND efforts to reinvigorate gaming, as well as efforts in academic political science to do the same. Yuna Wong can claim credit for suggesting back in 2013 that my frustrated mutterings about the field were the start of a dissertation, and one that should be conducted at RAND. My profound thanks to you all.

Beyond those formally on my committee, I found myself in the fortunate position of having many advisors and active participants from the gaming community, without whom this project would not have been possible. Peter Perla kindly agreed to serve as outside reader—his frank and constructive feedback were critical in clarifying this work. Arguments with and between Jon Compton, Stephen Downes-Martin, Margaret McCown, Ed McGrady, Phil Pournelle, and Chris Weuve were critical in shaping key claims. I also owe thanks to Matt Caffrey, Tim Wilkie, and Scott Chambers (Connections Wargaming Conference); Erik Lin-Greenberg, Reid Pauly, Andrew Reddie, and Jackie Schneider, (International Studies Association); Ivanka Barzashka and Anna Nettleship (King's College London's Wargaming Network); Mike Ottenberg (MORS Wargaming Community of Practice); and the War Gaming and Strategic and Operational Research Departments of the U.S. Naval War College, who provided important forums to present parts of this monograph and conduct workshops. I owe particular thanks to Jackie, who urged me to write the paper that became Chapter 3 and very much changed the nature of the project in the process. Beyond those named, to the many other member of the wargaming community who took the time to answer surveys, respond to emails, answer interview questions, attend workshops, and give feedback on presentations, my deep thanks.

At RAND, I found a community who believed in the value of study of gaming, in both theory and practice. Sally Sleeper, Jennifer Kavanagh, Yamit Feinberg, Marcy Agmon, Christine Wormuth, Mike McNerney, Mike Spirtas, Laura Baldwin, Megan McKeever, and Brien Alkire provided support and guidance from which this project benefited tremendously. The project also would not have been possible without the resources of the RAND Corporation Archive—Cara McCormick was essential in helping locate relevant historical materials, making them accessible to me, and overseeing review of the resulting research. Chapter 9 could not exist without her efforts. The RAND Methods Center for Gaming, the community of game designers at RAND,

and particularly my fellow “Dames of Wargames” provided essential fellowship and community. Finally, thanks to the many researchers who worked with me to create games at RAND—much of my thinking was shaped in project work conversations. Along with others already mentioned, particular thanks are owed to: Ben Connable, Abby Doll, Jeff Drezner, Shira Efron, Aaron Frank, Adam Grissom, Sarah Harting, Caitlin Lee, Igor Mikolic-Torreira, Chris Mouton, Karl Mueller, Jenny Oberholtzer, Dave Ochmanek, Joel Predd, Dave Shlapak, Geoff Torrington, and Becca Wasser.

I also own my thanks to the members of the Pardee RAND Graduate School community who saw me thought the process. Susan Marquis and Rachel Swanger were critical advocates on my behalf. Dave Baiocchi and Gery Ryan were rocks of stability, advice, and good cheer. I owe particular appreciation to Mary Parker, Terresa Cooper, and Amy Nabel on who I all too frequently depended on to correct the many mistakes I made along the way. Paul Dryer was an excellent mentor and tireless advocate for my presence at PRGS. Chris Nelson and Andrew Parker oversaw the independent study that underpinned how this work was conducted. Finally, to my fellow students, particularly my “office mates” Ify Edochie and Sara Turner, walk buddy Erin Duffy, baking buddy Diana Carew, defense partners-in-crime Nick Martin and John Speed Meyers, mentor Bonnie Triezenberg, and mentees Hilary Reininger and Damien Baveye.

The clarity of this document was greatly improved by David Adamson editing work.

To my friends and family, who all too frequently had to listen to yet another description of what it is I do for a living--thank you for your enthusiastic support and patience over the last five years.

Finally, this work is dedicated to the two people most responsible for it existing.

The first is my partner, Dave Kasten. Dave’s contributions to this project included countless conversations that clarified key ideas, editing much of the document, two cross country moves, and more requests from me to figure out dinner than I care to think about. I am profoundly grateful for his care and attention of me and my ideas through a sometimes-grueling process.

Second is Deirdre Hollingshed, who taught me how to do the work of gaming. From weaving multiple narratives into a single document to writing an email that wouldn’t get me fired, hers are the lessons I turn to day in and day out to do my work. I hope this work can offer new gamers a small portion of the help and good advice she provided me.

Chapter 1: Introduction: Games for National Security Policy Analysis and How to Improve Them

This chapter sets the scene for the monograph to come by answering three fundamental questions: What are national security policy analysis games? What uses are they put to? And why do I think they are in need of improvement? The first section defines what a national security policy analysis game is—first by offering existing terms and definitions, then by explaining why I use the term “national security policy analysis games” and what I mean by it. I then flesh out the concept by defining the key elements of a game and discussing how games compare to other analytical tools the reader may find more familiar. The second section turns to how games have contributed to national security analysis, starting with a historical analysis and moving to contemporary perspectives, including an overview of how different types of games have been traditionally defined. The section concludes with a discussion of the limits and potential for abuse of games, naturally leading to the third section on the potential to improve games for policy analysis. This final session discusses the limits on current approaches for defining what makes for a “good” game. I argue that the current dominant approach to games which treats design as primarily an artistic practice has contributed to the field. Instead, I contend that games for policy analysis would be better served by articulating scientific principles that ought to underpin game designs. The bulk of this monograph is devoted to articulating what such a scientific approach might consist of.

What Is a National Security Policy Analysis Game?

The U.S. Department of Defense official definition of wargaming is “representation of conflict or competition in a synthetic environment, in which people make decisions and respond to the consequences of those decisions.”¹ A game is a model of a particular national security problem with human players representing actors with competing goals attempting to develop a strategy to improve their position. These actors are able to make decisions to define and implement their preferred approach to managing the problem through processes framed within the game. These decisions then affect through a set of rules other actors and the broader environment in which the competition is taking place. Put concisely, a game involves human

¹ Joint Chiefs of Staff, *Joint Publication 5-0: Joint Planning* (Washington, DC: Joint Chiefs of Staff, 2017). p V-31. This definition was only added in the 2017 edition of the joint publication and takes as its primary inspiration Peter Perla’s frequently cited definition of games as “a warfare model or simulation whose operation does not involve the activities of actual military forces, and whose sequence of events affects and is, in turn, affected by the decisions made by players representing the opposing sides.” which previously served as a consensus position.

players representing actors, who make decisions in a competitive environment based on a set of implicit or explicit rules, and grapple with the potential consequences of their actions.

The representation of the relevant actors, environments, and rules of different games vary a great deal. A game might consist of a single player at a computer terminal directing animated military forces through realistic depictions of a real-world theater of operation; or 15 players periodically rolling dice on either side of a map while moving cardboard counters printed with military symbols; or hundreds of military officers sorted into small teams receiving occasional written messages about an escalating diplomatic crisis. All fall under the rubric of games.

This diversity has predictable consequences when it comes to developing a clear understanding of what is, and is not, a game. A gaming professional once quipped that the collective noun for a group of wargames ought to be an “argument,”² and nowhere is this as evident as in the inability of the field to come to a common definition.³ While few gamers would disagree with this core statement, many different permutations and interpretations exist, generating vigorous conflicts over what is, and is not, a game. For example, the doctrinal definition is situated within a discussion of course of action (COA) analysis, leading some to argue that only games exploring a proposed military action, using set procedures listed in the document, can properly be called wargames. Others point to the much wider range of applications observed in national security work. There is also debate whether “competition” requires a human adversary or if forces such as disease, natural disaster, or bureaucratic friction provide the necessary competition. Other debates focus on how concrete the decisions of players need to be and how much those decisions need to directly shape the consequences represented. While efforts have been made to resolve these tensions, consensus has been slow to emerge.

Why Use “National Security Policy Analysis Games”?

As an attempt to bring some clarity to this debate, I prefer the term “national security policy analysis game” rather than the more common “wargame” or “serious game” for several reasons (though I contract the somewhat unwieldy phrase to “game” or “wargame” in this work for the sake of brevity). First, the use of “national security” rather than “war” denotes a broader range of topics that better reflect the actual application of the tool to issues such as crisis management, measures short of war, and bureaucratic policy areas such as acquisitions and personnel that have major implications for national security beyond fighting major wars. It also is more inclusive of diplomacy and development community members who may find the term “war” off-putting culturally as well as not being descriptive of their work. At the same time, I include the

² Graham Longley Brown, *Successful Professional Wargames: A Practitioner's Guide*, ed. John Curry (The History of Wargaming Project, 2019). p 3

³ Indeed, some scholars of games, most notable Ludwig Wittgenstein, have argued that the difficulty in defining a “game” generically is inherent because the concept is one with “blurred edges” that is nevertheless useful. See: Ludwig Wittgenstein, *Philosophical Investigations*, trans. G.E.M. Anscombe, P.M.S. Hacker, and Joachim Schulte, Revised 4th edition ed. (Chichester, UK: Wiley-Blackwell).

descriptor “national security” to make it clear that I am not including other areas to which gaming has been applied under the rubric of “serious games” or “policy games” such as urban planning, environmental action, health and education.⁴ While the advice in this work may be helpful to practitioners outside the national security spaces, I have not made careful study of these other applications and look to others to evaluate the applicability outside the scope of this work.

Second, the addition of “policy analysis” makes it clear that these are games designed to inform policy decisions by generating a better understanding of or information about a real-world problem. I use “analysis” in the general sense to refer to efforts to better understand the elements and structure of a policy area. Put differently, I use the term to apply to a broad range of activities that are also sometimes referred to as research, inference, or studies—such work need not be quantitative (as is sometimes inferred) nor devoted only to approaches that decompose aspects of a larger problem (as is indicated in some formal definitions of the term “analysis”). The focus on “analytical” games excludes two substantial portions of the field: commercial and educational games.⁵ In the evocative words of Jon Compton “[games designed for different purposes] may look very similar, but much like the tool-box metaphor, just because there is a commonality of tools used for both plumbing and electrical, these tools are being used for completely different things.”⁶

Historically, the lines between commercial and policy games have been blurred with some well-known hobby designers being asked to consult for the government and other designers engaging in sustained careers in both spaces.⁷ Many well-respected national security game designers argue that conversance with a range of commercial game designs is critical as a repository of design approaches. However, although commercial games can be used as a means of supplying additional clever mechanisms and useful representations, these games are still fundamentally designed to entertain and, to a lesser extent, to make money. As such, commercial game designers are not under the same requirements as analytic game designers to tie their game design to the ability to produce credible knowledge about a policy problem.⁸ As a result, the

⁴ An assessable summary of some of these applications can be found in: Igor S Mayer, “The Gaming of Policy and Politics of Gaming,” *Simulation & Gaming* 40, no. 6 (2009).

⁵ For a discussion of game typologies that includes deeper discussion of educational and commercial games in relation to games for research, please see: Elizabeth M. Bartels, “Gaming – Learning at Play,” *ORMS Today*, 2014; Phillip E Pournelle, “Designing Wargames for the Analytic Purpose,” *Phalanx* 50, no. 2 (2017).

⁶ Jon Compton. “Analytical Gaming.” (2014). p 4.

⁷ For a history of the interactions of commercial and serious wargame design, see: Peter P. Perla, *The Art of War Gaming: A Guide for Professionals and Hobbyists*, ed. John Curry, 2nd ed. (History of Wargaming Project, 2011); Wittgenstein, *Philosophical Investigations*; *ibid.*; *ibid.* and Matthew B. Caffrey, *On Wargaming: How Wargames Have Shaped History and How They May Shape the Future* (Newport, RI: Naval War College Press, 2019).

⁸ Perla, *The Art of War Gaming: A Guide for Professionals and Hobbyists*. pp 175-176

design considerations that this monograph is most keenly concerned with do not hold the same weight in discussions of commercial game design.

Similarly, I do not consider games designed for education and training applications in this monograph. This is not to suggest that education games are not important to the national security space—wargaming is integrated into many national security education curricula⁹ and likely represent a major portion of all defense gaming. There is also important crossover between education and research games. Games designed as part of research efforts may have educational goals. For example, a goal of a game might be to share knowledge within the design team, a task I do consider as part of research since team-based research processes on any kind tend to involve a stage of work understanding the current understanding of the problem, requiring synthesis and other analytical tasks. Contrast this with games used in later stages of the project to teach the new knowledge generated during research to others, which strike me as more straight forward efforts to communicate knowledge to others, and thus sit closer to education games. However, the division between these is often a matter of perspective,¹⁰ and so other researchers may well debate where the line is drawn between research and education in specific cases. I have opted to keep my scope narrow.¹¹

Elements of a National Security Policy Game

National security policy games exhibit tremendous variation in form, but the definition offered above points to three primary components that must be present in all games—an environment that instantiates a policy problem, actors who can shape the course of events, and rules that govern how actors influence one another and the environment.¹²

Games feature a synthetic environment that presents the context of a national security challenge. Depending on the game design, the design of the environment could consist of a scenario that lays out the story of how we got to the policy crisis in question, but might also present on the game board, maps, pictures, data sets and other tools that help players understand the context for their decisions. Second, human players are asked to represent key actors, who have different goals and abilities, and can make decisions that impact the policy challenge.

⁹ Chairman of the Joint Chiefs of Staff, "Officer Professional Military Education Policy," (Washington, DC2015).

¹⁰ William M. Jones. "One View of Games, Simulations and Analogs." (Santa Monica, CA: RAND Corporation, D-12290-PR, 1964). p 6

¹¹ In an earlier iteration of the framework I did include educational games, but any game that had both educational and research objective shared the same characteristics as a game that only had the research objective. However, the presence of the educational objective provided an "out" to designers to declare a game successful even if it did not produce the desired information. I opted to remove them from the framework to more clearly focus attention on designing games to produce desired information. I am indebted to discussion with the U.S. Naval War College's War Gaming Department in general, and Shawn Burns in particular for clarifying this point.

¹² Other designers have opted to subdivide these features further: for example, "objectives, scenario, database, models, rules and procedures, infrastructure, participants, analysis, culture and environment, and audiences" from: Christopher A. Weuve et al., "Wargame Pathologies," (Arlington, VA: CNA, 2004). p 1

Sometimes these details are provided as part of the game, in other cases players bring their own expertise to fleshing out the intentions and capabilities of the actors they represent. Third are the rules that shape how actors can affect one another and the environment by determining what actions they can take and those actions' plausible effects over time—in other words, they map out the potential causal relationships between actors and the environment. These can take the form of lists of formalized rules but also may leverage the expertise of subject matter experts to shape them dynamically. Taken together, these tools allow the game designer to build an artificial world, populate it with human decisions, and enable those humans to experience meaningful consequences of their choices.

While it is helpful to speak about these elements as separate when discussing design tradeoffs, it's important to note that in practice the line between different parts of the policy system is often less crisp. For example, it may be difficult to disentangle where the line between the “resources” of an actor ends and the rules that govern their actions begin—a game might opt to represent finite resources of an actor by limiting the number of actions they can take. Similarly, actors not represented by human players might be treated only as part of the environmental context provided in the scenario. The divisions between elements are intended to promote consideration of all aspects of design, not to insist on how those elements are eventually packaged together.

A great deal of existing work on game design is dedicated to enumerating different ways in which a designer might opt to represent these core elements. For example, games have traditionally been defined by whom they draw on to serve as players (e.g. senior leader seminar), the range of actions available to players (e.g. fixed vs open), the medium of the game (e.g. hex-and-counter¹³ or computerized games), the method of adjudication (e.g. rigid rules vs. umpired games), the role of computers (man-machine vs free-form games) or some combination of these factors. Former Dean of Naval War College War Gaming, Tom Culora, distinguishes among large multiplayer games with fairly open adjudication, small iterative games with more rigid adjudication, “regency” games focused on senior leader education, and massive online games.¹⁴ Other commentators link the style of adjudication to the medium of the game: umpired games, rigid manual games, or computerized games.¹⁵

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, a sponsor who wants to define the participants, format of the game, or means of adjudication too early can overly constrain designers' choices, preventing the designer

¹³ That is, a map-like game board with a super-imposed hexagonal grid to tabulate range and movement for cardboard printed counters (game pieces) containing information about and representing specific forces or game-relevant states.

¹⁴ Thomas J Culora, "A War-Gaming Renaissance," *Proceedings*, 2016.

¹⁵ Andrew Wilson, *The Bomb and the Computer* (London: Barrie and Rockliff 1968). pp 46-47

from developing the most appropriate game possible to answer the analytic problem at hand. Furthermore, advice to game designers that consists mostly of potential design elements might be helpful in expanding the options considered by experienced designers but are unlikely to assist new designers in selecting which design elements are most appropriate to their specific analytic challenge.

Defining Games in Relation to Analytical Tools

Within national security analysis, games are often positioned in association with and in distinction to three other sets of analytic techniques: workshops and seminars, modeling and simulation (M & S), and exercises. The detailed distinctions among these three other approaches are somewhat obscured by imprecision in common usage, but the general characteristics are relatively well agreed to. An exercise is “a military maneuver of simulated wartime operation,”¹⁶ and is distinguished from a wargame by the use of actual military equipment and movement of forces. In contrast, a workshop “involves subject matter experts gathering to discuss a problem”¹⁷ but does not have the focus on driving the participants to make concrete decisions or dealing with the projected results of those actions. While workshops can bring together subject matter experts to synthesize findings, the lack of structure “sometimes lead[s] to a discouraging diffuseness and inconclusiveness when these discussions are summarized.”¹⁸ Finally, models refer to a “logical representation of a system, entity, phenomenon or process”¹⁹ and a simulation is a “model acted over time.”²⁰ In common defense use, these most often refer to mathematical models, usually dependent on computers to operate, and thus contrast with games because they do not involve actual humans making decisions.

While these distinctions seem fairly clear in theory, in practice the lines are rather blurry.²¹ For example, some exercises involve the participation of an actual staff in a simulated engagement against an active red adversary. Particularly in command post and staff exercises, where the roles being rehearsed are focused on planning and decision making, it might be quite hard to clearly distinguish the line between exercise and game. In the case of workshops, the question becomes how clear must decisions be, and how direct must be the link between decision

¹⁶ William L. Simpson Jr., “A Compendium of Wargaming Terms (Updated),” Military Operations Research Society Wargaming Community of Practice, (2017). p 19

¹⁷ Ibid. p 49

¹⁸ 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, D-3164-RC, 1955). p 3

¹⁹ Simpson Jr., “A Compendium of Wargaming Terms (Updated).” p 29

²⁰ Ibid. p 39

²¹ It is perhaps worth noting that the blurring is hardly a new phenomenon—the confusing between terms, and the communities limited ability to enforce disambiguation have long been a source of frustration. For example, see: Martin Shubik, “On Gaming and Game Theory,” (Santa Monica, CA: RAND Corporation, P-4609, 1971).

and consequence? For example, if an event features a pre-set script, in which players make choices about how to respond to a crisis, but the behavior of outside actors is precalculated and considered dominant in determining outcomes, the proper categorization might be very blurry indeed. Similarly, a “human-in-the-loop” model or simulation is a category of M & S where a human must interact with a model to make decisions at key points—a process that would be indistinguishable from a computerized wargame. As a result, it’s generally best to conceptualize these tools on spectrums, and recognize that different practitioners will opt to draw the line between the use of terms in somewhat different places.

As often as not, the choice of term used for any particular event is driven by bureaucratic considerations rather than a desire for analytic clarity. For some communities the use of “game” invokes frivolity rather than serious analytic pursuits. A designer working in such a space may opt to use “exercise” or “seminar” to evoke a more serious tone. Alternatively, communities that value quantitative research may push for “simulations” over “games.” At the same time, the availability of resources for particular tools or a term’s cachet in the organization may encourage the use of one term over another as a means of securing additional resources and attention for the work. Finally, objections to the use of one term over another might be a form of policing the quality of the activity—for example, the pejorative term BOGSAT (“bunch of guys sitting around a table”)²² is often used to claim that an event should not properly be considered a game, sometimes because it lacks a definitional element, but sometimes simply due to poor quality execution. Games in which outcomes are determined by expert judgement might be deemed “BOGSATs” if the quality of the expertise or transparency of decisionmaking is not sufficient. The reality is that while a designer should always be clear in their own mind what tool they are using, and what the attending limitations ought to be for analysis, the use of these terms is likely to be somewhat fuzzy in practice.

Games also occupy a small, but historical place within social science research on national security issues, particularly within international relations in political science.²³ Confusingly, in this context games are most often referred to as simulations. Here, games are often compared to, and distinguished from, 1) case study approaches that draw on archival records and interviews to build a rich understanding of a historical case, 2) formal models that use game theory and other mathematical tools to model behavior, and 3) lab and survey experimental approaches. However, in contrast to the other comparisons listed above, in the context of political science, gaming and

²² Simpson Jr., “A Compendium of Wargaming Terms (Updated).” p 51

²³ For a historical perspective, see: Harold Guetzkow et al., *Simulations in International Relations: Developments for Research and Teaching* (Englewood Cliffs, NJ: Prentice-Hall, Inc., 1963). and Lincoln P. Bloomfield, “Reflections on Gaming,” *Orbis* 27, no. 4 (1984). A new wave of political science research leveraging games can be seen in work like: Reid B.C. Pauly, “Would U.S. Leaders Push the Button? Wargames and the Sources of Nuclear Restraint,” *International Security* 43, no. 2 (2018)., Erik Lin-Greenberg, “Game of Drones: What Experimental Wargames Reveal About Drones and Escalation,” *War on the Rocks*, 2019., Jacquelyn G. Schneider, 2018; Andrew W. Reddie et al., “Next-Generation Wargames: Technology Enables New Research Designs, and More Data,” *Science* 362, no. 6421 (2018).

simulation are far less established as tools for research—most literature and active practice focuses on gaming as a tool for teaching, rather than research, and thus falls outside the scope of this study.

Application of National Security Policy Games

Having defined what a game is (however messy and imperfect that definition might be), we can now turn to examining how games are used in national security policy research and analysis. Within national security policy spaces, games are applied to a wide range of purposes and problems, the utility and appropriateness of which are debated.

From a historical perspective, early, highly abstract games like set, chess, and go were used for elite education.²⁴ 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, most notably using the Prussian staff “Kriegspiels”—that is, war games. Through the late 19th and early 20th centuries, both educational and exploratory applications of gaming grew and were used to consider issues ranging from adversary behavior to emerging technology on the battlefield.²⁵ 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.²⁶ Particularly in the US, many of these games were informed by the growing commercial board-game sector, which developed new mechanics for representing historical and contemporary problems that were then adopted by those designing games for the Department of Defense.²⁷ Non-military application also grew over the latter half of the 20th century as games were used to explore issues including urban planning and health care systems.²⁸ Authors stress that all of these applications can be found today, both in the US and abroad.

In addition to such historical approaches that demonstrate the value of gaming, a number of leaders and practitioners offer compelling narratives that often serve as the field’s touchstones about what gaming can contribute to national security research. These perspectives focus on

²⁴ The history of wargames and the purposes they have served in national security policy are relatively well documented. For a particularly excellent recent history, see: Caffrey, *On Wargaming: How Wargames Have Shaped History and How They May Shape the Future*. Important older histories include: Perla, *The Art of War Gaming: A Guide for Professionals and Hobbyists*; Thomas B. Allen, *War Games : The Secret World of the Creators, Players, and Policy Makers Rehearsing World War Iii Today* (New York: McGraw-Hill, 1987); Wilson, *The Bomb and the Computer*; Mayer, “The Gaming of Policy and Politics of Gaming.”

²⁵ Caffrey, *On Wargaming: How Wargames Have Shaped History and How They May Shape the Future*.

²⁶ A more detailed history of the use of games in this period at RAND will be offered in Chapter 9.

²⁷ The relationship between commercial and professional games is particularly well covered in: Caffrey, *On Wargaming: How Wargames Have Shaped History and How They May Shape the Future*; Perla, *The Art of War Gaming: A Guide for Professionals and Hobbyists*.

²⁸ For a concise description, see: Mayer, “The Gaming of Policy and Politics of Gaming.”

different aspects of the value of gaming—their potential to influence policy, their ability to provide insights on often opaque aspects of decisionmaking, and their impact on participants.

Current conversations include testimonies of key senior leaders who have advocated for the use of wargame to shape defense decisionmaking. Former Deputy Secretary of Defense, Bob Work, and Vice Chairman of the Joint Chiefs of Staff, Gen. Paul Selva, articulated a strong vision for the ability of games to support defense innovation:

Wargaming is one of the most effective means available to offer senior leaders a glimpse of future conflict, however incomplete. Wargames provide opportunities to test new ideas and explore the art of the possible. They help us imagine alternative ways of operating and envision new capabilities that might make a difference on future battlefields.²⁹

In other words, Work and Selva argue games are valuable because they provide an opportunity for new ideas to be generated and socialized in the department. Such innovation is critical in times of strategic uncertainty—so, for example, as the DoD looked to respond to the rise of peer competitors after a long period of focus on the wars in Iraq and Afghanistan, games filled a critical role supporting decisionmakers. In other words, the support games could provide to senior leaders is what defines their worth and so justified increased resources and attention to them.³⁰

Another perspective focuses on the ability of games to tackle questions that few other methods can shed light on. Nobel Prize winner Thomas Schelling argued that games have the unique property of studying the interactions between different decision centers and so enables the study of interactions in decisionmaking. This makes games well suited to studying communications, intentions, perception and misperception, signaling and other such issues.³¹ He also highlights several incidental benefits of games, such as creating an opportunity for information exchange leading to useful discoveries of both people and information players or analysts might not have otherwise been exposed to. Often these benefits occur later in participants professional lives, when ideas and contacts from past games prove relevant to current problems.³² He also highlights that games can produce “useful principles” of frequent trends in human behavior,³³ similar to the type of insights granted by game theory models such as the Prisoner’s Dilemma.

²⁹ Robert Work and Paul Selva, "Revitalizing Wargaming Is Necessary to Be Prepared for Future Wars," *War on the Rocks*, December 8 2015.

³⁰ Robert Work, Memorandum, February 9 2015.

³¹ 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, P-7719, 1991). p 32

³² Ibid. pp 24-25

³³ Ibid. p 30

A final view focuses attention on games' unique effect on their participants. For example, Peter Perla and Ed McGrady emphasize the role of the players in the game. According to these authors:

wargaming's power and success (as well as its danger) derive from its ability to enable individual participants to transform themselves by making them more open to internalize their experience in a game...there is an undercurrent of something less tangible than factors or models that affects fundamentally the ability of a wargame to transform its participants...a particular connection to storytelling.³⁴

Put differently, games are valuable because of their power for “sharpening and refining the stories we tell ourselves”³⁵ sometimes by generating new stories, sometimes by socializing a story more broadly within the department. The perspective argues that the value of games is not in the facts they create but rather in shaping how we understand facts.

While these descriptions have been highly influential in shaping how gamers define the value of their work, taken together these works inspire, but do not necessarily provide a clear, systematic vision of the appropriate application of games to national security policy analysis. These works do not present a systematic description of what problems games are, and are not, suited to address. In turn, that makes it easy to generate rules of the road about what questions games should and should not tackle. The next sections focus on existing attempts to define the bounds of what games can achieve in support of research.

Typologies of Applications of National Security Policy Games

In parallel to these influential perspectives on the value of gaming, other work has tried to define what games can contribute by mapping out the different types of games. Works describing these typologies are generally short (frequently a matter of a few sentences, rarely longer than several pages), and somewhat unsatisfying as guides to game design. First, the very short length of these discussions prevents authors from spelling out the implications of the typologies on game design and analysis. While other frameworks do offer more specific definitions, designers have generally opted to create their own typologies without reference to past efforts, creating a sizable number of overlapping categorizations with ambiguous relation to one another. As a result, it is often not clear what is at stake in using any particular set of terms, either in terms of practical game design work or in terms of advancing the method from a theoretical perspective.

One group of typologies focuses on the nature of the problem that is being examined, and how tractable it is to analysis. For example, one can divide games between those that deal with “analyzable” and with “non-analyzable problems”—arguing that problems that can be decomposed and for which optimal solutions can be found produce a different type of learning

³⁴ Peter P. Perla and ED McGrady, "Why Wargaming Works," *Naval War College Review* 64, no. 3 (2011).

³⁵ ED McGrady, "Getting the Story Right About Wargaming," *War on the Rocks*, November 8, 2019.

than games for which the rules, moves, and outcomes of play cannot be anticipated in advance. As early RAND research framed this division:

Learning, in the former case, has to do with handling of simple inputs in complex ways; in the latter, with handling complex inputs in simple ways. In the analyzable game universe, the point is to discover hidden strategic implications involved in repeating and combining simple elementary moves. In the non-analyzable game universe, it is to discern the ‘strategic’ pregnancy of the immediately given situation. In the analyzable game, one creates strategic opportunities, in the non-analyzable game, one uses them.³⁶

While this framing is conceptually attractive, in practice policy problems are, more often than not, complex. While a team might opt to define a narrower piece of the problem in order to make it more tractable to analysis, that is a matter of how the problem is scoped by the team rather than something intrinsic to the phenomenon.³⁷ In other words, the distinction being drawn reflects a decision about scope, rather than the type of knowledge to be produced.

An alternative approach seeks to define the purpose of the game from the perspective of how the information produced contributes to different Department of Defense concerns such as concept development, capabilities development, science and technology foresight, senior leader engagement and operational decisions, and training and education.³⁸ Such typologies are often attractive to sponsors and can be a helpful means of collecting together games dealing with similar topics. However, these categories are quite broad and thus will contain games with quite different issues. As a result, there is a great deal of diversity of game design within any of these DoD concerns so do not serve as particularly helpful guides to the designer. Furthermore, the specific research question that may fall into any of these categories may or may not be appropriate to game. A game to stress test potential weaknesses in a proposed concept would be appropriate while a game to “validate” a concept would not. A game to explore how a new technology might change adversary perceptions would be highly appropriate. A game to measure the effects on numbers of casualties from the use of a new armor would not. In short, these categories are potentially misleading to sponsors and consumers of games by suggesting much broader research programs than games can effectively support.

The majority of these typologies are oriented around the analytical tasks that a game might be asked to support.³⁹ For example, according to longtime RAND game designer Milton Weiner, games are used for 1) organizing knowledge held by a range of researchers, 2) research and

³⁶ Paul Kecskemeti. "War Games and Political Games." (Santa Monica, CA: RAND Corporation, D-2849, 1955). p 7

³⁷ Horst W. J. Rittel and Melvin M. Webber, "Dilemmas in a General Theory of Planning," *Policy Sciences* 4, no. 2 (1973).

³⁸ Yuna Huh Wong et al., "Next Generation Wargaming for the U.S. Marine Corps: Recommended Courses of Action," (Santa Monica, CA: RAND Corporation, RR-2227-USMC, 2019). pp 5-9

³⁹ These frameworks all include training, education, or both, which I have omitted from this discussion since these applications fall outside the bounds of the dissertation.

evaluation of what factors are important and how they related to one another, and 3) theory building.⁴⁰ In the case of research games, Weiner further defines three subtypes. The first treats games as an opportunity to observe the course of events in order to suggest what factors and relationships are particularly important. The second, modeled on experiments, depends on the power of comparison to evaluate the effect of a single change in context on the outcomes of the game. The third plays out a particular plan, policy, or weapon to get a sense of its strengths and weaknesses in a particular situation.⁴¹ A similar framework from the same period instead lists forecasting, innovation and strategic inventiveness, and the revelation of poorly understood dynamics ripe for further study as the primary purposes of games.⁴²

More recent additions to the literature also echo these themes. For example, Ed Parson makes a distinction between games for experiments, to promote creativity and insights, and for the integration of knowledge, not all of which he considers likely to produce useful knowledge.⁴³ Peter Perla divides research games into three classes: developing or testing strategies and plans, identifying issues, or building consensus among participants.⁴⁴ Two other frameworks that do somewhat similar work are Graham Longley Brown's division between games to understand, to generate insights, and evaluate⁴⁵ and Stephan Downes-Martin's categorization of experiential, comparison, and analytic games.⁴⁶ Without necessarily disagreeing with the types identified in these texts, the distinctions and differences among them are underdefined in current works. More fundamentally, a sense of what is at stake in game design is missing—in other words, how do you design a game given that you want to achieve these ends? This monograph is dedicated to answering these questions.

The Limits and Misuses of Policy Gaming

The flip side of the debate over the appropriate applications of games is a similarly heated discussion about what applications games are *not* suited to. On one hand, this is not surprising—all methods are better suited to answering some questions than others, and the limitations of games are often shared with other approaches. Discussing these limits is important for developing norms within any field. On the other, the depth of concern expressed about games,

⁴⁰ 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

⁴¹ Ibid. p 28

⁴² 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

⁴³ Edward Parson, "What Can You Learn from a Game?," in *Wise Choices: Decisions, Games, and Negotiations*, ed. Ralph L. Keeney Richard J. Zeckhauser, James K. Sebenius (Boston: Harvard Business School Press, 1996).

⁴⁴ Perla, *The Art of War Gaming: A Guide for Professionals and Hobbyists*. p 181

⁴⁵ Longley Brown, *Successful Professional Wargames: A Practitioner's Guide*. pp 89-91

⁴⁶ Ibid. p 91.

and the disagreements within the field about how games can be useful, are extensive and troubling. In many cases, these debates boil down to several key concerns. The first is that researchers are pushing games to answer questions to which they are not suited—that is, that games have limits which are not being observed. The second is that games are designed and executed in such a way as to undermine their ability to answer the research question at hand—in short, games that *could* be successful are mis-designed, mis-executed, and mis-analyzed and so fail to achieve their potential. Regardless of the cause, the common theme is that a greater-than-acceptable number of games fail to meet their objectives and do not contribute productively to the enterprise of policy research and analysis in the department.

A core set of critiques centers on the artificiality of games—since both the decisionmakers and environment are synthetic, they risk presenting compelling narratives which may not bear any resemblance to reality.⁴⁷ Some of these limitations are shared with modeling and simulation efforts, but those related to human players bare special mention since they are a key element of a game that differentiates games from other techniques. Humans are “playing” at making decisions in artificial environments are inevitably different then the behavior of decisionmakers in the real world, facing real stakes. The question then becomes how these artificialities effect what can be learned from a game. As early as the 1950s RAND gamers expressed concerns about the potential impact of the artificialities, arguing:

There is clearly a difference between ‘mere playing’ (an activity which leaves the welfare of the participants largely unaffected except insofar as they derive enjoyment from the play activity as such) and ‘fighting in earnest,’ where the welfare or existence of the participants depends on the outcome of the ‘game.’⁴⁸

Particular concern has long been expressed about the quality of role-playing of adversaries, often referred to as the “red” teams.⁴⁹ While defenders of games argue that these artificialities are no more limiting than those present in a range of other research techniques,⁵⁰ the problem of how “real” results of a game can possibly be if players are “merely playing” bedevils those seeking to use games for analytical purposes.

A similar debate has been waged over the ability of games to support forecasting and prediction. While warnings against treating game results as predictive are ubiquitous, gamers are also quick to cite Admiral Nimitz’s assertion that the games played during the interwar years had prepared the Navy for everything they saw in the World War II Pacific campaign except the kamikaze pilots.⁵¹ Those warning of the inability of games to predict often highlight two key

⁴⁷ Levine, Schelling, and Jones, "Crisis Games 27 Years Later : Plus C'est Deja Vu."

⁴⁸ Kecskemeti. "War Games and Political Games." p 1

⁴⁹ Wilson, *The Bomb and the Computer*. pp 60-64

⁵⁰ Levine, Schelling, and Jones, "Crisis Games 27 Years Later : Plus C'est Deja Vu."

⁵¹ Caffrey, *On Wargaming: How Wargames Have Shaped History and How They May Shape the Future*. p 81

points. The first is that games present a plausible “future history”⁵² of events—but the plausibility of a scenario is not sufficient to inform decisions that must hold up under a reality that may not match the scenario.⁵³ This concern is compounded by a second concern—that the immersive nature of games is “seductive.”⁵⁴ After vividly experiencing one potential future, players may be inclined to inflate its likelihood.⁵⁵

This tension represents a fundamental challenge for analysis—to be analytically useful to decisionmakers, games must illuminate causal relationships that can provide a useful guide to navigating future decisions. That is, they must at some level produce relevant understanding that can be transported into the future. However, at the same time, games only present a small number of specific futures—usually just one. Often, designers select futures seen as particularly dangerous—if we consider our expected distribution of potential futures, we are picking from an extreme, rather than the central tendency of our prediction. In theory, this dichotomy is easily managed--games illuminate indicative⁵⁶ patterns and trends that can be transported to other contexts, but the specifics of the game context and events should not be expected to appear and so games should not be treated as a precise prediction of the future. However, in practice the seductive nature of games makes this division hard to police and thus remains a subject of concern.

Another common tension is between different approaches to managing the complexity of the policy problems games represent. One approach seeks to break problems down into component parts—often driving studies to consider narrower, more technical problems that are more tractable. But to consider larger issues, one must either combine such detailed representations or simplify the dynamic at play to the point of absurdity.⁵⁷ The tendency towards more detail is particularly problematic in game design because more often than not, games are run early in the process of research, before the boundaries and components of a phenomenon are well understood. As a result:

We must be careful... that our detail and complexity are compatible both with our knowledge of the real world and with the purpose of the game. Otherwise we run the risk of specifying a number in the third decimal place when we are ignorant of whether the number is positive or negative.⁵⁸

⁵² Harvey A. DeWeerd, "A Contextual Approach to Scenario Construction," (Santa Monica, CA: RAND Corporation, P-5084, 1973).

⁵³ Levine, Schelling, and Jones, "Crisis Games 27 Years Later : Plus C'est Deja Vu." pp 3-4

⁵⁴ Ibid. p 1

⁵⁵ Ibid. p 4

⁵⁶ A phrasing I have adopted from my colleagues Stacie Pettyjohn and Becca Wasser.

⁵⁷ This concern is not unique to gaming. For well-known examples of related concerns from modeling and simulation, see: Paul K. Davis, "The Base of Sand Problem : A White Paper on the State of Military Combat Modeling," (Santa Monica, CA: RAND Corporation, N-3148-OSD/DARPA, 1991).

⁵⁸ R. D. Specht, "War Games," (Santa Monica, CA: RAND Corporation, P-1041, 1957). p 6

Conversely, games that are too simple may produce erroneous results or results that are so general they are unproductive for research purposes. A variation of this latter point is the complaint that games often simply replicate conventional wisdom and thus produce insights that are trivial at great expense.⁵⁹ A game designer must therefore strike a balance—too simple, and you are liable to not produce insights; too complex, and you'll offer false precision and risk missing the forest for the trees. Depending on the research topic, the distance between the two failure modes can be quite narrow.

Individually, these limitations are problematic, but taken together they raise a specter of the dramatic abuse of games.⁶⁰ The potential for targeted simplifications of an artificial environment that generates highly engaging results can make it all too easy to design a game to provide evidence for a predetermined outcome. If games are used to inform major defense decisions, the stakes for those who oversee the gaming process may be high enough to encourage fraud.⁶¹ Game design is inherently a series of many small choices about how to design the environment, rules, and actors of the game which can enable manipulation during design and execution, as well as game analysis, to shape results. As a result, games have received a troubling reputation as easy to manipulate.⁶² Game designers are quick to defend games, noting that the potential for abuse is not limited to games but rather is shared with many other approaches. Nonetheless, the concern is prevalent enough to effect the credibility of games, and so must be recognized.

In large part as a result of the potential for manipulation, there is often grave concern about using games for hypothesis testing, a use that many gamers reject absolutely as not appropriate.⁶³ However, like the issue of prediction, there is a tension between this common objection and the prevalent use of games for tasks such as course of action analysis⁶⁴ which aims to demonstrate the viability (or lack thereof) of a proposed scheme – which itself is a form of hypothesis testing. Some gamers have suggested the practice of “course of action falsification”⁶⁵—that is, not trusting games to provide strong support for hypotheses but recognizing their usefulness in

⁵⁹ Levine, Schelling, and Jones, "Crisis Games 27 Years Later : Plus C'est Deja Vu." pp 10 and 14-15

⁶⁰ Stephen Downes-Martin is conducting a fascinating line of research on how to design manipulative games. See: Stephen Downes-Martin, "Group Dynamics in Wargames and How to Exploit Them" (paper presented at the Connections North 2019 Wargaming Conference, Montreal, CA, 2019 of Conference). and "Preference Reversal Effects and Wargaming" (paper presented at the Connections North 2020 Wargaming Conference, Montreal, CA, 2020 of Conference).

⁶¹ "Your Boss, Players and Sponsor: The Three Witches of War Gaming," *Naval War College Review* 67, no. 1 (2014). p 32

⁶² For one frequently cited example, see: Micah Zenko, "Millennium Challenge: The Real Story of a Corrupted Military Exercise and Its Legacy," *War on the Rocks*, 2015.

⁶³ Levine, Schelling, and Jones, "Crisis Games 27 Years Later : Plus C'est Deja Vu." p 14

⁶⁴ Staff, *Joint Publication 5-0: Joint Planning*.

⁶⁵ Tom Mouat, personal communication, November 2019.

identifying failure modes as a potential compromise. But this has not been consistently adopted and thus the tension stands.

Beyond the disagreement about where games should not be used lies the question about more avoidable errors in the use of games. There is often very little distance between “reasonable” and “problematic” uses of games so the details of design, execution, and analysis become critical to maintaining quality. Designers and sponsors have long argued that games are not living up to their promise due to errors in design, execution, and analysis.⁶⁶ In general, there is a concern that games are often under-structured and too formulaic⁶⁷ to connect to the analytical objectives of the work.⁶⁸ There are also persistent worries that individuals outside the game team can interfere with the game in ways that undermine its analytic power.⁶⁹ These concerns about the quality of games risks the support of senior decisionmakers—if leaders are not getting games they feel support decisionmaking, they will use other tools.⁷⁰

Much of the current debate among gamers has focused on the need to produce better game designers. Senior gamers have long flagged that there are not enough good designers to meet the current level of demand.⁷¹ Recommendations from the community have generally focused on the need to educate more gamers and position experienced gamers better to advise senior leaders.⁷² These efforts are important steps toward professionalization but may be insufficient to offer the improvement in game quality needed, if conducted under the current model. The next section lays out why current efforts in isolation are likely to be insufficient and argues that a new approach to understanding how games work is needed to push positive change forward in the field.

Building Better Games for National Security Policy Analysis

The absence of assessment tools prevents clear differentiation between well-designed, executed, and analyzed games that produce useful information and those that are misleading. Without the ability to consistently distinguish good analytic games from poor ones, poor games can influence decisionmakers when they reinforce what sponsors wanted to hear and good games can be ignored if they produce unwanted findings. However, cracking the problem of the lack of measures is more difficult than it first appears.

⁶⁶ An excellent discussion of common shortcomings is compiled in: Weuve et al., “Wargame Pathologies.”

⁶⁷ Peter P. Perla, “Now Hear This—Improving Wargaming Is Worthwhile—and Smart,” *Proceedings Magazine*, January 2016 2016.

⁶⁸ Phillip E Pournelle, “Can the Cycle of Research Save American Military Strategy?,” *War on the Rocks*, 2019.

⁶⁹ Downes-Martin, “Your Boss, Players and Sponsor: The Three Witches of War Gaming.”

⁷⁰ Jon Compton, “The Obstacles on the Road to Better Analytical Wargaming,” *War on the Rocks*, 2019.

⁷¹ Stacie L. Pettyjohn and David A. Shlapak, “Gaming the System: Obstacles to Reinvigorating Defense Wargaming,” *ibid.* February 18 2016. *ibid.*

⁷² Perla, “Now Hear This—Improving Wargaming Is Worthwhile—and Smart.”

Two competing perspectives to game evaluation exist. The first, and more common, treats games as an art form. Measures of successful game design are tied to the experience of the final product of the game. This approach has led to frustration amongst both game sponsors and game designers that many games of lackluster, or outright poor quality are produced. An alternative perspective turns to more analytical, scientific approaches to assess whether the goodness of design has been soundly guided by the process that was followed during research. While recognizing the potential pitfalls of such an approach, I argue a scientific approach offers much-needed tools to understanding whether a game is good. As a result, it offers the potential for far-reaching improvement in the quality of national security policy games.

What Makes a Game Good?

First, there is the fundamental issue of defining the quality (or qualities) by which a game is assessed. Here, the nature and purpose of games pose an immediate difficulty. Games are often used to examine the future for which by definition the “right” answer is not known at the time the game is run and assessed because there is not strong empirical evidence from real-world events to compare results too. This point has most often been made regarding nuclear war, perhaps most famously via the quip of a RAND researcher: “General, I have fought just as many nuclear wars as you have.”⁷³ The point is true for many other areas of war and conflict, particularly when considering future conflicts using weapons that may not exist even in the lab today. Indeed, as discussed earlier in this chapter, games are not in the business of offering specific predictions of what will happen but rather provide indicative information about important trends or raising questions that had not previously been considered. In fact, the most influential games often shape decisionmaking precisely in order to prevent the crisis of the game from ever occurring in the real world. As a result, accurately forecasting the future is not an appropriate measure of game quality.

Another obvious measure of a good policy game is one that causes policies to be improved. Unfortunately, whether this result occurs is very difficult to know in all but the rarest cases. In part, this is a problem shared by many types of analysis, since after all: “policy-makers could make policy without the benefit of analysis, and the policy might be good. We don’t really know if it’s good or not until we put it to the ultimate test.”⁷⁴ Implicit in this are two challenges to creating good measures. First, it is difficult to evaluate the consequences of the game (or any other analysis) for policymaking in most circumstances. In some cases, decisionmakers will explicitly call out the role of a piece of analysis in shaping their thinking. But too often, the connection is unclear. Is a study cited because it shapes the perspective of the decisionmaker, or simply because its conclusions support a position they had already come to? What if a participant draws on experience from a game they participated in years earlier to shape how they make a

⁷³ Fred Kaplan, *The Wizards of Armageddon* (Palo Alto, CA: Stanford University Press, 1991). p 254

⁷⁴ Wilson, *The Bomb and the Computer*. p 105

snap judgement? Tracking such effects is likely to be more expensive (and politically sensitive) than is practical. The issue is further compounded by the second problem: evaluating the “goodness” of a policy is notoriously difficult. Without dismissing the importance of such evaluator efforts, they sit well outside the problem set of gamers hoping to improve the value of their work, and so I do not treat them here.

Given these limitations, it makes sense to focus assessment on gauging the quality of the decisions made in the design, execution, and analysis of the game. However, the wargaming community lack consensus perspective on how to evaluate a good game design. 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 often-cited handbooks on the design of games stress the importance of linking the choice of design elements to the purpose of the game, since a “wargame’s objectives should be the principal drivers of its entire structure.”⁷⁵ 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.”⁷⁶ Experienced designer argue that is because there are an infinite number of ways to make these connections, there is no defined process for a designer to follow.

Leading members of the field have suggested that a key solution is to use expert game designers who have long track records of producing good design to assist in quality control.⁷⁷ However, others claim that this solution is unworkable thanks to the relative scarcity of expert designers and the difficulty of determining who is a true expert.⁷⁸ If experts have the ability to usefully assess the quality of games but there are not enough of them to be able to successfully assess the volume of work produces, one solution would be building a tool that can capture some of the knowledge being applied by expert assessors to aid less practiced designers.

While in practice, senior game designers clearly can and do make this link on a routine basis, they are currently not readily able to articulate a general theory behind this process. For example, when surveying expert gamers in an attempt to catalog cues they use in evaluation, I found that while there were strong heuristics available to judge the quality of a game during its executions, even experienced gamers struggle to articulate the markers of a good design or analytical product.⁷⁹ Respondents described a need to have a “match” between purpose and design, but

⁷⁵ Weuve et al., “Wargame Pathologies.” p 13

⁷⁶ Perla, *The Art of War Gaming: A Guide for Professionals and Hobbyists*. p 181

⁷⁷ “Now Hear This—Improving Wargaming Is Worthwhile—and Smart.” and McGrady, “Getting the Story Right About Wargaming.”

⁷⁸ Stacie L. Pettyjohn and David A. Shlapak, “Gaming the System: Obstacles to Reinvigorating Defense Wargaming,” *ibid.* February 18 2016. *ibid.*

⁷⁹ I fielded a survey in May and June of 2017 to expert members of the wargaming community and received 50 responses, most from highly experienced professionals. The aim of the survey was to collect information from gamers about how they conduct assessment before, during, and after the execution of a game. In addition, the survey

generally were not more concrete than characterizing the link as “understandable”, “logical” or “sensible” choices. Once again, we are left with the strong sense that the link between purpose and game design is highly salient but not well defined.

Two Competing Explanations for the Gap

Given that gaming has been used to support national security policy for decades, the continued persistence of this gap appears confusing. In fact, the lack of clear linkages between purpose and design has its roots in a divide within the community between practitioners who see the fundamental nature of gaming as an art and those believe that the foundations of design for research and analysis must be based in science. These two perspectives identify different ways to think about assessment and so offer different practical approaches. An artistic approach evaluates work based on “level of competency in the medium deemed appropriate by the community.”⁸⁰ In contrast, a scientific approach argues game assessment should focus on whether the process that was followed during research is sound. While historically, those holding games as an art have dominated the profession, advocates for a base of science has long been present. I argue this latter approach, if properly framed, offers much promise for improving the state of practice across the field.

Today, the majority of gamers think of their work as an art. This view tends to focus on games as a means of communication,⁸¹ and in particular, for storytelling about “what is possible, what is unexpected, what worked and led to victory, or what failed and led to defeat.”⁸² However, the preference to treat games as a craft has come with it serious concerns about game quality. Advocates of the artistic approach⁸³ and their critics⁸⁴ both raise concerns about number of poor games that are produced for defense audiences, and the inability of the community to effectively policy quality. Two issues are particularly key: the difficulty of identifying and training new gamers and the inability to develop effective means for the community to determine what is, and is not, competent practice.

included demographic information and general questions about game design practice to contextualize answers. Detail available in: Elizabeth M. Bartels, "Insights from a Survey of the Wargaming Community" (paper presented at the Military Operations Research Society Wargaming Community of Practice, Alexandria, VA, 2017 of Conference).

⁸⁰ Sawyer Judge, "The Wargaming Guild: How the Nature of a Discipline Impacts Its Craft and Whether It Matters" (Georgetown University, 2019). summarizing Howard Gardner, *The Arts and Human Development; a Psychological Study of the Artistic Process* (New York, NY: Wiley, 1973). p 117

⁸¹ Perla, *The Art of War Gaming: A Guide for Professionals and Hobbyists*.

⁸² McGrady, "Getting the Story Right About Wargaming."

⁸³ Perla, "Now Hear This—Improving Wargaming Is Worthwhile—and Smart."

⁸⁴ Compton, "The Obstacles on the Road to Better Analytical Wargaming."

The artistic approach tends to focus on the individual genius of the designer and their ability to recombine elements to striking effect. New designers looking for direction have long been frustrated, with early RAND gamers noting:

There is no theory of operational gaming, in the sense of a systematic discipline which tells you exactly what purpose gaming has, what rules you have to follow in setting up a game, and how you go about achieving the stated purpose on the game has been constructed. Gaming, in other words, has not yet attained the status of a scientific method but is still very much in the nature of a craft.⁸⁵

This often frustrates new gamers entering the profession who must navigate a guild system that can be of deeply uneven character. Too often, prior exposure to commercial board games are used as an indicator of potential talent, even though the purpose of games designed to entertain is radically different than games for research and analysis. This can result in researchers who might make for excellent designers not receiving the training they need.⁸⁶ A limited throughput of new gamers may be acceptable in periods during which few games are run, but under current conditions the demand for high quality games outstrips the supply of designers making training of new gamers a priority.⁸⁷

Additionally, determining the quality of games rests to a great degree on the judgement of designers, sponsors, and clients. In line with the practices of artistic research, proponents of the artistic view of gaming argue that the best hedge against bad games is for good designers to call out bad practices.⁸⁸ However, this solution has encountered a range of practical barriers, due to the nature of the gaming community without formal credentials or other means of defining who is, or is not a credible source of judgement. Problems discussed in the past include the limited number of gamers experienced enough to conduct these evaluations⁸⁹ and the limited ability of the community to effectively signal who is, and is not, competent to outsiders.⁹⁰ A survey of experienced members of the community also reveals potential flawed approaches to evaluation, including assuming the quality of a game based on who designs them (even though even the best gamers will admit to the occasional poor design), preferring games that are engaging for players even if the fact base of the game is insufficient for sound research, or approving of games that reify already held positions.⁹¹ This is far from a new problem—even in the 1970s wargamers

⁸⁵ Olaf Helmer-Hirschberg, "Strategic Gaming," (Santa Monica, CA: RAND Corporation, P-1902, 1960). p 1

⁸⁶ Elizabeth M. Bartels, "Building a Pipeline of Wargaming Talent: A Two-Track Solution," *War on the Rocks*, 2018.

⁸⁷ Stacie L. Pettyjohn and David A. Shlapak, "Gaming the System: Obstacles to Reinvigorating Defense Wargaming," *ibid.* February 18 2016. *ibid.*

⁸⁸ ED McGrady, "Getting the Story Right About Wargaming," *ibid.* November 8, 2019 2019.

⁸⁹ Stacie L. Pettyjohn and David A. Shlapak, "Gaming the System: Obstacles to Reinvigorating Defense Wargaming," *ibid.* February 18 2016.

⁹⁰ Elizabeth M. Bartels, "Building a Pipeline of Wargaming Talent: A Two-Track Solution," *ibid.* 2018.

⁹¹ "Short Insights from a Survey of the Wargaming Community."; Weuve et al., "Wargame Pathologies."

expressed concerns that: “Free-form gaming.... has few good practitioners and a product that is very hard to measure, making it extremely difficult to ascertain whether the art form has improved in the last few years.”⁹² Put simply, the commitment to an artistic approach to games has made it difficult to set clear standards about what makes for a good game and prevent charlatans from producing bad games that mislead decisionmakers, and this problem has endured for decades despite the efforts of many researchers to improve the state of the field.

This monograph argues for an alternative approach: that game design for research and analysis should be grounded in logics of inquiry adopted from science. This approach is not a call to strip the art from game design: truly excellent games are works of creative genius in which the designer is able to work within the limitations to create compelling results. Additionally, a scientific foundation will not ensure a successful game—outside factors, particularly sponsor and player disposition can derail even the best designed games. Instead, I argue that a clearly articulated, scientifically sound logic should form the foundation of game design. This work presents such an approach, and why I believe that adopting these practices has the potential to improve the practice of gaming for policy analysis.

To be clear, game designers have good reasons to worry about the application of “science” to games—too often the approach to science favored by the Department of Defense is rigidly formulaic and lacks the tools to focus on the fundamentally squishy phenomena of human decisionmaking that are the focus of games. The first issue leads to a concern that gamers will be forced to replicate a few, not very strong designs.⁹³ The second leads to fears that a “scientific” approach will attempt to turn games into something they fundamentally are not—standardized processes that produce predictive, quantitative analysis.⁹⁴ In parallel, game designers who have observed misuse of other analytical tools by charlatans argue that appeals to science are insufficient to prevent deception and poor quality work from being presented to DoD.

However, without dismissing the validity of these fears, it strikes me that science has not been given a fair shake in these debates. The social sciences have developed a rich set of approaches to conducting inquiry that is systematic but has also prized the ability to tackle difficult research question using creative research designs that make the most of limited or uncertain data.⁹⁵ Games share many characteristics of other approaches to analysis used in social science—including a focus on human decisionmaking, particularly in groups. As such, it seems that games should be amenable to a scientific logic of design in which games are judged by their

⁹² Martin Shubik and Garry D. Brewer, "Models, Simulations, and Games--a Survey," (Santa Monica, CA: RAND Corporation, R-1060-ARPA/RC, 1972). p 6

⁹³ Perla, "Now Hear This—Improving Wargaming Is Worthwhile—and Smart."

⁹⁴ McGrady, "Getting the Story Right About Wargaming."

⁹⁵ I am indebted to Sawyer Judge for an excellent discussion that clarified many of these point for me.

ability to develop designs that are logically sound.⁹⁶ Such standards would allow for clearer means of assessing good game design by stakeholders by insisting that the logic of design be 1) clearly understandable and 2) transparent to a range of stakeholders. While such tools do not prevent unethical researchers and disengaged sponsors from generating and accepting poor work, they offer a suite of tools for those that want to do better. If this argument is current, a scientific approach of the right type offers the ability to improve the quality of games by offering clearer standards for design.

It is important to be clear that the goal of a scientific framework is to serve as a necessary guide to game design. Yet, these considerations are not, in and of themselves, sufficient to produce best-in-breed game designs. Rather, this discussion is intended to set a minimum requirement for game designs to credibly support research. If a gamer cannot produce a credible explanation about why their game produced information that is fit for the game purpose, such results can be dismissed from consideration for analytic purposes. Individual designers are still left with the task of designing the best game available by making clever design choices that enable players to fully engage. Truly masterful games do have a great deal of artistry in their design -- but this approach contends that it must be artistry built on the foundations of a sound logic. This work focuses on describing this foundational element.

Towards a Scientific Approach: Defining Logics for Game Design

This monograph seeks to advance the case for a scientific approach to gaming—one that offers guardrails for designers, sponsors, and consumers of games, without negating the need for games to remain focused on human interactions and decisionmaking. In the following chapters I lay out a set of logical approaches to conducting research using games based in social science. Chapter 2 presents my overall approach to research, including the range and limits of the data available to me. Chapters 3 and 4 provide the theoretical grounding for my argument. Chapter 3 applies recent discussions of philosophy of science to games and shows how the tool aligns with three different approaches to scientific research. Chapter 4 presents four archetypes of information that can be generated by games to support research. Chapters 5-8 consider the practical ramifications of the theoretical approach, by describing the types of tradeoffs that are likely to dominate game design for each of the four archetypes I develop in Chapter 4. Chapter 9 presents a narrative history of gaming at RAND, describing how the types of information games produce at RAND has changed over time. I offer concluding thoughts and recommendations to game designers, sponsors, and consumers in Chapters 10.

⁹⁶ 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).

Chapter 2: Study Approach

Building on the existing literature detailed in the previous chapter, this study develops a framework of different types of games with the goal of articulating different potential logical connections between a policy analysis game's purpose and design. The first task was to determine the type of framework—or classification scheme—that would provide the most utility to game designers. After determining that a set of archetypes would be most productive, I then surveyed the available sources of data to populate the types. Because of limited records of games, it was most feasible to focus on expert validation rather than a traditional classification activity, to refine the framework. As a result, I opted to conduct a series of expert interviews in individual and group settings to gather feedback about the ability of the framework to capture expert practice.

Research Context

As is often the case with research, the availability of data was fundamental to shaping the research design of this project. In this case, the project required data on how games are designed in practice. In seeking to understand how game designers conduct their work, two broad categories of data were available: written descriptions of specific game designs, and direct discussion with practicing game designers about their process. Both types of data exhibited systematic limitations that stem from the policy environment games are designed in and the processes they are intended to support. This section reviews these limitations, both in the interest of situating the research design that follows, and to inform scholars not embedded in the community about some of the limitations of the available data.

Limits of Existing Information on Specific Games

While game reports make for tantalizing reading, there are major lacunae in our written record of specific games. Public reports on games represent only a small portion of the total games run and the percentage and nature of what is available is difficult to characterize. Security classification, sensitivity, and professional cultures that do not prioritize open publication limit the available game documents. When game reports are accessible, they are often difficult to locate and survey. Once identified, available written reports often include very little information about the overarching logic of research and how it contributed to design. In short, exactly the type of data needed to develop or test a traditional classification scheme focused on research design in games is often missing from our written records, and what is available cannot be treated as representative. This section summarizes the gaps in the written literature and offers some possible explanations for the limitations.

First, results of some games are never captured in a formal report. For example, games that are intended to inform short-term decisions or to provide experiential learning to participants may be documented using informal tools such as briefings, memos, or emails, rather than longer written documents. Generating formal reports requires time and resources that offices may be unwilling or unable to expend if they cannot be justified by an explicit need for a written product. These practices may produce some level of documentation of games but pose major barriers to consistent capture of materials by a researcher, since salient details may only exist in the memory of designers and key players. As a result, these games are often lost from the available sample.

When games are documented, records may not be publicly available. Sensitivities, including, but not limited to classification, prevent us from developing a good understanding of what reports might be available, and what types of biases might be introduced by looking at only this partial record. If games are used to inform future planning, then it makes sense that adversaries would be invested in discovering what topics are being gamed as a potential indicator of future decisionmaking, available information on the adversary's tactics and capabilities, or emerging capabilities. Games that involve analysis of domestic or alliance dynamics may also be sensitive if they reveal weaknesses which could be exploited by an adversary or damage critical relationships. Information about these games is often restricted as a result. Other games are not proactively released because it would impinge on the ability to conduct future games. For example, games that reveal mistakes or poor judgement might be embarrassing to participants and sponsors if they were released, raising the stakes of participation and removing the benefits of games as a forum for low-risk experimentation. While sensitive reports may become accessible over time, or if directly requested, these processes are idiosyncratic and unpredictable. As a result, many game reports are not publicly accessible.

Even when game documentation is appropriate to disseminate publicly, often it is not formally published, and thus is extremely difficult to locate. In part, this is consistent with broader practices in the government and related industries where documents may be produced but only circulated internally. The reasons behind this practice vary but can include limited communities of interest best reached through direct circulation, barriers to accessing public forums such as publication reviews, and limited incentives for public engagement that discourage individual researchers from going through the additional effort of formal publication. This "gray literature" is sometimes available to researchers through direct contacts or lucky happenstance but is difficult to find and thus impossible to survey comprehensively. While recent efforts have been made to make this work more accessible, there are serious limits to the current system. For example, the Department of Defense has recently begun collecting game records in a central repository. While this system includes information about some 700 games,⁹⁷

⁹⁷ Garrett Heath and Oleg Svet, "Better Wargaming Is Helping the U.S. Military Navigate a Turbulent Era," *Defense One* 2018.

it is only available on restricted DoD systems, and no equivalent program is available to catalog unclassified reports across organizations.⁹⁸ As a result, this body of work remains difficult to reliably and equitably access.

When written reports are available, it is striking how few contain discussion of the design process. Study of current gaming practices suggests that the information contained in written reports varies widely due to different purposes, sponsor requirements and preferences, and business practices.⁹⁹ Several different phenomena could explain this gap. One concern is that many games are run by contractors or consultants who depend on repeat business to remain financially viable. For these designers, publishing detailed descriptions of games represents a potential loss of future business—a concern that the client will ask “if I can run the game based on the report, why do I need to hire the game designer?” Another issue is that the consumers of information from games are often far more interested in the substantive insights from the game than the methodological details of how the information was produced. This tendency to want to focus on the substance rather than the technical process is shared in other analytical disciplines. However, the lack of common technical, concise language to describe the logic behind design choices may exacerbate the tendency to short-change this reporting in game results, since there is little pressure from the gaming methods community to include the information and few ways to do so that are consistently comprehensible if designers opt to do so. When technical details are provided, they often focus on specific components of the game, such as the adjudication system, rather than the overarching logic of research and how it contributed to design. In short, exactly the type of data needed to develop or test a traditional classification scheme focused on research design in games is often missing from our written records.

Because of the wide-ranging set of reasons game designs are not documented well in archival sources, it is difficult to characterize the biases that shape what is available. For example, games on pressing topics of the day may require the use of sensitive information that restricts circulation of the report, but it also may be more important to sponsoring offices to publish reports that can influence different constituencies. For example, a game examining the application of emerging technology to military problem sets might be especially sensitive because of the technical projections used to represent the technology, and thus closed to the public, or games may be very open to try to attract academics and industry attention to gain access to cutting-edge research being conducted outside the military. Publication politics may depend on interpretation of guidance or office culture, which are highly dependent on personalities or specific context. All of these considerations are difficult to discern from the outside. As a result, while it is clear that any accessible sample is not likely to be a representative sample of games run to support policy, it is difficult to characterize how the sample is biased,

⁹⁸ Ivanka Barzashka, "Wargaming: How to Turn Vogue into Science," *Bulletin of the Atomic Scientists*, 2019.

⁹⁹ Wong et al., "Next Generation Wargaming for the U.S. Marine Corps: Recommended Courses of Action."

and must be careful in any analysis to recognize the presence of substantial, difficult to characterize, gaps in the written record.¹⁰⁰

Limits of the Expert Community

For some of the same reasons it is difficult to account for what games are run to support national security policy, it is also difficult to characterize the community of gaming practitioners. Because there is no common academic training, certification for wargaming required for practice, or professional credential (such as a bar or board), there is not a central listing of professional wargamers. Many different organizations practice wargaming, and not all are known to one another. This problem is compounded by the dominance of contractors in the field, which increases the number of organizations involved in gaming. The general lack of public publishing exacerbates this lack of awareness. As one prominent convener put it, wargamers seeking to engage with the broader field must “navigate an archipelago of excellence”¹⁰¹ in which many different centers work in partial isolation from one another.

Several initiatives are designed to create a stronger community, but their impact is imperfect. The two most established are the Military Operations Research Society’s (MORS) annual conference’s wargaming track and the Connections Wargaming Conference. Both serve as broad hubs for the gaming community. These conferences provide regular forums for gamers to present research and to network. However, because of the time and expense associated with conferences, they attract only a portion of the field. Both have spun off other efforts, including the MORS community of practice series of lectures and international Connections Conferences, that aim to make the network more accessible. Additionally, efforts like the Wargaming Repository that try to track games from a central position in the Pentagon have the potential to offer a different hub to identify gamers.¹⁰² However, submission of games to the repository is not mandatory, and access to the repository is restricted and has limited utility. All of these efforts bias toward U.S. DoD game designers, with uneven participation from other sectors and nationalities. Other biases may well be present—for example, contractors must justify charging overhead, making them less likely to participate in time-consuming events unless they can demonstrate benefit to supervisors or sponsors. As is the case with publications, the scope and extent of these biases is difficult to predict—because the full shape of the community is unknown, it’s hard to know who is missing.

¹⁰⁰ For an example of thoughtful treatment of bias in a sample of games, see: Pauly, “Would U.S. Leaders Push the Button? Wargames and the Sources of Nuclear Restraint.”

¹⁰¹ Tim Wilkie, National Defense University

¹⁰² Heath and Svet, “Better Wargaming Is Helping the U.S. Military Navigate a Turbulent Era.”

Study Approach

Instead of attempting to draw conclusions from categorizing limited empirical data, I developed an alternative approach which depended on expert validation and illustrative examples to inform iterative refinement of the framework. This entailed moving through four phases: 1) understanding policy game design in scientific terms, 2) framework design, 3) expert validation, and 4) example elicitation. The following sections lay out my process for undertaking each, making special note of the limitations and constraints that shaped my research design and the resulting limits that may impose on my results.

Phase 1: Understanding Policy Game Design in Scientific Terms

Before diving into the substance of the framework, I first found it productive to consider where games might fit within existing understandings of science. As described in Chapter 1, existing discussion of the role of games tends to eschew formal consideration of philosophy of science, ontology, and epistemology, and the limited texts that do exist have not seen wide circulation or sustained engagement. In some cases, this is because the writer takes the position that wargaming is an art.¹⁰³ In others, it is because a practical field like gaming has limited patience for pure philosophy. Furthermore, the diversity of intellectual backgrounds among games limits shared access to precise language and common definitions to use in such debates. As a result, little current writing on gaming engages with core concepts from philosophy of science that seek to understand what can be learned from observing the world.¹⁰⁴

Given this gap in the literature, any attempt to develop a more systematic approach to games based on science would benefit from a careful understanding of what science is. First, I started with examinations of science drawn from the social sciences (and international relations in particular) because they share the focus on humans, their decisions, and the organizations they build that characterize the topics of national security policy games. This literature lays out not one approach to science, but several distinct traditions. I then turned to the existing literature by game designers describing their practice, looking for evidence that game designers follow one or more of the philosophical commitments laid out in the literature on science. I used this work as illustrative examples to demonstrate the viability of not one type of scientific gaming, but several different types. I also considered how the insights from other discussion of social scientific research might apply to how we learn from games.

Phase 2: Initial Framework Design

In the next stage of research, I developed a framework of game types, drawing on both the existing literature and my own experience as a game designer. Completing this task had three

¹⁰³ Most notably Perla, *The Art of War Gaming: A Guide for Professionals and Hobbyists*.

¹⁰⁴ Barzashka, "Wargaming: How to Turn Vogue into Science."

fundamental elements: establishing the scope of the framework (that is, what aspects of games were to be typed), deciding the format of the framework (that is, how to most usefully develop categories), and populating the framework with initial content. While these evolved to some degree in later stages of research, this initial framing did much to influence the final form of the framework.

Scope

The goal of this work is to explore if different logics could be found to connect game purpose to the design choices to the eventual findings and, if these logics could be identified, present them in a practical way so that they are accessible to designers, sponsors, and consumers. Initially I drew heavily on the literature on logics of inquiry common in the social sciences. The concept of a “logic” is helpful because it points to the researcher’s responsibility to develop a persuasive argument that connects the research process to the credibility of the resulting findings. This term is used to refer to the argument that researchers make to explain why the study they conduct can inform the reader about the real world generally, and about causality of key phenomena specifically.¹⁰⁵ While not a settled point, most authors agree that more than one logic is available to research designers, and that different logics will be appropriate depending on the type of evidence available and the type of information the researcher wants the study to produce. However, early interviews clearly demonstrated that the language of “logic of inquiry” or “logic of design” was not generally and consistently understood by game designers. This monograph undertakes the task of explaining logics, and their potential value, in the hopes that such a framework will be of use.

In searching for an alternative, more understandable language to describe the load that I wished the typology to support, I moved through several iterations before finally settling on one that differentiated the information that games are designed to generate. One concept was typing the purpose and objectives of the game. This would be consistent with existing game literature.¹⁰⁶ While sensible in theory, in practice the purpose and objectives of the game are set in consultation with the game sponsor. This results in the use of unclear language and multiple

¹⁰⁵ For some key texts expounding on this point, see: Gary King, Robert O. Keohane, and Sidney Verba, *Designing Social Inquiry: Scientific Inference in Qualitative Research* (Princeton, NJ: Princeton University Press, 1994); Alexander George and Andrew Bennett, *Case Studies and Theory Development in the Social Sciences* (Boston, MA: MIT Press, 2005); Derek Beach and Rasmus Brun Pedersen, *Causal Case Study Methods: Foundations and Guidelines for Comparing, Matching, and Tracing* (Ann Arbor, MI: University of Michigan Press, 2016); Stephen L. Morgan and Christopher Winship, *Counterfactuals and Causal Inference: Methods and Principles for Social Research* (New York, NY: Cambridge University Press, 2007); Henry E. Brady and David Collier, eds., *Rethinking Social Inquiry: Diverse Tools, Shared Standards* (New York, NY: Rowman & Littlefield Publishers, Inc., 2004); Gary Goertz, *Multimethod Research, Causal Mechanisms, and Case Studies: An Integrated Approach* (Princeton, NJ: Princeton University Press, 2017); Jason Seawright, *Multi-Method Social Science: Combining Qualitative and Quantitative Tools* (New York, NY: Cambridge University Press, 2016).

¹⁰⁶ For example see: Parson, “What Can You Learn from a Game?.”

objectives.¹⁰⁷ Designers often do considerable work to translate from the “official” objects recorded in the game documentation to their own understanding of what is desired from the game, which then drives design.

Instead, I have opted to focus not on the purpose and objectives of the game itself, but rather on the desired end point of the project of which the game is an element—what information needs to be produced in the game.¹⁰⁸ While in some sense the difference is semantic since the purpose and objectives of the game should state what the desired outcomes of the game are, this backward logic of starting with the desired end point resonated with interviewees. For example, this approach mirrors current teaching that data-capture plans should be based on what information needs to emerge from the game. The game is then designed around these requirements.¹⁰⁹ As a result, I opted to frame the archetypes around the types of information the game is designed to produce, rather than focusing on purpose and objective.

One additional change to the scope of the framework that occurred in the process of the research project was removing games aimed at education or communication from the typology. I discovered that any game that had both educational and research objectives shared the same characteristics as a game that only had the research objective. The presence of the educational objective provided an “out” to designers to declare a game successful even if it did not produce the desired information. In other words, no additional requirements were imposed on design as a result of adding educational objectives. As a result, I opted to remove it from the framework to more clearly focus attention on designing games to produce desired information.

Format

In parallel to scoping the purpose of the framework, it was also necessary to identify what type of framework would be most helpful. Several different types of classification frameworks are regularly employed in policy analysis, the most common of which are taxonomies and typologies. However, I determined that the slightly less common approach of archetypes had several advantageous characteristics, and thus opted to use that for my framework.

Generally, categorization systems are developed in one of two ways, either top-down by defining theoretical distinctions that can then be tested by sorting the population of interest, or bottom-up. While terminology differs somewhat between fields, the distinction is usually made between typologies, which are driven by theory, and taxonomies that start with an empirical base to define categories. Both types of systems have proven helpful in policy contexts, though the empirical bases of taxonomies may be somewhat more defensible in policy contexts.¹¹⁰

¹⁰⁷ Downes-Martin, "Your Boss, Players and Sponsor: The Three Witches of War Gaming."

¹⁰⁸ This insight arose during an interview with Ed McGrady, to whom I am indebted.

¹⁰⁹ Interview with Jeff Applegate, game designer, Monterey, CA, August 2018.

¹¹⁰ Kevin B. Smith, "Typologies, Taxonomies, and the Benefits of Policy Classification," *Policy Studies Journal* 30, no. 3 (2002).

However, for the grounded approach to be reasonable, there must be a diverse, representative set of observations to class. As described above, the records of games are too incomplete to provide such a basis, making a taxonomic approach impractical for this study.

However, on closer examination, several aspects of traditional typologies are also problematic in the context of gaming. When well designed, typology categories should be mutually exclusive and comprehensively exhaustive—that is, any given item to be classified can be placed in one, and only one category. However, much of the gaming literature, as well as practical experience, emphasized that games have multiple objectives. As a result, a more flexible framework seemed more appropriate, and ultimately more useful, to the task of classifying games.

Archetypes -- a variant of typologies -- seems more promising. Used in fields ranging from philosophy, psychology, and literary criticism, archetypes feature “ideal forms.” In the field of policy analysis, they are perhaps most closely associated with systems thinking, which defines “system archetypes”¹¹¹ as broad patterns of behavior that reoccur in many different contexts. Beyond these applications, archetypes have a long history of use in policy analysis as a tool for communicating complicated results to broad audiences, suggesting that the approach may help make my findings more accessible to non-expert audiences, particularly game sponsors. Because archetypes are ideals, the expectation is that few if any observed examples will be fully described by the archetype. Rather, it’s a tool for identifying patterns, which may occur in combination, and more or less strongly, across cases.

The concept of pattern detection seemed a particularly good fit given the initial survey I conducted of game designers. When surveyed, expert gamers describe a well-designed game as one for which the game design “matches” the purpose of the game but they fail to explicitly describe what this match might consist of.¹¹² The limited language that experts use to describe the design process has interesting parallels in the literature on expert decisionmaking, and in particular the Recognition-Primed Decision Making (RPD) model.¹¹³ This model centers the role of pattern recognition in decisionmaking—that is, rather than comparing competing options or choices, experts draw on a bank of experiences to quickly assess what is typical about a decision and its context and then use this determination, as well as anomalies in the decisionmaking context, to develop a viable plan of action. In other words, the RPD model would posit that when experienced gamers are presented with a new game’s purpose, they quickly identify what games the new project is similar to, based on their experience and how the new project may diverge from “typical” projects, and use that pattern recognition as a basis for making design decisions.

¹¹¹ Peter M. Senge, *The Fifth Discipline: The Art and Practice of the Learning Organization*, Revised and Updated edition ed. (New York, NY: Doubleday, 2006).

¹¹² Bartels, "Short Insights from a Survey of the Wargaming Community."

¹¹³ Gary Klein, *Sources of Power: How People Make Decisions* (Cambridge, MA The MIT Press, 1998). pp 147-153.

Archetypes play into exactly that pattern by providing examples for designers to compare to that may be outside of their own experience because they represent a range of extreme types. A designer can use archetypes to compare to the current situation and determine on what dimensions the problem is similar to any given type and use that information to refine a design. It also gives designers a common language to talk about their games—designers may not have observed the same set of games, but the archetypes provide a shared pool of references that they can utilize to aid communication.

Populating the Initial Framework

I generated the initial framework drawing on several key sources of information. First, I surveyed the existing literature (surveyed in chapters 1) for existing frameworks and ideas. Second, I considered the interaction of those concepts with my own game design practice. This included careful consideration of how well games I have designed fit into existing categories, as well as careful consideration of major decisions in my own design process. This process generated five categories of game types, as well as an initial set of characteristics that differentiate them.

In a second phase of ideation, I worked with the framework at a conceptual level to clarify both the categories of games and how they could be differentiated from one another. My goal in this was to enhance the clarity of the archetypes, as well as my own ability to articulate them to my validation audiences. My primary approach to this work was to consider the conceptual challenges of designing a game that shared attributes of multiple archetypes. I considered what it would mean to combine those game objectives in a single design with particular attention to areas of design I felt would produce tension. These areas of tension could then be highlighted as differentiating characteristics of each type. By repeating this process for each possible combination of two types (that is, considering all 10 potential combinations) I was able to eliminate one type and refine my characterization of the remaining framework elements.

At the end of this stage, I had generated an initial framework, with short descriptions of each type, as well as a table characterizing the types along 10 different dimensions. This short document served as the initial basis for discussions in Phases 3 and 4.

Phase 3: Validation

In the third stage of the project, I elicited direct feedback on the framework from experienced game designers and sponsors to ensure both validity and utility. “Validity” would mean that the patterns I capture in the framework map onto patterns experts recognize from their own practice. By “useful” I mean that the presentation is clear and applicable to the types of game problems confronted by designers, frequent participants, and sponsors. While different designers found different tools helpful and reflective of their own practice, my aim was for the majority of designers to recognize the utility of the framework in describing their own practice and output.

The primary approach I used to conduct validation was semi-structured interviews¹¹⁴ with game designers and sponsors in which I walked participants through the framework. I then actively elicited whether the overall framework approach, and the specific archetypic categories, captured the subject's understanding of games. I also asked if they thought the framework would be useful, both to themselves, and to other stakeholders. Generally, feedback took one of three forms: 1) general statements that the framework aligned with their understanding of design, 2) specific concerns with one or more categories, and how they interacted with their own experience, or 3) concerns about how to differentiate two or more categories.¹¹⁵ Based on feedback of the later two types, I then made revisions to the framework, prior to conducting the next interview, restarting the cycle.

I conducted interviews with a range of game designers and sponsors. Over the course of the study I interviewed over 30 individuals in one-on-one and small group settings. Individuals were recruited from my professional network, recommendations from other subjects, and calls for interested participants made at major gaming conferences. Table 2.1 summarized the institution and equities represented in these conversations.

In addition to individual interviews, I also conducted a broader "validation workshop" at the Connection Wargaming Conference in July, 2018. This session was held during a workshop track of the conference during which participation was determined by self-selection. Nearly 40 individuals opted to participate in the session. While detailed demographic information was not collected, the group included participants from the US Army, Air Force, and Navy, intelligence community, contractors working across the joint and service communities, and UK MoD. The workshop included both a presentation of the framework featuring a somewhat abridged version of the type of feedback offered in the individual interviews.

Participants also engaged in a "typing" activity, in which they were given a description of a game purpose and were asked to assess how appropriate each of the archetypes would be to apply. This approach was modified from techniques for consensus elicitation such as the RAND/UCLA appropriateness method in which participants engage in multiple rounds of scoring about the appropriateness of an approach, given a stated problem.¹¹⁶ This approach has the advantage of providing more structured feedback that focuses on presenting a unified view of the expert community. The goal of this mixed format was to balance rich, free-flowing discussion with more structured data collection.

¹¹⁴ "semi-structured" interview protocols consist of a set of pre-determined questions, but in contrast to a structured approach the research is free to insert follow up questions, drop questions, or change the order of the discussion based on the natural flow of the conversation.

¹¹⁵ For the initial decision to focus on distinguishing characteristics, I am indebted to discussions with Stacie Pettyjohn (interview, Arlington, VA, May 2018)

¹¹⁶ Kathryn Fitch et al., *The Rand/Ucla Appropriateness Method User's Manual* (Santa Monica, CA: RAND, 2001).

Table 2.1: Affiliation of interview and selected workshop subjects

Organization	Office	Designers	Sponsor
Joint	Deputy Secretary of Defense		X*
	Office of the Under-Secretary of Defense, Policy	X*†	X*†
	Office of Cost Assessment and Program Evaluation, Department of Defense	X	X*†
	Office of Net Assessment, Department of Defense	X*†	
	Joint Staff Studies, Analysis and Gaming Division	X	
	RAND	X†	
Army	Army Command and Staff College	X	
	Center for Army Analysis	X	
Navy	Naval War College (NWC) Wargaming Department	X†	
	NWC Strategic and Operational Research Department	X	
	NWC Halsey Group	X	
	Naval Postgraduate School	X	
	CNA	X*†	
Air Force	Air Force Research Laboratory	X	
	RAND	X†	
Marines	Group W	X†	
Interagency	Central Intelligence Agency	X*	X
	Defense Intelligence Agency	X	
	Department of Homeland Security	X†	
International	UK Ministry of Defense	X†	

* indicates interview was with a former (<5 years out) member of the office

† indicates interview included a contractor or federally funded research and development center researchers

Phase 4: Examples

The fourth phase of research focused on gathering a repository of sample games, which could be used first to refine the framework, and then to help illustrate it. In the first phase of this work I collected game reports from archival reports and interviews with game designers and categorized them based on the framework archetypes. When the designer was available and amenable, I included them in this typing process.

Originally, the goal of typing these sample games was to determine 1) whether the framework provides the “types” needed to classify common game designs, 2) what alternative categories are needed to account for successful games, and 3) to provide empirical support for the descriptive and prescriptive description of each type of game. However, as I discovered the fragmented nature of these records and as the framework solidified, the goal of collecting games became more a matter of illustration. By ensuring a diverse sample of games from across the community and across framework types, I hope to provide readers with a range of example games that may reflect their own practice and give them access points for applying the framework.

Expert Elicitation Interviews

The primary source of sample games was drawn from semi-structured interviews with a range of game designers and sponsors. Often, but not always, these discussions were combined with the expert validation interviews described above. In order to elicit specific information about game design, I used an approach modified from critical decision methods.¹¹⁷ This approach to eliciting information about decisionmaking focuses interviews on recounting specific, non-routine incidents. By focusing on specific games, experts are more likely to provide detailed descriptions of the tradeoffs and decisions they made, alternatives they considered, and potential pitfalls they navigated.

Interviews generally proceeded in two parts. The first gathered a general description of the types of games the individual designed or had otherwise been involved with to provide context and situate the interview subject within my sample. The majority of the interview time was then spent discussing a single game that the subject is particularly proud of. While other prompts were occasionally used (such as “a game you would go back and rewrite,” “most typical,” and “a game you were proud of”) this prompt was the most consistent in providing fast recall of a specific game and produced the majority of the game descriptions recounted in chapters 5-8. These interviews were often recorded (and when necessary transcribed) to create a set of rich descriptions of game design choices.¹¹⁸

Archival Evidence

Use of written records as a basis for analysis comes with several advantages. It allows for the study of games over a far longer historical period than living practitioner memory, ensuring greater diversity of evidence to incorporate into the framework. It also provides a more transparent body of information, since other researchers will be able to access the same records to independently verify or dispute findings. However, as noted above, the available game record is only a small snapshot of the range of games produced, and there is no reason to believe there are not systematic biases that shape what is available. As a result, an archival-focused approach is not viable.

That said, written records of specific games offered a useful complement to expert interviews. In some cases, interview subjects specifically referenced reports they felt had useful details about game designs, particularly when looking at older projects. As a result, I’ve used archival materials to supplement interview-based descriptions for several games. In a small number of cases for which I could find well-documented, unclassified game reports that presented a perspective that was missing in interviews (particular on historical gaming), I also cite example games that I draw entirely from written records. However, I make no pretense of

¹¹⁷ Klein, *Sources of Power: How People Make Decisions*. p 189

¹¹⁸ In some cases, the setting of the interview could not accommodate recording, and contemporaneous notes by the interview were used instead.

these being anything but a convenience sample based on the records that were available to me. Thus, the majority of these games are drawn from the RAND Corporation Archive.

Professional Practice

In addition to drawing on games designed by others, I also include examples drawn from my own practice which are publicly available. In part, this is simple expedience, since there is no body of games in which I can understand the tradeoffs made as well as the ones where I made those choices. As a result, I can provide rich descriptions of these games to help illustrate some of the tradeoffs which might feel abstract otherwise. I also include discussion of these games because their design was a key testing ground for working out the framework presented in this book. Often it was the synthesis of more theoretical efforts, conversations with other designers, and then observing how I made choices in my own design work that most clarified my thinking.

Framework Refinement

Taken together, the archival and interview records provided data for two types of analysis to support the refinement of the framework. The first approach was to check the descriptive power of the draft framework. Games were generally identifiable as one or more types, though the process of talking through the typing for a specific game with other designers early in the process was helpful in clarifying the archetype descriptions. Descriptions of games also helped to refine the tradeoffs detailed in each of the framework's types. Examples of actual design decisions suggested different tradeoff considerations or alternative strategies which were incorporated into the framework.

However, these data and the resulting analysis have some key limitations. The first is that the results are not representative—written descriptions of design and strongly recalled games focus on games for which design was more complex, novel, or otherwise remarkable. Further, the uncertain representativeness of both the sample of game practitioners and of game documentation also introduces risks of missing elements that should be included in the framework. Phase 3 was designed to mitigate some of these risks by enabling direct feedback on the framework from interview. However, I believe it will be sufficient that the framework be broadly useful, even if it is not able to demonstrate that types of the framework are mutually exclusive and collectively exhaustive (a frequent standard for typologies that need not apply to archetypes).

Chapter 3: Towards a Social Science of Policy Games

What and how do we learn from games designed to support research and analysis? Surprisingly, this is not a settled question among policy gaming practitioners. A debate has raged over whether games can best be considered an art or a science, and what that might mean for what types of conclusions can be drawn from games.¹¹⁹ Advocates for seeing games as an “art” emphasize the experience of the participants of a game, and how games have the ability to cultivate new thoughts and build new understanding in the minds of players.¹²⁰ Defenders of the position that games are a science argue that, if games are going to contribute to analytic projects, they should be held to the same standards as other types of research.¹²¹ While both positions are represented among professional gamers, based on analysis of a recent survey of gamers, those using artistic language and design principles like “player enjoyment” outnumber the supporters of viewing games from a scientific perspective.¹²²

In this chapter, I explore the traditional argument between those who see games as an art or a science and why the distinction matters for the use of policy games. While recognizing that truly masterful game design requires creativity and art, I argue that in order to inform policy analysis, games must be grounded in science. However, that “science” is less monothetic than the standards frequently invoked in current debates. Drawing on social science, which is also interested in the role of humans and their interactions, and thus is a better base of comparison than the natural sciences, I identify multiple distinct “sciences” that analytic gaming can support. Each is defined by a “philosophy of science”—that is, a set of key claims about how we can learn about the world, which in turn dictate a logic by which we conduct research. These logics are critical, because they provide a template for assessing the claims made in research and analysis. To the extent that the way a game generates information is consistent with a logic of a philosophy of science, that game is able to provide insights within the bounds of that philosophy. When games deviate from the selected logic, the resulting claims become less credible.

To make these logics more accessible to game designers, sponsors, and consumers, this chapter leverages an existing typology of philosophies of science from international relations, which considers questions of political and military decisionmaking that are often the focus of national security policy analysis games and demonstrates how existing theory about how games

¹¹⁹ Peter P. Perla, “The Art and Science of Wargaming to Innovate and Educated in an Era of Strategic Competition” (paper presented at the King's College London Wargaming Network Lecture, London, UK, 2018 of Conference).

¹²⁰ Perla and McGrady, “Why Wargaming Works.”

¹²¹ Barzashka, “Wargaming: How to Turn Vogue into Science.”

¹²² Bartels, “Short Insights from a Survey of the Wargaming Community.”

tell us about the world fall into several of the articulated logics. In other words, designers are already using multiple “scientific” approaches to gaming; they have just not positioned themselves in a way that recognizes the existence of multiple, valid logics. In contrast, a pluralistic approach recognizes that not all games need to apply the same logic, arguing for tailored assessment of games stemming from each of the philosophical traditions. Put concretely, games can be judged based on their ability to generate knowledge through a scientific process, but that process, and thus the standard of judgement, will be different for different games. Conversely, applying the standards of one philosophy of science to a game designed to support inquiry in a different logic will not be sound.

Policy Games as Art: A Dominant Paradigm with Problems

A sizable portion of game designers today identify their work as something more akin to art or craft than a science. Advocates of this position hold that while the facts and models that underpin the game are based on research that may be scientific in nature, the work of building, running, and making sense of a game are an art.¹²³ Proponents of this perspective believe that there is an “undercurrent” of storytelling—or more forcefully put, a “*story-living* experience”—that taps into both the human ability to be moved by stories and our ability to learn through experience.¹²⁴ In this perspective, “designing and playing a wargame inherently involves storytelling. There can be many different types of stories being told, but in general they circle around what is possible, what is unexpected, what works and led to victory, or what failed and led to defeat.”¹²⁵ Following this logic, games improve decisionmaking by shaping the stories senior leaders tell about policy problems and their potential solution. In particular, games provide an opportunity to question the stories that form conventional wisdom about national security policy problems and elevate new stories into senior leaders’ discussions. These stories then shape subsequent analysis and experimentation.

Often, wargamers who subscribe to this approach center their practice on the key role of human players in games. In the words of ED McGrady: “There is a difference between social models of behavior and a person across the table trying to beat your strategy. Wargames draw this out... This presence forces players to deal with their own perceptions, feelings, ambitions, and concepts, while also having to deal with the opponent.”¹²⁶ This focus on the experience of players, and the narratives that result from that experience, orients the designer to think about the type of experience they want participants to have, and the elements of game design that are likely

¹²³ McGrady, “Getting the Story Right About Wargaming.”

¹²⁴ Perla and McGrady, “Why Wargaming Works.” p 112

¹²⁵ McGrady, “Getting the Story Right About Wargaming.”

¹²⁶ Ibid.

to promote such an experience. This move argues that the work of the game designer is inherently artistic, rather than analytical. To quote McGrady again:

In building the design the designer is creating a terrain that the players will explore. Since there is no hard and fast rule about how, or why, the players will explore various parts of the terrain, the designer is producing an incomplete design. The design is only completed by the interaction of the players with the design.

I say the designer is creating a narrative path for the players in the game. The players then come and add to or change that path in some ways. So it's a collaborative art form in both design and execution. Given the collaborative nature of the enterprise, it strikes me as odd that you would talk about the "science" of games, since any science input to the design will quickly be overcome by the decisions of the players.¹²⁷

In other words, in this view the process of game design is artistic—made of choices about how to create a particular experience for players rather than that of a scientist, trying to make observations to learn about how the world works. As an artistic endeavor, the "goodness" of a game is defined by its ability to create a space for participants to engage with and generate compelling narratives. In this view "good games are ritual spaces inside which play becomes real to the participants,"¹²⁸ encouraging measures of success like "player engagement" that are common among practicing gamers.¹²⁹

Proponents of this perspective highlight the limits of games to argue that it is dangerous to place an inappropriate analytic burden on games. They emphasize the artificialities of the game environment, the inability to truly repeat or control the interactions of groups of players and adjudicators, and warn about the limits of meaningful data captured, given the complexity of game play. While sometimes these caveats are applied only to treating games as experiments¹³⁰ other researchers raise these concerns more generally. Many of these concerns are well founded. In particular, the emphasis on humans tends to focus games on questions that are difficult to quantitatively measure and resist simple statistical approaches to analysis commonly used to study the physical behavior of weapons systems. However, as is discussed later, these features are not limited to games, but rather are shared by other fields studying the behavior of humans and institutions in the social sciences and have been overcome in a range of ways.

Finally, it's worth highlighting that some members of this community have recently taken inspiration from the emerging field of artistic research.¹³¹ This emerging approach picks up on an

¹²⁷ ED McGrady, personal correspondence, February 2020.

¹²⁸ James Fielder, "Reflections on Teaching Wargame Design," *War on the Rocks*, January 1 2020.

¹²⁹ Bartels, "Short Insights from a Survey of the Wargaming Community."

¹³⁰ McGrady, "Getting the Story Right About Wargaming."

¹³¹ Originally proposed in Perla, "Short The Art and Science of Wargaming to Innovate and Educated in an Era of Strategic Competition." this line of research has been expanded to excellent effect in: Judge, "The Wargaming Guild: How the Nature of a Discipline Impacts Its Craft and Whether It Matters."

earlier articulation of two fundamental ways of viewing the world. The first is the personal narrative, in which subjective interpretation is dominant. The second seeks to “classify, schematize, and analyze” in a “paradigmatic approach.”¹³² While research is traditionally linked with the latter, artistic research seeks to harness the former approach to produce new knowledge which offers the potential for a research approach that can integrate many different aspects of experience. Artistic research seeks to do so from the perspective of the artistic practitioners as a researcher rather than a subject of research.¹³³ Recent work by Sawyer Judge has examined the approach as a potential model for gaming in an artistic mode that recognized features such as the role of community, position of the designer, and location of game validity in the skill of execution of the medium.¹³⁴

Pitfalls of Artistic Approaches to Policy Analysis Gaming

The commitment to treating policy gaming as an art comes with costs. United States Department of Defense officials have stated that greater systemization of gaming is necessary for games to have their desired impact on policy.¹³⁵ Conversely, others also express frustration at the large number of games that fail to contribute meaningfully to defense analysis and decisionmaking.¹³⁶ Spend time at a gathering of gamers and it is normal to hear concerns that the results of games are warped or discarded by sponsors and stakeholders,¹³⁷ and that inexperienced gamers are developing unsound products.¹³⁸ In other words, there is a concern by both game designers and sponsors that many games are not good, but there are few well-understood ways to evaluate the quality of games, beyond the direct judgement of highly experienced practitioners.¹³⁹

In large part, this absence of communicable standards can be rooted in the identification of games as art. The use of the term “art” conjures up a degree of subjective judgement—standards for “good” or “useful” can be fluid. While advocates of treating games as art argue that the validity of games comes from the work games do when they “reshape our conception or shed light on things we had not seen before”¹⁴⁰ in practice this has not produced satisfactory means of

¹³² Jerome Bruner, *Actual Minds, Possible Worlds* (Cambridge, MA: Harvard University Press, 1986).

¹³³ Kathleen Coessens, Darla Crispin, and Anne Douglas, *The Artistic Turn: A Manifesto* (Leuven, Belgium: Leuven University Press, 2009). pp 42-43

¹³⁴ Judge, "The Wargaming Guild: How the Nature of a Discipline Impacts Its Craft and Whether It Matters." Pp 24-27.

¹³⁵ Work and Selva, "Revitalizing Wargaming Is Necessary to Be Prepared for Future Wars."

¹³⁶ Jon Compton, "The Obstacles on the Road to Better Analytical Wargaming," *ibid.* 2019.

¹³⁷ Downes-Martin, "Your Boss, Players and Sponsor: The Three Witches of War Gaming."

¹³⁸ Pettyjohn and Shlapak, "Gaming the System: Obstacles to Reinvigorating Defense Wargaming."

¹³⁹ Bartels, "Short Insights from a Survey of the Wargaming Community."

¹⁴⁰ Perla, "Short The Art and Science of Wargaming to Innovate and Educated in an Era of Strategic Competition."

delineating good and bad games. For example, when experienced designers are asked to articulate how they judge a “good” game, many heuristics focus on player engagement and emotion.¹⁴¹ Standards that focus on engagement of players risk producing games that are enjoyable and thus engaging to play but are disconnected from the real-world shape of policy problems they seek to inform. Another approach to assessment depends on experienced game designers calling out bad games,¹⁴² and the somewhat converse practice of gauging the quality of a game based on the reputation of the designer.¹⁴³ Both approaches are quite consistent with other artistic practices that posit “the simple idea that if it looks like art, it is. All that is required is a level of competency in the medium deemed appropriate by the community.”¹⁴⁴

Such an approach to evaluation has been criticized over the history of gaming. First, standards for assessment that are focused on the immersive experience of participants are hard to achieve by anyone other than participants, but as discussed in Chapter 1, there is also a concern that participants may be misled by games that are compelling, but fundamentally flawed. This creates a tension where the individuals viewed as best able to value a game are also seen as unreliable. Moreover, in practical policy contexts individuals beyond the players seek to consume the results of games—and this approach offers few tools for other stakeholders to gauge whether game results are a sound basis for decisionmaking.

Similarly, dependence on the community to determine “competency” is beset by practical challenges. Practically, there are not enough good gamers out there to directly assess all games run by the department,¹⁴⁵ and even if there were issues like classification and competition between rival contracting firms limit the opportunities for direct observation of the work product of our peers. Furthermore, the question of how a sponsor might identify a good game designer to perform assessment creates a chicken and egg problem. Without knowing if the games a designer creates are good, how can we determine that the designer is expert enough to judge the quality of games?

Finally, judging the worth of games on artistic grounds opens games up to criticize on aesthetic grounds. On one hand, no game designer will deny the importance of “chrome” and “fluff”—terms borrowed from the commercial hobby gaming word to describe elements not necessary to the game’s core mechanics, but that add to the narrative engagement and excitement.¹⁴⁶ On the other hand, most designer have also had the experience of wrestling with a game sponsor who wants a game to look a specific way because it aligns with their aesthetic

¹⁴¹ Bartels, "Short Insights from a Survey of the Wargaming Community."

¹⁴² McGrady, "Getting the Story Right About Wargaming."

¹⁴³ Bartels, "Short Insights from a Survey of the Wargaming Community."

¹⁴⁴ Judge, "The Wargaming Guild: How the Nature of a Discipline Impacts Its Craft and Whether It Matters." summarizing Gardner, *The Arts and Human Development; a Psychological Study of the Artistic Process*. p 117

¹⁴⁵ Pettyjohn and Shlapak, "Gaming the System: Obstacles to Reinvigorating Defense Wargaming."

¹⁴⁶ Rex Brynen to Paxsims, June 26, 2019, <https://paxsims.wordpress.com/2019/06/26/setting-the-wargame-stage/>.

preferences even when the format cuts against the nature of the problem. For example, a game exploring issues at the nexus of political, economic, and military decision making is not likely to be served well by a hex and counter approach that only represents conventional military forces, even if that is a type of game the sponsor finds appealing. At best, such convictions add considerable stress to the design process. At worse they lead to poorly designed games or games whose potentially useful findings are ignored because of the aesthetic preferences of the consumer.

The Artistic Case Against Science

Given these pitfalls, the benefits of adopting a scientific process, which provides a logical approach and means to fairly judge the utility of analysis, seems intuitively appealing. So why is there so much resistance to viewing games as a scientific approach? Ivanka Barzashka's recent article in the *Bulletin of the Atomic Scientists* calling for a more scientific approach to gaming offers some explanations. Most critically, she argues, many gamers have a narrow understanding of how science is practiced.¹⁴⁷ Many gamers, as well as the broader national security analysis community, are strongly influenced by military operations research (OR) and associated engineering and business efficiency approaches to research. While some practitioners define the boundaries of OR around the goal of using the scientific process to improve organizational decisionmaking,¹⁴⁸ a large portion of the field centers itself on quantitative approaches to analysis, particularly optimization.¹⁴⁹ Often these approaches are associated with an interest in standardization as a way to gain quantitatively measurable efficacies, often without sufficient consideration of whether rigid processes are appropriate.¹⁵⁰

Unsurprisingly, this rigidly quantitative approach to science invested in standardization is a poor fit with the human-centric, exploratory nature of games. Modern American OR tends to see wargaming as an inferior tool—possible appropriate in early stages of research, but always severely limited.¹⁵¹ In John Hanley's memorable phrasing "Gaming is both a grandparent and orphan of operations research."¹⁵² It is therefore unsurprising that gamers resent the attempt to demarcate their field as "unscientific" and thus less valuable than other approaches to policy

¹⁴⁷ Barzashka, "Wargaming: How to Turn Vogue into Science."

¹⁴⁸ Wong, "Preparing for Contemporary Analytic Challenges."

¹⁴⁹ John T. Hanley, "On Wargaming" (University of Maryland, 1991). p iv

¹⁵⁰ Joseph M. Hall and M. Eric Johnson, "When Should a Process Be Art, Not Science," *Harvard Business Review*, 2009.

¹⁵¹ In part this is due to debates in the 1950s and 1960s as civilian researchers with "scientific" expertise fought to replace the dependence on military judgement in defense planning. For discussion of this history and its impact on gaming, see: Perla, "Short The Art and Science of Wargaming to Innovate and Educated in an Era of Strategic Competition." and Sharon Ghamari-Tabrizi, "Simulating the Unthinkable: Gaming Future War in the 1950s and 1960s," *Social Studies of Science* 30, no. 2 (2000).

¹⁵² Hanley, "On Wargaming." p v

research. As a result, many gamers fear that movement towards “scientific” games denudes games of key aspects of the tool—applying standards and practices that are not consistent with the core focus on human decisionmaking. The parallels to artistic communities that “fear that a thinking culture which insists of rigorous analysis might interfere with the skill and open-endedness, the pre-noetic, deeply intuitive and intensely felt quality of experience that constitutes an artistic performance”¹⁵³ are obvious and unsurprising.

Some members of the gaming community have argued that there are alternative forms of scientific practice that offer different demarcations about what is scientific and may better accommodate the nature of gaming. For example, recent work by Yuna Wong has situated gaming in a broader history of “soft” operations research, which accepted a much broader range of approaches to research.¹⁵⁴ Other recent work by scholars and practitioners such as Compton,¹⁵⁵ McCown,¹⁵⁶ Bartels,¹⁵⁷ and Barzashka¹⁵⁸ has pointed to the social sciences as a potential model, given those disciplines’ need to grapple with the same types of human decisionmaking that are the focus of games.

This work follows their lead, suggests the potential value of a broader exploration of science and how gaming might fit its paradigm. To paraphrase a scholar of intelligence analysis engaged in a similar debate: “rather than ask whether intelligence analysis [or in our case, gaming] is an art or a science, more productive answers will come from asking what kind of science [gaming] is or could be?”¹⁵⁹ To answer this question, I turn to a brief exploration of philosophy of science in the context of gaming.

Towards a Science for Gaming

Drawing on social science, which is also interested in the role of humans and their interactions, and thus is a better base of comparison than the natural sciences, I identify multiple distinct “sciences” that analytic gaming can support. Each is defined by a “philosophy of science”—that is, a set of key claims about how we can learn about the world, which in turn

¹⁵³ Coessens, Crispin, and Douglas, *The Artistic Turn: A Manifesto*.

¹⁵⁴ Wong, “Preparing for Contemporary Analytic Challenges.”

¹⁵⁵ Compton. “Analytical Gaming.”

¹⁵⁶ Bartels, McCown, and Wilkie, “Designing Peace and Conflict Exercises: Level of Analysis, Scenario, and Role Specification.”

¹⁵⁷ Elizabeth M. Bartels, “Games as Structured Comparisons: A Discussion of Methods” (paper presented at the International Studies Association, San Francisco, CA, 2018 of Conference); Goertz, *Multimethod Research, Causal Mechanisms, and Case Studies: An Integrated Approach*.

¹⁵⁸ 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).

¹⁵⁹ Aaron Frank, “The Philosophy of Science and Intelligence: Rethinking Science in Support of Intelligence” (paper presented at the International Studies Association Annual Conference, San Diego, CA, 2012 of Conference).

dictate a logic by which we conduct research. These logics are critical, because they provide a template for assessing the claims made in research and analysis that can be applied relatively naively—once one understands the logics, it is possible to apply that standard to a game design without being an expert in the approach (though of course an expert's perspective will be more nuanced). To the extent that the way a game generates information is consistent with a logic of a philosophy of science, that game is able to provide insights within the bounds of that philosophy. When games deviate from the selected logic, the resulting claims become less credible within the bounds of the selected philosophy.

A cursory survey of the philosophy of science literature reveals greater diversity of thought than the standard “scientific method.” Limiting our consideration only to major debates within the 20th century, the “standard” description of science as a systematic process to falsify broad law-like claims is certainly present.¹⁶⁰ However, alternative voices argue that this approach is rarely used in practice, instead arguing that “normal” science as practiced in most day to day research instead focuses on solving puzzles that operate within the constraints of major theories.¹⁶¹ Rather than falsifying theories, evidence that counters existing principles is interpreted carefully within the confines of major theories, making truly disruptive work “extraordinary.” Other major figures argue for moving away from the work of falsifying individual theories to instead consider “research programmes” of interrelated claims, which can be examined retroactively to see if they are progressing or degenerating over time and across debates.¹⁶² Taken together, these perspectives argue that there is more than one way available to do science. Based on this diversity of available approaches, gamers may need to search more broadly to find an appropriate approach to model.

In considering a philosophy of science of gaming, it is helpful to refer to the treatment of the topic in international relations (IR) literature. The similarities in topics of study between international relations scholarship and national security policy games makes this a natural bridge point into more academic literatures. Furthermore, the role of IR as a science has long been contested, providing opportunities to see the costs and benefits of different definitions of science, as opposed to adopting the position that the discipline is not scientific.¹⁶³ The eclectic intellectual traditions and substantive concerns in this field also make for a rich discussion of different philosophy of science considerations and concerns.¹⁶⁴ Finally, the focus on rare events

¹⁶⁰ Karl Popper, *The Logic of Scientific Discovery* (New York, NY: Routledge, 1992). p 92

¹⁶¹ Thomas Kuhn, *The Structure of Scientific Revolution* (Chicago, IL: University of Chicago Press, 1970).

¹⁶² Imre Lakatos, "History of Science and Its Rational Reconstructions," in *The Methodology of Scientific Research Programmes*, ed. John Worrall and Gregory Currie (Cambridge, UK: Cambridge University Press, 1978).

¹⁶³ Colin Wight, "Philosophy of Social Science in International Relations," in *Handbook of International Relations*, ed. Walter Carlsnaes, Thomas Risse, and Beth A. Simmons (London, UK: SAGE Publications Ltd, 2013). p 25

¹⁶⁴ I also owe thanks to conversations with several scholars for early formulations of this discussion, most notably Tony Rivera, San Francisco, March 2018 and Jacqueline Schneider, Toronto, March 2019.

like wars and treaties, decisionmaking undertaken under conditions of secrecy, and role of intangible factors in decisionmaking under stress all create stresses for approaches to science that emphasize reproduction of results in controlled environments. As a result, IR puts a premium on creative research designs drawing on a range of approaches to collect and analyzed data, thus recognizing more variability in approach to research than the rigidly standardized approaches described earlier in this chapter. However, the value placed on clever design *does not* remove the requirement that research designs align with a logic of inquiry—the logic is seen as a necessary condition for solid research, which can be improved by the addition of art.¹⁶⁵

Jackson's Typology of Philosophies of Science

Patrick Thaddeus Jackson offers a particularly coherent framework summarizing alternative perspectives according to their grounds for making claims in international relations.¹⁶⁶ Jackson distinguishes two major “wagers,” or fundamental beliefs about how we can know things, that when intersected define four major traditions. On one hand he is concerned with the situation of the researcher in relation to the world, distinguishing between two traditions: One that sees the mind as separate, and therefore able to objectively view events (formally “mind-world dualism”), and one that sees the observer as inherently a part of the world, and thus observation as subjective (“mind-world monism”). On the other hand, he distinguishes between traditions that believe in generating knowledge based on observation alone (“phenomenalism”) from those that argue we can learn about the unobservable (“transfactualism”). When intersected these principles characterize four distinct philosophy of science traditions--positivism,¹⁶⁷ critical realism, analyticism, and reflexivity--illustrated in Figure 3.1.¹⁶⁸

Each of the types Jackson describes captures a particular philosophical position that has been used to generate knowledge in international relations. Below I summarize each of these positions --positivism, critical realism, analyticism, and reflexivity.

¹⁶⁵ I am indebted to Sawyer Judge for clarifying my thinking on this point.

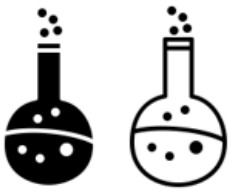



¹⁶⁶ Of course, Jackson's approach is not the only framework available. See: Wight, "Philosophy of Social Science in International Relations." for a summary of alternative, as well as a discussion of their short-comings.

¹⁶⁷ Jackson opts to use the term “neopositivist” to describe this position, however many proponents of the tradition object to the label. I use the more common “positivist” throughout my discussion to increase accessibility.

¹⁶⁸ Jackson, "The Conduct of Inquiry in International Relations: Philosophy of Science and Its Implications for the Study of World Politics." pp 34-39

Figure 3.1: Jackson's typology of philosophies of science

Source: Jackson, p 37

	Phenomenalism	Transfactualism
Mind-world dualism	 <p>Positivism</p>	 <p>Critical Realism</p>
Mind-world Monism	 <p>Analyticism</p>	 <p>Reflexivity</p>

Positivism

The most common of the positions is positivism. Positivists are interested in understanding whether general, law-like statements about causality between discrete factors can correctly describe observed patterns.¹⁶⁹ 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. This perspective links explanation with predictions, since once a causal relationship between factors is established it can be generalized to other relevant contexts.¹⁷⁰ Jackson highlights the space in this tradition for qualitative evidence and modes of analysis that focus on intervening mechanisms, making for a much broader view than the narrow perspective sometimes ascribed to this approach.¹⁷¹ While Jackson's work notes the dominance of this position within academic settings, it's worth noting its even greater dominance in policy circles. The ability to generalize a causal pattern offers decisionmakers the possibility of prediction, to see into the future and correctly project how a policy is likely to play out and enable that knowledge to inform their decision today.

¹⁶⁹ Ibid. p 108

¹⁷⁰ Ibid. p 111

¹⁷¹ Ibid. p 109

Critical Realism

In contrast to the positivist approaches, critical realist accounts deviate from the core claims of phenomenalism to argue that real, but unobservable, phenomena ranging from quarks in physics to social structures in social sciences can be studied scientifically through a process of *abduction*. In order to draw inferences about these unobservables, 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.¹⁷² This approach moves away from the normal scientific assumption of generalizability, to focus instead on building an understanding of the “specific, contingent, and complex.”¹⁷³ This also means that adherents of critical realism argue that theories cannot predict, they can only demonstrate the limits of what is possible, which is valuable if previously unrecognized.¹⁷⁴ While critical realism is a far less popular frame for policy analysis than positivism, it has been attractive to some because of its interest in mechanisms that are a good fit with studies of processes.¹⁷⁵

Analyticism

Analyticism moves away from positivism in a different direction than critical realism by rejecting the separation of mind and world. Instead, the approach argues that theory is an act of sensemaking that tries to explain what is being observed.¹⁷⁶ Researchers in this mode immerse themselves in a problem and then develop an “oversimplification” of the observed complexities which can then be used to produce a case-specific narrative of causality.¹⁷⁷ In other words, researchers in the frame develop models that are simple, and thus inherently non-representative, of the true complexity of the world, but are useful to the researcher 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.¹⁷⁸ If a model is not sufficiently similar to the case to be useful, the model

¹⁷² Ibid. pp 82-83

¹⁷³ Frank, "Short The Philosophy of Science and Intelligence: Rethinking Science in Support of Intelligence." p 38

¹⁷⁴ Jackson, "The Conduct of Inquiry in International Relations: Philosophy of Science and Its Implications for the Study of World Politics." p 111

¹⁷⁵ For some recent examples of work in this vein, see: Phil McEvoy and David Richards, "Critical Realism: A Way Forward for Evaluation Research in Nursing?," *Journal of advanced Nursing* 43, no. 4 (2003). Megan Lourie and Elizanth Rata, "Using a Realist Methodology in Policy Analysis," *Education Philosophy and Theory* 49, no. 1 (2017).

¹⁷⁶ Jackson, "The Conduct of Inquiry in International Relations: Philosophy of Science and Its Implications for the Study of World Politics." p 114

¹⁷⁷ Ibid. p 142

¹⁷⁸ Ibid. p 143-144

may be updated, or the researcher might generate an argument using the specifics of the case as to why the model does not apply.¹⁷⁹

While this approach is less dominant than positivism, the prominence of Max Weber¹⁸⁰ and John Dewey's work in the model has served to keep the tradition alive in policy studies.

Reflexivity

The last of the four types, reflexivity, consists of efforts by researchers to extend their own experience, particularly their experience as researchers and the values revealed by personal approaches to the work, to make claims about the broader context in which they work.¹⁸¹ Generally, this work takes the form of criticism—uncovering current social structures as a means of allowing people to redefine the elements of their context that they find unsatisfying. This approach to analysis, while it frequently policy-relevant, is rarely popular with pragmatic policymakers in part because the tendency of the claims of reflexive work to be read as “partisan interventions or simple statements on behalf of one group or another” because of their grounding in the personal experience of the researcher.¹⁸² As a result, while these approaches to science are common in the academy, they are less frequently used in other empirical research,¹⁸³ and are unlikely to be attractive to analysts with close ties to policymakers.

The Case for Pluralism

Beyond offering this clarifying typology, Jackson's work is focused on defending these traditions as equally viable, though differentiated, ways of conducting science. Work produced in one frame should abide by that frame's logic and should not be judged by the standards of the other frames.¹⁸⁴ This position has since been adopted by advocates of multi-method research, who argue that the same methods may be used across frames but must follow the logic dictated by the frame to be valid. In other words, this tradition argues for pluralism but a pluralism that recognizes and respects distinctions rather than attempting to absorb all tools.¹⁸⁵

¹⁷⁹ Ibid. p 147

¹⁸⁰ Ibid. p 114

¹⁸¹ Ibid. pp 156-159

¹⁸² Ibid. p 168

¹⁸³ Beach and Pedersen, *Causal Case Study Methods: Foundations and Guidelines for Comparing, Matching, and Tracing*. p 11

¹⁸⁴ Jackson, "The Conduct of Inquiry in International Relations: Philosophy of Science and Its Implications for the Study of World Politics."

¹⁸⁵ Beach and Pedersen, *Causal Case Study Methods: Foundations and Guidelines for Comparing, Matching, and Tracing*. P 15

Philosophies of Science for Gaming

Given this pluralism, the question becomes: Can gaming be used to advance science in the first three of these traditions (as noted, excluding reflexivity), and if so, how? As other works on methods using Jackson's typology have established, the same method can be used across multiple philosophical frames, but the application of the method is often different.¹⁸⁶ In examining the sparse theoretical literature on policy gaming, I find evidence of positivist, critical realist, and analyticist approaches to gaming for policy research. While there may be a case that educational games can use reflexivity, it is not often used in games for policy analysis, and thus falls outside the scope of this project, but is briefly discussed below as a point of interest. I also find evidence of exactly the type of cross-frame dismissal of other approaches that Jackson opposes. In the following section, I present the argument for three separate philosophical frames for gaming and argue that each should be treated as a separate, but valid, form of scientific gaming.

First, it is important to be clear that despite frequent claims to the contrary, policy games are designed to produce information about causality that can, to some degree, be transferred to potential events in the real world, including those in the future. These claims stem not from the nature of games themselves, but rather from their application in policy settings. If we claim that games are helpful to decisionmakers in navigating the future,¹⁸⁷ they must in some way arm decisionmakers with correct information about cause and effect that can inform future decisions.¹⁸⁸ That is different than a guarantee of successful prediction of a specific future, since the complexity of many interacting events, many outside the control of the decisionmaker, leads to outcomes that are influenced by more than just their decision.¹⁸⁹ Existing work opts to frame this limitation as games providing *indicative* rather than *predictive* information,¹⁹⁰ an approach with which I generally agree. However it is important to be clear that the work being done is fundamentally about establishing causal relationships. Too often, gamers' work (including me in my own past work¹⁹¹) attempts to be modest about the certainty of our claims and introduce hesitancy about the nature of the claims we are making rather than expressing confidence in their strength.

¹⁸⁶ Ibid. pp 11-13

¹⁸⁷ Work and Selva, "Revitalizing Wargaming Is Necessary to Be Prepared for Future Wars."

¹⁸⁸ Robert C. Rubel, "Epistemology of War Gaming," *Naval War College Review* 59, no. 2 (2006). p 110

¹⁸⁹ Ibid. p 110

¹⁹⁰ Ibid. p 112 and Stacie Pettyjohn and Becca Wasser, "The Promise of Structured Strategic Wargames: Moving Beyond the Seminar" (paper presented at the International Studies Association, San Francisco, 2018 of Conference).

¹⁹¹ Elizabeth M. Bartels, "Games as Structured Comparisons: A Discussion of Methods" (ibid. San Francisco, CA, of Conference).

Beyond the value provided by clearly acknowledging the work gaming is asked to perform, recognizing that games are used as a means of studying causality points us towards a well-developed literature on studying causality in the social sciences. Jackson's work synthesizes a wide range of this literature. I will also return to this point in Chapter 10, where I draw on related literature on the design of studies for causal inference.

Positivism

Perhaps unsurprisingly given the dominance of positivist thought in other areas of empirical social science and policy analysis, there is a substantial community of gamers operating in this mode. In particular, a sizable number of "experimental," "quasi-experimental," and "structured comparison" games attempt to demonstrate the influence of a specific factor on decisionmaking and other outcomes of interest by systematically varying game conditions and observing the effect on player discussions and choices.¹⁹² Generally, these games focus relatively narrowly on demonstrating the connection between a difference in a single key factor and outcomes (for example, linking the presence of a drone vs. piloted aircraft with decisions that were more or less escalatory¹⁹³) or the connection between the type of analysis provided to decisionmakers with the arguments used in decisionmaking (for example, the impact of broad vs. deep analysis on decisionmaking¹⁹⁴). In other words, analysis from these games seeks to provide evidence of a simple causal relationship by tracing patterns of behavior in the game and making claims about other cases where the pattern might hold.

Many within the game design community have disputed the validity of using games in this frame. However, often these concerns have more to do with specific limitations of the approach rather than the appropriateness of the underlying philosophy. Perhaps the most frequent complaint is that the artificiality of game scenarios and role-playing prevents appropriate generalization of game results onto real-world settings.¹⁹⁵ However, this problem is hardly unique to games, since many laboratory experiments also take place in artificial environments.¹⁹⁶

¹⁹² For examples 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); Reddie et al., "Next-Generation Wargames: Technology Enables New Research Designs, and More Data."; Lin-Greenberg, "Game of Drones: What Experimental Wargames Reveal About Drones and Escalation."; Elizabeth M. Bartels et al., "Do Differing Analyses Change the Decision?: Using a Game to Assess Whether Differing Analytic Approaches Improve Decisionmaking," (Santa Monica, CA: RAND Corporation, 2019); Dominic D. P. Johnson et al., "Overconfidence in Wargames: Experimental Evidence on Expectations, Aggression, Gender, and Testosterone," *Proceedings of the Royal Society* 273 (2006).

¹⁹³ Lin-Greenberg, "Game of Drones: What Experimental Wargames Reveal About Drones and Escalation."

¹⁹⁴ Bartels et al., "Do Differing Analyses Change the Decision?: Using a Game to Assess Whether Differing Analytic Approaches Improve Decisionmaking."

¹⁹⁵ Levine, Schelling, and Jones, "Crisis Games 27 Years Later : Plus C'est Deja Vu." pp 2-12 and Parson, "What Can You Learn from a Game?." p 237

¹⁹⁶ For example, consider widespread debates over the generalizability of behavior research based on populations of college students.

In fact, as one game designer eloquently frames the issue, games replicate more of the actual decisionmaking interactions than other laboratory environments and thus produce findings that are more generalizable because they can better mimic interpersonal interactions and environmental complexity. The counter to this is that the researcher loses the level of control typically associated with experiments.¹⁹⁷ Because of inherent variation in people, and in interactions between them, full control across cases is impossible. Again, this concern is not unique to games—techniques such as case study research have devoted considerable attention to accounting for alternative explanations.¹⁹⁸ Finally, the argument is made that the focus on crises and other extraordinary events inherently focuses games on “novelty and uniqueness”;¹⁹⁹ there is a limited call among game sponsors for generalizability. While it is true that the scope of application may be somewhat limited, such concerns have hardly prevented positivist work from occurring using other policy analysis tools. Taken together, the majority of arguments made against positivist approaches to games are concerns about how such work is done rather than the viability of the philosophical approach.

Critical Realism

An alternative approach to gaming focuses instead on games as tools for hypothesis generation through abduction—hallmarks of the critical realist approach to science. The most notable example of this approach can be found in Jon Compton’s work, which stresses that the complexity of war is best understood as a system where “the whole is greater than the sum of its parts.”²⁰⁰ Part of the utility of games stems from being able to observe the system created by competing actors in a specific environment. As a result, rather than trying to separate out individual factors as in a positivist approach, this approach argues that games work best when they consider broader complexes of causal factors and the processes by which these factors cause different outcomes. In other words, this approach is focused on causal mechanisms rather than causal factors. These mechanisms also do not have to be directly observed to be real. For example, a key output of games in this mode is a “theory of success”—that is, a causal argument about what sets of actions are likely to produce the desired result in a specific conflict.²⁰¹ The underlying strategy may not be directly articulated by players, but the individual components and consequences can be observed and the causal force of the strategy analyzed as a result.

In addition to articulating the core understanding of causality espoused by critical realism, this approach to gaming also articulates a number of other claims consistent with this philosophical perspective. For example, Compton also argues that games should not be seen as a

¹⁹⁷ Interview with Jacqueline Schneider, Newport, RI, June 2018.

¹⁹⁸ Bartels, "Short Games as Structured Comparisons: A Discussion of Methods."

¹⁹⁹ Parson, "What Can You Learn from a Game?." pp 238-239

²⁰⁰ Compton. "Analytical Gaming." p 8

²⁰¹ Interviews with Jon Compton, Washington, DC, August 2018 and Phil Pournelle, Washington, DC, March 2019.

deductive or inductive process, but rather as following an abductive logic where a theory is postulated as the best explanation for the available evidence.²⁰² He stresses that this means that games are a tool for hypothesis generation but cannot contribute to proving an abducted theory since a plausible explanation can still prove to be wrong.²⁰³ He also argues against attempts at broad generalization, arguing for narrow generalization to similar cases and stresses that games show what “can” or “may” happen if those specific conditions occur rather than offering any type of law-like generalization.²⁰⁴

Analyticism

Third, another group of texts focuses on games as a type of model following many of the forms of argumentation advocated in the analyticist mode of science. Rather than describing games as an opportunity to observe differences or trace mechanisms that can advance our understanding of causality, this perspective sees games as an opportunity to construct a model of the key causal forces at play. In effect, games yield artificial political-military histories about how events could unfold that are built by “examin[ing] why these events occurred—the combinations of player decisions and umpire determinations that produced them”²⁰⁵ in order to generate a causal narrative. For example, game observations can lead to narratives about how groups make competitive decisions, which can then be considered as an ideal description that might be helpful in explaining real world decisions.²⁰⁶ In other words, the outcome of analysis based on this type of game is the model of the problem developed both by the initial game design and by the contributions of players which flesh out how it evolves over time.

Similar to other work in the analyticist mode, this perspective stresses that valid games are those that produce “useful” knowledge for a specific purpose, rather than making any general claim about games producing “true” information.²⁰⁷ In this model of inquiry, game designers and participants intentionally “distill” a problem by simplifying it enough that it becomes tractable and useful.²⁰⁸ So long as the game attempts to “represent reality to the degree necessary to explore the warfare phenomena in which we are interested,”²⁰⁹ these simplifications do not prevent us from advancing understanding through the use of games. However, as a result of the focus on the game as a simplified mode, this view also stresses that information from games is

²⁰² Compton. "Analytical Gaming." p 6

²⁰³ Ibid. p 6

²⁰⁴ Ibid. p 5

²⁰⁵ Rubel, "Epistemology of War Gaming." p 117

²⁰⁶ A famous example of this type of finding is found in Levine, Schelling, and Jones, "Crisis Games 27 Years Later : Plus C'est Deja Vu." pp 28-30

²⁰⁷ Rubel, "Epistemology of War Gaming." pp 109-110

²⁰⁸ Ibid. p 114

²⁰⁹ Ibid. p 113

conditional²¹⁰--it may be helpful in other contexts, but there should be no assumption that it will describe a generalized causal relationship.

Reflexivity

As noted above, reflexive approaches to research are relatively uncommon in policy analysis, due to their negative framing as advocacy or criticism without concrete policy recommendations for improvement. However, the concept of drawing on personal experience as a way of surfacing broader structures does occur within the literature on policy games but is limited to the role of games as an educational tool, which falls outside the scope of this project. That said, it is worth noting here that the role of games in developing personalized understanding as a means of producing structural critique is well documented in other literatures on gaming.²¹¹

Three Separate but Equal Approaches

The arguments presented above offer three separate arguments about the way games generate useful information for policy analysis. Each follows a clear logic that is consistent with other scientific work, but they produce distinct types of claims, as summarized in Table 3.1.

Table 3.1: Implications of philosophies on games-based research²¹²

Philosophy	Implications
Positivism	<ul style="list-style-type: none"> • Preference for parsimonious theories in which a small number of factors drive differences • Research focused narrowly on the role of specific causal factors • Interest in application of the theory to a broader set of potential cases (generalizability)
Critical Realism	<ul style="list-style-type: none"> • Focused on causal mechanisms that may not be directly observable • Complex causal stories in which multiple causes interact to produce outcomes • Results applicable to a small, carefully bounded population
Analyticism	<ul style="list-style-type: none"> • Focused on sense-making in a specific context • Generates models that are useful in a specific context • May be helpful to explaining other cases, but goal is to produce something useful for the specific case rather than a broad class of conditions

Existing texts tend to present themselves as offering singular, correct ways of producing knowledge from games, setting themselves in opposition to other approaches. Jackson's arguments in favor of a pluralistic approach to philosophy of science argues for a different conclusion—that there is more than one approach to scientific gaming, but that to produce valid

²¹⁰ Ibid. p 114

²¹¹ For examples of approaches to game design explicitly oriented to using personal experience as a means of understanding broader systems, see: Mary Flanagan, *Critical Play: Radical Game Design* (Cambridge, MA: MIT Press, 2009). and Elizabeth Sampat, *Empathy Engines: Design Games That Are Personal, Political, and Profound* (CreatSpace Independent Publishing Platform, 2017).

²¹² Table is modeled on a discussion of case studies in each philosophical frame offered in: Beach and Pedersen, *Causal Case Study Methods: Foundations and Guidelines for Comparing, Matching, and Tracing*. p 12

information, gamers must work within a specific, explicit logic which stipulates how they draw information from a game. If a designer attempts to blend multiple logics, the competing foundational arguments about the relationship between mind and world and the ability of science to make claims about the unobservable with generate a tension that undermines the logic of the work. Thus, there is more than one potential logic but only one logic can be in operation at a time.

The corollary of this claim is that when we assess the ability of a game to tell us something about the world, that standard must be tailored to the specific philosophy of science being used. In other words, the failure of a positivist game to make claims consistent with an analytic logic does not make the game findings invalid; it just makes them poorly suited to an analytical research project. This logic behooves designers to be explicit about what logic they are following. This is not only important to ensure that a game is correctly assessed, it also prevents findings that are valid in one philosophical framework from being imported into another without appropriate consideration and refinement. For example, findings from an analyticist game might be helpful in generating a hypothesis for positivists testing, but they would not stand as causal evidence of a relationship in positivist research. Clear labeling of the philosophy in which research is conducted can minimize these types of errors.

It is also worth noting that within a series of analytical efforts, more than one logic may be at play. In fact, recent work on multi-method research argues that the most productive combination of approaches are those that produce different types of knowledge that can be integrated together to form a broader understanding.²¹³ As a result, it is common to link multiple efforts together that use different logics, specifically because of the different work each is doing. This topic will be returned to in Chapter 10.

Producing Scientific Knowledge with Games

Having established that games can produce information that contributes to scientific knowledge in multiple ways, I now turn to a consideration of what type of knowledge is produced and how it is generated. Three particular characteristics are key: (1) the nature of the problems that are tractable to gaming, (2) the decision about whether to treat the evidence from games as empirical or formal, and (3) the decision about whether to look for evidence of causality through a difference-based approach or through causal mechanism. While these latter distinctions need not align with the philosophical approach to inquiry selected, they often do.

²¹³ Goertz, *Multimethod Research, Causal Mechanisms, and Case Studies: An Integrated Approach*.

The Nature of Problems Suited to Gaming and the Need for a Bayesian Approach to Certainty

One way to consider the type of information games generate is to compare games to other types of analysis. One of the most adopted frameworks for situating wargaming among other defense simulation tools comes from the work of John Hanley, who placed different techniques on a spectrum of indeterminacy associated with the problem.²¹⁴ Hanley contrasts mathematical approaches that are determinist or feature only statistical or stochastic indeterminacy with those that feature more structural indeterminacy. Mathematical approaches require that a problem space be clearly defined, that persistent data be available, that units of measurement be understood, and that relationships be determined in advance of analysis. Such problems have solutions that can be determined mathematically to produce either a point or distribution.²¹⁵ More complex, but still mathematically tractable, problems feature strategic indeterminacy in which competitive dynamics between actors come to the fore (such as can be modeled in game theory).²¹⁶ In contrast, structural indeterminacy deals with problems for which “the bounds of the problem, what elements to include, and unknown relationships and data needed to perform mathematical calculations”²¹⁷ is unknown. It is this latter class of problems to which games are best suited.²¹⁸

The structural uncertainty inherent in the questions posed for games to answer has deep implications for the appropriate level of certainty and confidence to have in game conclusions. It has long been argued that games do not prove anything.²¹⁹ However, all three philosophies of science argue against treating the results of inquiry as settled fact. Positivists generally would argue that we can falsify, but not prove. Critical realists would argue that unobservable phenomena will always be an abducted theory as new observations change our understanding. Analyticism is interested in utility rather than truth for its standard and is thus unconcerned with proof. In other words, according to all three theories, games may not prove, but neither do other forms of scientific discovery.

Instead, games can add to knowledge using any one of the three philosophical approaches described earlier in this chapter, but because our understanding of the problem features fundamental indeterminacies, we should be modest in our certainty about claims-making. When a game generates evidence that appears to support for the existence of a causal relationship, it

²¹⁴ Hanley, "On Wargaming." p 13

²¹⁵ "Some Theory and Practice of Serious "Futures" Games" (paper presented at the Connections Wargaming Conference, Carlisle, PA, 2019 of Conference). pp 6-7

²¹⁶ Ibid. pp 6-7

²¹⁷ Ibid. p 9

²¹⁸ Parson, "What Can You Learn from a Game?." p 234

²¹⁹ For a famous formulation of this position see: Levine, Schelling, and Jones, "Crisis Games 27 Years Later : Plus C'est Deja Vu." p 15

can be easy to focus on uncertainty about the nature of the postulated relationship. However, given that we ought to assume a high level of uncertainty about the causal relationship, it may be more productive to focus on the degree of confidence we have that what we see in the game is actually evidence of a particular causal relationship.²²⁰

One approach to articulating this type of uncertainty is to adopt a “folk Bayesian” approach to qualifying analysis.²²¹ Taking its cue from Bayesian statistics, which focuses attention on how evidence for causality shifts confidence in our beliefs, this approach recommends that we carefully work to understand what pieces of evidence from the game mean for our causal argument. Such an approach requires careful consideration of what information could be generated by the game that would support or refute a particular causal claim before the evidence is collected. Once the evidence is gathered, we then need to assess whether the evidence collected actually supports or refutes our core claims.²²² In other words, we need to think first about what the evidence could tell us; then, after setting those standards, use the standards to guide what we should take away from the game. In contrast to traditional Bayesian approaches, these results are unlikely to be a quantitative measure of certainty, but instead might resemble the types of confidence assessments common in the intelligence community.²²³

Game Evidence: Empirical or Formal?

One of the most discussed characteristics of a game is its artificiality—in the words of Bob Levine, games feature both ersatz history (that is, fictitious scenarios) and people (in the form of role-playing players).²²⁴ When we use a game to inform us about the potential motivations of a leader in a fictitious crisis or course of action employed for a military capability that does not exist, the task of a game is fundamentally estimative or speculative,²²⁵ making claims for which we do not yet (and may never have) observed historical events to serve as a check. As a result, for games to have policy utility, we must explain to an (often skeptical) audience what can be learned in the face of the artificial nature of games.

Two different understandings of what types of observations we are generating in a game are in evidence. The first treats games as a data-generating event, in which player decisions can be observed as events occurring in the world, and the resulting observations analyzed as empirical

²²⁰ Put more formally, our uncertainty is epistemological, not ontological in nature

²²¹ Adapted from Beach and Pedersen, *Causal Case Study Methods: Foundations and Guidelines for Comparing, Matching, and Tracing*. Chapter 6

²²² Ibid. p 156

²²³ Richard J Heuer and Randolph H. Pherson, *Structured Analytic Techniques for Intelligence Analysis* (Washington, DC: Congressional Quarterly, 2011).

²²⁴ Levine, Schelling, and Jones, "Crisis Games 27 Years Later : Plus C'est Deja Vu." pp 3-12

²²⁵ To use the terms employed by noted intelligence thinker Sherman Kent in: Sherman Kent, "Estimates and Influence," in *Sherman Kent and the Board of National Estimates: Collected Essays*, ed. Donald P Steury (Washington, DC: Central Intelligence Agency, 1994).

data. While findings must account for potential artificialities introduced by the synthetic environment and role-playing, this approach focuses on the fact that in the game real people are making decisions and experiencing consequences.²²⁶ The game design is treated much as the experimental setup in a laboratory experiment might be—as infrastructure needed to generate the phenomenon and record data about it. The game design is the subject of study only to the extent that documentation is required for another researcher to reproduce the event or understand the data-generating process in order to interpret the game’s findings. Observations of game play are analyzed as empirical evidence of decisionmaking, just as historical records of decisions might be treated.

An alternative perspective views games as models that incorporate humans as part of the simulation and which generate observations of the logical implications of the game’s structure. In this view, game observations are not empirical—they are the logical extensions of the representations built collaboratively by designer and players that can be studied for insights in the same way as a technical drawing or computer simulation. In this frame, the game design is simply the designer’s contribution to the model, which is completed once the players introduce their understanding and see how the system of game and players evolves together over time. Game analysis attempts to describe what is happening in the model, which can then be related back to the real world. Inferences can be drawn from the process about the logical implications of the model, but such findings are based on formal rather than empirical grounds.

While these distinctions do not map absolutely onto the divisions between the philosophical positions differentiating the philosophies of science, as a general rule advocates of mind-world monism will be unlikely to view games as empirical. If independent observation is not possible, then it makes far more sense to treat both game designer and players as part of the theory-generating unit, rather than treating the observers of the game as a separate entity capable of existing outside the system. In contrast, while positivist and critical realists may opt to treat game data as a model, these approaches prioritize data observed by an outside actor and thus will tend to encourage game analysis to treat game data as empirical observation.

Basis of Game Analysis: Differences or Mechanisms?

Scholars of research methods in the social sciences have long differentiated approaches to analysis that are based on measuring differences to demonstrate causation from those that focus on the process that connects cause and effect. While often advocates of one position argue for the superiority of their chosen approach, the reality is that they are different means of exploring

²²⁶ A more moderate version of this perspective argues that only a subset of types of observations from a game can be treated as empirical data. For example, a body of work by scholars at the Naval War College argued that command and control decisions were a unique are of games in which actually communication and decisions, rather than simulated ones, occurred in game and thus were appropriate to treat as empirical evidence. For more details on this argument, please see: Perla, Markowitz, and Weuve, "Game-Based Experimentation for Research in Command and Control and Shared Situational Awareness."

causality which will persuade different people.²²⁷ Both can be applied to games, and while again there is not necessarily a one-to-one relationship between the selection of an approach and the philosophical frame of the work, there are strong tendencies that tend to link them.

Difference-based approaches make the fundamental argument that we can understand causality by comparing what happens when a cause is present and absent, and comparing the difference in outcome. In practice this can be done either by comparing across cases, or by comparing a case to a logical argument about what would have transpired.²²⁸ In the social sciences, because there are usually multiple sources of variation, pains must be taken to attempt to control other sources of variation to make difference-based claims compelling, and when such control is not available to explain logically why potential confounding explanations are less credible than the primary causal argument.²²⁹ The disadvantage of this approach is that while it may demonstrate clear evidence of cause and effect, it cannot provide much information about how the cause actually produces an outcome,²³⁰ which is often of great interest to policy makers attempting to construct new interventions or anticipate second order effects of strategies. Critics are especially likely to attack counterfactors that are based on logic rather than observation, since often compelling arguments can be put forward for alternative scenarios that also appear plausible.²³¹

In contrast, a focus on causal mechanisms seeks to lay out how causation actually occurs by laying out the process that connects cause to effect. This must be more than a narrative about what events occurred in what order; instead we must map out a system that can explain why “causal force” is transferred through the causal mechanism to produce the observed result.²³² In other words this approach focuses on the activities that link different parts of the system, rather than focusing on factors that may be present or absent—the system as a result is more than the sum of its parts and is liable to be misunderstood if it is atomized.²³³

Based on these descriptions, it is perhaps not surprising that as a general rule, positivists and analyticists tend to gravitate towards difference-based explanations, and critical realists towards mechanistic explanations. However, there are some key nuances that complicate this division. While both positivist and analyticist approaches lean on counterfactuals to explore differences that can flesh out causal relationships, in practice counterfactuals are used differently by each approach to science. In positivist research, counterfactuals are used to explore how a causal

²²⁷ Beach and Pedersen, *Causal Case Study Methods: Foundations and Guidelines for Comparing, Matching, and Tracing*. pp 27-28

²²⁸ Ibid. p 28

²²⁹ Ibid. p 30

²³⁰ Ibid. p 31

²³¹ Ibid. p 40

²³² Ibid. p 35

²³³ Ibid. pp 37-38

relationship works across different cases. In contrast, analyticist approaches use counterfactuals to explore alternatives to what happened in a single case—for example, by highlighting the differences between a model and a case as a means of helping us understand what happened in the specific case.²³⁴ This alternative form of counterfactuals is not at all interested in numerically measuring differences across cases, but rather in imagining how the narrative of events would be altered by different conditions based on the researchers' prior experience. Put differently, in analyticism, causal factors are those we cannot imagine getting the outcome in question without.²³⁵ This approach fundamentally depends on imagining the counterfactual—"if this factor was absent, could this outcome have occurred?" However, at interest is not a measured difference in value between the two outcomes, but rather a causal narrative—how did events unfold differently in different conditions. At the same time, some positivists argue that mechanisms may be explored practically by breaking up the mechanistic process into a series of smaller causal relationships, which can then be investigated using a difference-based method.²³⁶ Put simply, it's possible to construct arguments that draw on both difference-based and mechanistic style arguments, and so it is worth being explicit which approach, or combination of approaches is used in analysis to ensure that the logic of argumentation is clear, and claims can be properly evaluated.

Conclusion

Fundamentally, this chapter argues for the value of a scientific approach to gaming that can generate logical standards for designers to use as a guide to building policy games. To do this, I draw on philosophical traditions from social science approaches to studying international relations to argue that science, far from being the monolith that is so often presented by the defense analysis community, actually has multiple viable logics that can be uses. I identify three of these logics, positivism, critical realism, and analyticism, within the current literature on game design. Recognizing these logics as distinct shifts the discussion from arguments about how game design ought to work to establishing distinct, internally consistent standards for work in each tradition. In other words, a positivist should not reject critical realist work simply because it does not adhere to the standards of positivism, but rather evaluate the work using the logics laid out by critical realism.

Conversely, work in a given tradition should only make claims that are consistent within that tradition. For example, a game that follows an analyticist approach cannot produce a validated theory of causality that is generalizable to other contexts, but rather can generate a useful model

²³⁴ Jackson, "The Conduct of Inquiry in International Relations: Philosophy of Science and Its Implications for the Study of World Politics." p 115

²³⁵ Ibid. pp 148-149

²³⁶ Beach and Pedersen, *Causal Case Study Methods: Foundations and Guidelines for Comparing, Matching, and Tracing*. p 36

of a specific context of interest. These differences suggest that games produced under different philosophies might best serve different analytical purposes, because they produce different types of information. Considering the purposes of games, the relative ability of different philosophical frameworks to support them, and some of the practical difference in study and game design that might result is the focus of the next chapter.

Chapter 4: Four Archetypes of Games to Support National Security Policy Analysis

The previous chapter discussed how we can learn from games by leveraging different philosophies of social science. I argue that current discourse around gaming aligns with three distinct philosophies of science: positivism, critical realism, and analyticism, and that each is governed by a different logic. These logics are all internally consistent, but each is quite distinct since they depend on different claims about how the researcher relates to the topic of study and whether we can learn about phenomena we are not able to observe. Positivism seeks to identify causal relationships through direct observation, with the goal of establishing the role of causal factors as a generalizable relationship. In contrast, critical realism focuses on proposing causal mechanisms which cannot be directly observed under current conditions in order to generate proposals about how the multiple factors that make up a specific context drive suggest a specific causal story is promising. Finally, analyticism focuses on generating a model of a specific context—the goal is a useful tool to promote understanding, rather than any claim to generalizable truth. As a result of these differences in understanding how we learn, the three philosophical perspectives suggest different processes and standards for judging claims about how the world works. As a result, games produced to align with one logic are unlikely to satisfy the conditions of a different logic. Implicit in this argument is that games guided by the different philosophies will generally produce different types of information. The question then becomes how these different philosophies can be leveraged to improve the actual practice of game design. While the previous chapter presented the philosophical claims of each approach, these can seem too abstract to guide pragmatic discussions with sponsors about what information a game must produce to be useful.

This chapter seeks to use the philosophical foundations of the previous chapter as a basis for a pragmatic discussion of the primary differences between categories of analytical games. This chapter presents a framework of four archetypes that describe the types of information that can be generated by a game designed for research. As described in greater detail in Chapter 2, I have developed this framework from existing literature on game design for analysis and research more generally, refined and extended based on my own experience as a designer, interviews with other game designers, and publicly available game design reports. I begin by presenting an overview of the archetypes and the characteristics that differentiate them. I then clarify the connections between the game archetypes described below and the philosophies of science described in Chapter 3. Finally, the chapter presents some design considerations and tradeoffs that characterize each archetype.

The premise of the framework is that in designing a game, one often works backwards. First, one considers what information would be helpful for the game to generate in order to answer the

research questions at hand.²³⁷ From there, the designer can then consider how to best structure the game to produce the desired information. Of course, this process is also informed by a range of constraints, including available time, resources, materials, and prior research. Much of the designer's work is to design a game that will generate information as useful as possible for answering the research question given the constraints. However, while constraints will inherently be context-specific, we may recognize patterns in the type of information that games are asked to produce. That is, following the logic above, we can define game types by the information that they generate, which then has clear implications for what design choices and tradeoffs ought to be made.

The information that we desire the game to produce is, ideally, another way of stating the game's purpose and objective. However, experienced policy gamers frequently note that the purpose and objective of a game are a frequent point of sponsor intervention, leading to vague or cluttered guiding statements.²³⁸ Ideally, this is solved by the designer guiding the sponsor to generate tight, focused statements of intent about what information the game ought to produce. However, in practice, game designers are often forced to accept unfocused objectives, and opt to develop a more defined scope for the deliverables with the sponsor informally.²³⁹ In recognition of this reality, I have adopted the convention of talking about the information to be generated, rather than the purpose and objective, to focus designers on the ultimate goal of the work rather than the artifact of bureaucratic processes.

As discussed in Chapter 2, this project develops an *archetype*-based classification scheme. Archetypes provide ideal types that may be used as a model or extreme point of comparison. Few, if any, games will perfectly match only a single archetype—that is, the categories presented below are not intended to be mutually exclusive. Furthermore, while the framework seeks to be comprehensive in describing games for research the fragmented nature of the gaming community and sizable gaps in the publicly available records makes it difficult to ensure this standard is met. New game types may be found in the historical record or identified by future practitioners. Instead, these types are intended as reference points. Thus, it is completely valid to describe a game as more or less like a specific archetype, or indeed, to have characteristics of more than one type.

Without claiming to describe existing games in distinct categories, these types are intended to serve as guides for designers. For example, it is possible to design a game that seeks to generate multiple types of information. However, the design of such a game will be complicated by the tensions between the types. A skilled designer may be able to successfully navigate these

²³⁷ This design process has long been advocated for by Jeff Applegate in his courses on game design. This framing of course assumes that the research question is appropriately answered with a game. As Chapter X discusses in some detail, this is not always the case, and designers should always be on the look out to steer sponsors away from inappropriate combinations of research question and method.

²³⁸ Downes-Martin, "Your Boss, Players and Sponsor: The Three Witches of War Gaming."

²³⁹ I'm indebted to Ed McGrady for working through these ideas with me.

tensions to create useful and credible games (and historically several have), but this framework highlights the difficulty of this task. Thus, designers may find it useful to discipline themselves to scoping games so that they only attempt to produce one type of information.

Overview of the Four Archetypes

This chapter lays out a framework of four archetypes, or ideal types, to describe the type of information that a national security policy analysis game can generate to support research. These types are: system exploration, alternative conditions, innovation, and evaluation.

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 problem. This type of game aligned with past descriptions of games that are used to organize the knowledge of a research team,²⁴⁰ games to suggest what factors and relationships are important,²⁴¹ games to reveal poorly understood dynamics,²⁴² and games to integrate knowledge,²⁴³ identify issues, and build consensus among participants.²⁴⁴

Alternative Conditions: These 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 model 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 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 context²⁴⁵ and well as providing a grounds for theory building about the role of the environment.²⁴⁶ Descriptions of games as experiments are sometimes of this type (though often also describe applications for evaluation, which I treat separately).²⁴⁷

²⁴⁰ Weiner, "An Introduction to War Games." p 25

²⁴¹ Ibid. p 25

²⁴² 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

²⁴³ Parson, "What Can You Learn from a Game?."

²⁴⁴ Perla, *The Art of War Gaming: A Guide for Professionals and Hobbyists*. p 181

²⁴⁵ Weiner, "An Introduction to War Games." p 28 and Longley Brown, *Successful Professional Wargames: A Practitioner's Guide*. p 91.

²⁴⁶ Weiner, "An Introduction to War Games." p 25

²⁴⁷ Parson, "What Can You Learn from a Game?."

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,²⁴⁸ producing innovation and strategic inventiveness²⁴⁹ and to promote creativity and insights.²⁵⁰

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,²⁵¹ testing strategies and plans,²⁵² evaluation, and analysis.²⁵³ Experimental games may also fall in this type, but do not always do so.

Differentiating the Types

These archetypes differ from one another in a number of ways. However, the four types can most clearly be differentiated by two characteristics: the primary focus of investigation and the audience for the game. As with the overall archetypes, these characteristics are not, in fact, mutually exclusive alternatives with a clear outline but rather represent a spectrum. “Ideal types” are not intended to ignore that middle ground, but rather to offer “pure” examples that make factors that are present to a greater or lesser degree more apparent.

The first delineation is between games that are focused on building an understanding of the nature of the policy problem versus those examining potential solutions. Games that focus on the problem seek to better understand the policy system in question, including how different

²⁴⁸ Perla, *The Art of War Gaming: A Guide for Professionals and Hobbyists*. p 181

²⁴⁹ 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

²⁵⁰ Parson, "What Can You Learn from a Game?."

²⁵¹ Weiner, "An Introduction to War Games." p 28

²⁵² Perla, *The Art of War Gaming: A Guide for Professionals and Hobbyists*. p 181

²⁵³ Longley Brown, *Successful Professional Wargames: A Practitioner's Guide*. pp 89-91





stakeholders think about the issue and how the system could evolve over time.²⁵⁴ While such games will often include consideration of what decisions key actors could make, the focus is on understanding how these actors fit into the broader system. In contrast, games that focus on solutions develop and judge potential interventions to understand how they might interact with the broader system. In other words, the focus is flipped, to place more attention on the players' decisions and their influence on the environment and rules of the game, rather than the impact of the environment and rules on the actor's decisionmaking. In policy contexts, this distinction also implies a difference in game purpose—games that focus on the problem are more likely to be descriptive—that is, laying out how different factors are linked together—while games focusing on solutions are more likely to lean to prescriptive recommendations.

The second division is between different audience. Early stage research that is intended to inform the research team and game sponsor is focused on the needs of a relatively narrow audience that is directly engaged with the game. These games general shape decisions about addition research and relatively small-scale investments that are within the sponsoring organizations purview. Alternatively, the analytic output of games can be designed to produce information that would influence on stakeholders beyond the sponsoring organization. The games tend to fall later in a series of research projects and, since they need to persuade a broader range of stakeholders who are more likely to not observe the game directly, the findings tend to devote more attention to being facially credible and definitive. As a result, these games are often intended to have a persuasive effect. Games that seek to influence outsiders credibly should provide greater transparency regarding the path leading to results, or risk being accused of attempting to manipulate consumers by obscuring unfavorable results.

By combining each set of characteristics, we can define relationships between the four archetypes. System exploration and alternative conditions games focus attention on developing an understanding of the problem, while innovation and evaluation games focus on potential solutions. System exploration and innovation games are early stage research generally intended for internal audiences, while alternative conditions and evaluation games further develop research to inform a broader audience. Figure 4.1 illustrates how the four archetypes align with these characteristics.

²⁵⁴ I am indebted to Chris Chivvis (interview, McLean, VA, March 2018) and Margaret McCown (personal correspondence, March 2018) for helping clarify the ways in which my understanding of policy problems in the contest of games depends on a systems approach to understanding problems. For a treatment of systems analysis applied to policy problems, see: Bob Williams and Richard Hummelbrunner, *Systems Concepts in Action: A Practitioner's Toolkit* (Stanford, CA: Stanford University Press, 2011).

Figure 4.1: Game archetypes organized by defining characteristic

	Develop an understanding of the problem	Develop strategies to address the problem
Early research to inform research team and sponsor	System Exploration  Elicit and synthesize mental models of the problem	Innovation  Develop new decision options that break from the status quo
Mature research to influence external stakeholders in a decision process	Alternative Conditions  Detect similarities and differences in decisionmaking based on different stating conditions	Evaluation  Judge the outcomes of player decisions based on a normative standard

It is important to note that these distinctions come with some important corollaries. Because of the artificial nature of game environments and the limited degree of experimental control over players and their interactions, many gamers are deeply uncomfortable using games to support causal or predictive analysis that seeks to extrapolate game results into the real world. However, games focused on solutions, and games providing information to external audiences (and most especially evaluation games that do both,) begin to wade into these dangerous territories. For example, because the designer has a great deal of control over how the game's environment and rules are shaped, an unscrupulous designer could set up a game designed to produce information favorable to a particular position to advocate for a particular solution to outside organizations. Less maliciously, a designer unaware of the potential biases introduced by a specific group of players could over-generalize results from one game to a much broader set of real-world decision contexts, offering poor predictions. As a result, many designers caution that games of these types are difficult to execute well and require care and due modesty in their analytical claims to be credible. Thus, most designers would assert that the different types of games vary in difficulty from easier in the upper left (systems exploration) to harder in the lower right (evaluation).

Beyond the direct consequences of these two distinguishing characteristics of games, there are several other related characteristics that might be used to describe how games of these types differ from one another. These are summarized in Table 4.1. First, the different types of games present different core design and analysis challenges—that is, because they aim to generate

different types of information, there are different tradeoffs that need to be considered. Second, the maturity of the research—that is, how developed our understanding of the issue is—tends to be different across game types. Similarly, the stage of decisionmaking as the focus of game analysis will differ. Finally, the target audience for the knowledge generated by the game differs across types. Unpacking these differences helps to better differentiate between the types.

Table 4.1: Secondary distinguishing characteristics of each game archetype

	System Exploration	Alternative Conditions	Innovation	Evaluation
Core design challenge	Elicit expert thinking in an accurate and traceable way	Maintain control over factors that are not being intentionally varied	Change status quo conditions enough to enable new thinking while retaining realism	Develop processes to generate credible decision outcomes
Core analysis challenge	Capture expert thinking, accounting for analyst bias	Account for the potential influence of factors on similarities and differences	Initial assessment of idea quality	Develop appropriate measures and instruments
Maturity of research	Early exploratory	Mid-term exploratory	Mid-term exploratory	Late refinement and confirmatory
Focus of findings	Framing the problem system	Decisionmaking processes	Decisions	Potential outcomes of interacting player decisions
Target audience for knowledge	Stakeholders trying to understand the problem set	Decisionmaker responding to or setting conditions	Investor in future research	Policymaker selecting course of action

Core design challenge. Depending on the type of information the game is intended to produce, designers grapple with different tradeoffs in making design choices. Weighing the core design challenge to produce each type of information can be a useful guide for design (more detailed consideration of design tradeoffs for each type of game are discussed in the following chapters). In the case of systems exploration games, the core design challenge is how to build a game that elicits expert understanding in a way that is understandable to other players – to allow for exchange of ideas during the game to generate synthesis – and to the research team – to facilitate analysis (discussed below). Alternative conditions games are challenging because of the need to control as many possible sources of confounding variation as possible during design. This is made particularly difficult by the reality of human players interacting both with one another and with the rules in organic ways that are difficult to anticipate. Put differently, where system exploration games are hard because of the need to leave space open to bring in player ideas, alternative conditions games require much more structure in order to enable comparison. The design of innovation games focuses on the challenge of relaxing constraints sufficiently for new ideas to emerge. If the designer simply reproduces the current system, new ideas are less likely to emerge, whereas if constraints that policymakers cannot change, such as the laws of

physics, are removed, the ideas generated by the game will not be feasible as practical actions. In contrast, evaluation games require great attention to designing a credible adjudication system to ensure that game outcomes meaningfully reflect potential real-world outcomes. This difference in design focus across the four types leads to designers to make different tradeoffs, which are explored in more detail in the later chapters of this book.

Core Analytic Challenge. Flowing from these game design challenges, game analysis to produce different types of information varies considerably. In system exploration games, the analytic challenge is often moving beyond simply reporting player discussion to determining how best to update the model of the system based on information gained from the game about how different players understand the policy problem. For example, if players disagree about how a process works, what is the best way to reconcile the differences in perspective? Alternative condition game analysis must grapple with the limits of design—that is, where unplanned variation may complicate our ability to draw conclusions simply through direct comparison. For example, if one group of players had far more interpersonal conflict than the other group, how might it have impacted team performance in ways independent of the key differences in the problem intended by the designer? Analysis of innovation games must struggle to provide a helpful screening of which ideas should be pursued farther: Dismiss ideas too quickly and good options could be discarded; present too many impractical ideas or offer only mild tweaks to the status quo and the sponsor will lose faith in the value of the game. Finally, analysis of evaluation games often struggles to measure game results in a clear and accurate way to support assessment. For a evaluation to be seen as a fair test, a skeptical audience must understand how player decisions were evaluated.

Research maturity. We can also consider the maturity of research associated with each type of game. Games are often combined into broader studies that include either multiple games or games coupled with other techniques.²⁵⁵ While there is not a hard and fast sequence of where games of different types may appear in the cycle of research, generally systems analysis games are run when first trying to understand the nature of a problem, while evaluation games are run later once there is a good understanding of the problem and of potential solutions under consideration. Alternative condition and innovation games fall somewhere in between—they require a somewhat structured understanding of the problem in order to identify factors to manipulate but are only useful when there are still substantial gaps in our understanding of decisionmaking. As a general rule, if we look at Figure 4.1 we expect games in the upper left to occur earlier in a research project than those in the lower right.

Distinctions about where games fall in the cycle of research also suggest some of the tensions that will exist when a game is used to produce more than one type of information. For example, a game that seeks to both develop new solutions and judge their utility is likely to require a fairly

²⁵⁵ For a longer discussion of games in multi-method studies, see: Elizabeth M. Bartels, "Adding Shots on Target: Wargaming Beyond the Game," *War on the Rocks*, 2017.

advanced understanding of the space for innovation and produce relatively immature judgements because there are still gaps in our understanding. At its extreme, this principle also suggests that it will be very difficult to successfully produce information from the same game that explores a system and evaluates policy options. If there are still questions about the nature of the system fundamental enough to justify a game to explore it, it is then unlikely that we would have the necessary level of understanding to produce good judgments about potential interventions in the system. While there may be exceptions to this trend, it certainly holds well enough to serve as a heuristic for designers and sponsors about when a game is being asked to do too much.

Focus. Because the different types of games produce different types of information, often one aspect of the game process is of particular interest. As a general rule, the action of a game flows in a particular order: players receive information about the decisionmaking context, they debate what information matters to their decision and why, they make a decision, and then they observe the outcome of the decision to understand their new context and begin the cycle again. As the name implies, systems exploration games focus on the nature of the problem, so analytic attention focuses on what aspects of the game context matter to players and why. These contextual aspects may include how stakeholders view the same context differently as well as how those understandings change over time. Alternative condition and innovation games examine player choices and the processes by which they are made, placing focus on the second two stages of the game. In the case of alternative conditions games, more focus may be placed on how game conditions influence player decisions, while in innovation games often there is a bit more focus on the decision itself though that can vary.²⁵⁶ Evaluation games focus on the potential outcomes of decisions so have a unique focus on the last phase of the game process. While most games will still include all four phases of game play, one or more may be attenuated in design because it is of lesser importance to the game's focus.

Audience. Finally, there is a pattern in the profile of who the information stemming from the game is intended for and what they intend to do with it. The results of system exploration games tend to inform stakeholders who are trying to understand a problem set. Generally focused on the sponsor (and to a lesser degree, players), these games are about understanding decisionmaking contexts rather than about supporting a specific, immediate decision. The narrower focus of alternative conditions games provides stronger guidance to a decisionmaker to understand the potential effect of conditions on decisionmaking. This may be helpful in anticipating the second order effects of decisions (for example, how an adversary might respond to competing policy choices, or how different stakeholders might react to a change in bureaucratic processes). It may also help decisionmakers understand how decisions will fare in different potential futures. Results of innovation games are most likely to be useful to investors who are determining areas for future research and development, whether that be in technological systems or bureaucratic

²⁵⁶ For example, a game that focused on innovation in the process for decision making might shift the focus of analysis.

solutions. These games help the sponsor decide where to invest in additional research, but on their own are usually insufficient to make the case for major investments. Finally, evaluation game results can inform policymakers trying to select a course of action. The scale of the decision, and thus the degree to which the information from the game must be persuasive to external audiences, may vary, but generally there is a sizable persuasive element in communicating the results of these games.

Games with Characteristics of Multiple Types

As stated earlier, it is possible for a game to resemble more than one archetype. While a skilled designer may be able to balance the competing goals of multiple types to build a game that can successfully generate multiple types of information, as a rule I argue that it is better to limit a game to a single type of information. Because games represent a substantial investment in staff time, the instinct to cover as many issues as possible is natural. However, the discussion of the differences between different types of games suggests why this instinct might be problematic. The differentiating characteristics discussed above highlight some key tensions. For example, the game types in the bottom row of the matrix depend on a more refined base of research than those in the top row. Attempting to answer both types of questions at the same time suggests a mismatch between the maturity of research and one of the two purposes. Take for example a game that tries to generate both systems exploration and alternative conditions information. Either the understanding of the problem is immature, risking that whatever factors are selected in the alternative conditions games are not very useful, or the understanding is more mature, and relatively little can be added by the incorporation of players' mental models. In either case, the game is much more able to generate one type of information than the other, and thus resembles one archetype much more closely than the other. As a result, I argue that it is better to try to focus on only one type of information, rather than over promise and under deliver.

Connecting Philosophy of Science for Games to the Archetypes

As discussed in Chapter 3, games may reflect several philosophies of science, each with its own logic. I identify three philosophies that have been employed by game designers; positivism, critical realism, and analyticism. It is therefore reasonable to ask how the proposed typology interacts with these different philosophies of science. As I argue below, some archetypes fit more harmoniously with some philosophies, while others give rise to tensions. Without going so far as to say any archetype is incompatible with any philosophy, there are clearly types that fit more naturally with certain logical approaches to knowledge production. Since designers tend to naturally gravitate to one philosophical position, these suggest that some types of games may be more natural for a designer to produce.

System Exploration

To a positivist, system exploration games have profound limitations. The openness needed to allow players to contribute their mental models of the problem makes it difficult to clearly isolate key factors or causal relationships during the game. Much like the positivist view of a single case study, games might be useful as a source for inductively generating hypotheses²⁵⁷ but are unlikely to advance causal claims.²⁵⁸ However, they have the added defect of artificiality, making them inferior to a real-world case for this purpose.²⁵⁹ Thus, one strongly inclined toward positivism is likely to only use systems exploration games as a means of hypothesis generation about phenomena for which no real-world case study can be generated.

In contrast, system exploration games align quite well with the logic of analyticist research. 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,²⁶⁰ practitioners of analyticism will have no problem making a case for the pragmatic usefulness of games to explore issues.

Critical realists are likely to find more value in systems exploration games than their positivist counterparts, though perhaps not be as strong advocates as their analyticist counterparts. Because games allow for the observation of specific processes, they are ideal for tracing out causal pathways. This makes them an attractive option for mechanism-based research for which rich contextual data enabling abduction is critical. However, the inherently unobservable nature of critical realist mechanisms may make the artificial nature of games more of a concern. It is one thing to abductively infer an unobservable phenomenon from real world observations, but to do so from the interactions of an inherently fictitious environment and actors may reduce confidence in the value of the causal claim. Thus, games are likely to be seen by critical realists as an imperfect means for generating causal claims, with appropriateness dependent on the specific topic.

²⁵⁷ Levine, Schelling, and Jones, "Crisis Games 27 Years Later : Plus C'est Deja Vu." pp 12-14

²⁵⁸ King, Keohane, and Verba, *Designing Social Inquiry: Scientific Inference in Qualitative Research*. pp 210-212

²⁵⁹ Parson, "What Can You Learn from a Game?." pp 237-238.

²⁶⁰ For example of a famous assertion of the utility of gaming, see: Levine, Schelling, and Jones, "Crisis Games 27 Years Later : Plus C'est Deja Vu." p 23

Alternative Conditions

The alternative conditions approach fits extremely well with the positivist research agenda. 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,²⁶¹ this can be managed in ways that are consistent with quasi-experimental traditions that are popular in a wide range of positivist research projects.²⁶² 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.

In contrast, for both critical realists and analyticists, this approach sits uneasily within their logics of inquiry. Both traditions are interested in complex contexts with many interacting factors that need to be considered holistically. Attempts to isolate and then vary specific factors in isolation breaks this commitment. For critical realists interested in causal mechanisms, not counterfactuals, the comparative project here provides relatively little additional leverage to help understand the causal process. Multiple cases may help illustrate how a mechanism works across a small subset of cases, but intentional variation is really only helpful in defining scope conditions for the universe that can be generalized to. This most likely can be done better using other logical tools. For analyticists who are not invested in generalization as a goal of science, cross-case comparison offers no advantage for inference. While it may be interesting to see if the same model is generated across multiple game settings, that is more appropriately the role of multiple games analyzed together rather than any type comparison across games with structured variation. In short, this archetype generates comparative information that is less valued by these two philosophical approaches.

Innovation

Much like system exploration games, the hypothesis-generation focus of innovation games makes them a somewhat uneasy fit with positivist approaches. Because of the artificial nature of games, inductively generated hypotheses may gain less traction with adherents of the approach. Furthermore, novel approaches are unlikely to be reducible to discrete factors; if the solution was that simple it likely would have been suggested in the past. In these cases, the task would be more a matter of eliciting existing but neglected good ideas from participants, and thus more properly be thought of as a system exploration game rather than a true innovation game. As a result, innovation games are not likely to be popular with positivist practitioners.

²⁶¹ Parson, "What Can You Learn from a Game?." p 238.

²⁶² Elizabeth M. Bartels to Paxsims, 2015.

Perhaps the most natural alignment for innovation games is with the critical realist approach. 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.

Analyticist approaches to innovation games share some characteristics with system exploration games, but do not align as well. On one hand, the pragmatic orientation of analyticism is well suited to the task of developing new approaches that are simple enough to be easily communicated outside the game. On the other hand, because that model cannot be assumed to apply elsewhere, the value of the game for policymakers is constricted. As a result, the specific problem at hand will likely deeply influence analyticists' assessment of the utility of an innovation game.

Evaluation

Evaluation games have an imperfect fit with positivist 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 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 judge 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.

Different Games for Different Philosophies

Taken together, the arguments above suggest that some archetypes are better aligned with some of the philosophies, as summarized in Table 4.2. That is not to say that poor alignment prevents a clever researcher from developing a game-based approach using that philosophy, but it does suggest they will have to do more work to defend why the approach is logically consistent to other researchers who adhere to that philosophy. Given the dominance of positivism in other approaches to policy analysis, consideration of potential objections from that logic may be particularly important for a designer working outside that philosophical tradition, or for a positivist designer using a game archetype that is a less comfortable fit for the positivist logic of inference.

Table 4.2: Degree of alignment between the three philosophies and four archetypes

Archetype	Positivism	Critical Realism	Analyticism
System Exploration	<i>Weak</i>	Some	Strong
Alternative Conditions	Strong	<i>Weak</i>	<i>Weak</i>
Innovation	<i>Weak</i>	Strong	Some
Evaluation	Some	Some	Some

Another implication of the differences in how each game type supports the logic of the different philosophies is that designers who adhere to a specific philosophy may find one style of game more natural to design than a type that fits less well into their philosophy. For example, a strong positivist may be inclined to see system exploration as a less useful enterprise than an analyticist scholar would. In contrast an analyticist is likely to see alternative conditions games as needlessly fussy, while a positivist would see their structure as critical to support cross-case comparison. These tendencies are not absolute—many designers are capable of adopting different logics or of identifying value for a game type within their logic. However, the tendency is useful to note, if only to be sensitive to the potential to dismiss too readily the utility of games that exist outside our preferred philosophical frame.

Design tradeoffs

There is no fixed recipe for moving from the game's purpose and objectives to its design. It is left to a designer to assemble mechanics, data, and people to craft an appealing game.²⁶³ However, that is not to say that no guidance can be offered to designers to steer them towards better and away from worse choices to achieve their objectives. Here, it can be helpful to think about the designer's job in terms of trade-offs. No matter the purpose of the game, the design process seeks to build a game that instantiates a model of the problem at hand. Design choices can either align with, or deviate from, that model, making for better or worse design.²⁶⁴ However, because games must be run in the service of practical ends, available resources in a wide variety of areas impose constraints that a designer must also work within. As a result, much of a designer's work requires making tradeoffs between what is dictated by the ideal research approach and what is feasible given constraints.²⁶⁵

Given this frame, one way to provide guidance to designers is to explore what tradeoffs are likely to be more or less problematic to the usefulness of findings, given a particular goal of a game. Because it is not usually possible to run a game in which no practical compromises are made, identifying the tradeoffs that are most likely to undermine the findings can allow for smarter design choices. When problematic choices cannot be avoided (as is often the case), advanced consideration can sometime develop mitigations within the research design, or at least allow for thoughtful discussion as part of analysis.

In the following four chapters, I discuss design tradeoffs inherent in each of the four archetypes, illustrated by historical examples. I organize the discussion of tradeoffs along the three key design elements that make up the model of the game: the environment, actors, and rules. The environment refers to the setting of the game that frames the central problem players seek to resolve. This includes not only the narrative scenario that traditionally describes the events leading up to the start of the game but also the information provided to players about the state of the world during the game. This comes in a range of forms, including narrative, visualizations and databases that together create the player's understanding of the setting.

Second is the actors, represented by players, who have resources they can use in an attempt to resolve the problem to their favor. The modeling of the actor includes the frame provided by the designer, such as the decision of which actors are represented and what level of aggregation, and what guidance is provided about each actor. However, perhaps more important are the human players who fill the role, whose mental models fundamentally shape what choices they

²⁶³ Perla, *The Art of War Gaming: A Guide for Professionals and Hobbyists*.

²⁶⁴ Bartels, McCown, and Wilkie, "Designing Peace and Conflict Exercises: Level of Analysis, Scenario, and Role Specification."

²⁶⁵ It is important to note that this is not an issue unique to gaming. For example, medical studies are often notable for quite small numbers of participants—researchers use the minimum number needed to test for a particular effect, and no more, to save on time and costs of the study.

make in the game. Taken together, these components describe what actors are in the game, their objectives, the resources available to them to pursue those objectives, and what decisions they can make in the course of game play.

Third come the rules that structure how the actors' decisions interact with one another and the environment. These include both rules that structure interaction during the game (for example, who may communicate with whom and how) and rules that are used during adjudication to decide the outcomes of decisions. These rules may range widely from very rigid to open, and from complex to simple depending of the needs of the game model. They can also be implicit or defined by the players, such as when players assert that a particular action is not permissible. Regardless of format, the game rules will shape how game play evolves, including how players learn information, how they make decisions, and the consequences for both actors and the environment of those decisions.

The following chapters provide a more detailed discussion of tradeoffs in each archetype. To make the practical consequences of these choices clear, each chapter presents several games to serve as illustrative examples. These are drawn from archival research, interviews with senior game designers, and my own work as a policy game designer. These chapters do not (and in fact cannot) offer a ridged "recipe" for game design, but instead illustrate how different designers have navigated tradeoffs within a class of games.

Chapter 5: Designing Games for System Exploration

As detailed in Chapter 4, system exploration games elicit and synthesize how players understand a problem in order to develop a better model of the policy issues, opportunities and constraints. These games are common in early stages of research, intended to form a foundational understanding of the problem system for later studies and analysis. As such, they are usually of most value to the immediate research team, players, and sponsors. The outputs of systems exploration games align best with analyticist approaches to research with their focus on building useful models. Additionally, some critical realists might appreciate system exploration games as a means of developing hypotheses about the nature of the policy problem that can then be further examined using other means. Positivists are likely to see this as an expensive approach to hypothesis generation that does not offer much improvement over other, cheaper tools.

Because the primary finding from a systems exploration game is based on the mental models provided by game participants, this type of game is strongly defined by the players charged with representing different actors. Since participants in the game provide the mental models to be captured, the quality and diversity of player understanding is critical for strong game results. Weak players will provide a poor model of the policy system. Game environment and rules must channel player input towards a common problem while enabling elicitation and documentation of player beliefs. If the structure of the game is too rigid, players do not have sufficient freedom of action to contribute their understanding of the problem. The game results will closely mirror the designer's understanding so little will have been added by playing, rather than simply building, the game.²⁶⁶ On the other hand, if the game is under-structured, it can lose focus—players address different problems and end up talking past one another, discussion turns away from decisionmaking and becomes more academic, and key data will not be captured. As a result of these considerations, systems exploration game designs will generally prioritize flexibility and player engagement compared to games designed for other purposes.

As a result of these tendencies, games for this purpose are often closely identified with “free-form”²⁶⁷ or seminar-style game formats. This conflation is somewhat misleading since these loosely structured formats can be used for other purposes while more structured formats may also be used to explore systems. Similarly, “political-military” games which focus on the relationship between political and military decisionmaking at the strategic level often, but not always, focus on developing a better understanding of how different stakeholder see complex problems. Therefore they tend to align to a considerable degree (but not absolutely) with the

²⁶⁶ Which is not intended to neglect the importance of the game development process as a tool for research by the design team, a point that will be returned to in Chapter 10.

²⁶⁷ William M. Jones, "On Free-Form Gaming," (Santa Monica, CA: RAND, N-2322-RC, 1985).

types of information produced by system exploration games. To the extent writing on free-form, seminar-style, and political-military games focus on the specific tasks of gaining an understanding of a policy problem, they can inform approaches to designing system exploration games I make reference to some key texts throughout the following discussion.

This chapter discusses these design tradeoffs in greater detail, using two games to illustrate design decisions in practice. The chapter introduces the example games: the Project Sierra games run by RAND for the U.S. Air Force in the 1950s to look at limited war in Jordan and a recent RAND game focused on better understanding “grey zone” competition. I then draw on these games to illustrate general arguments about better and worse design tradeoffs related to the environment, actors, and rules of systems exploration games.

Overview of Example System Exploration Games

While a definitive census of all games run by the Department of Defense is not available, consensus among practitioners is that system exploration games are likely the most prevalent category of games supporting policy analysis. Moreover, because of their use to study a wide range of problems, there is likely greater diversity in these games than in others. As such, a small number of games is not representative. Instead, the games below were selected to illustrate some of the diversity present. Project Sierra featured multiple games, focused on military operations between two teams, with most players drawn from the research staff of the project. In contrast, the Grey Zone game examined a wide range of options short of war in competition between three teams in more structured game play. Thus, these games show different approaches to resolving common issues in the design of system exploration games.

Exploring Across Escalating Scenarios: U.S. Air Force/ RAND Project Sierra Middle East Games—Jordan Series

Project Sierra was an early RAND effort to explore limited war—conflict that involved the U.S. but did not involve strikes on the homeland. Over the course of the four-year project, game design varied somewhat as teams experimented with new approaches—here I have opted to highlight the design of later games, focused on conflict in the Middle East run in 1957 and 1958, to provide a consistent snapshot of the game design approach.²⁶⁸ Characteristically for the project, multiple, intentionally varied games were run. Across the games, different political limitations determined what military actions were allowed in order to look at how these changes shaped limited war. Along with more traditional military decision-making, game play also focused on political, economic, logistics, and intelligence factors and key findings reflected these

²⁶⁸ A more general discussion of the series as a whole is offered as part of the history of RAND gaming in Chapter 9.

categories. The goal of the effort was for the research team to develop a better understanding of the problem of limited war, that could then inform additional research and analysis.

Seven games examined an invasion of Jordan by Syria (supported by the USSR) several years in the future with U.S. support to Jordan prior to hostilities ranging from limited logistics support in the early games to the authorization to use nuclear weapons in the last.²⁶⁹ The games featured two teams—one representing Syria and its allies (the red team) and the other the U.S. supporting Jordan (the blue team) both staffed by members of the research team with a range of expertise in different operational areas. There were also umpires moderating the course of play.

Play proceeded in several stages. After receiving an update from the control team on the current state of the game world, each team deliberated to establish their estimate of the situation, political and military objectives, and the general plan for achieving those objectives.²⁷⁰ This general approach was then approved or disapproved by the control team. After approval, players developed a more detailed implementation of the plan with particular attention to the actions to be taken by the Air Force in support of approved objectives. After reviewing the detailed plans of both sides, control assessed the outcomes of attacks as well as what resulting information would be available to both sides. Since both teams operated with only the information that would reasonably be available to them, control had to also maintain a view that includes the adjudicates state of the world (sometimes called “ground truth”) throughout game play.²⁷¹ This process is illustrated in Figure 5.1. Players were provided detailed inputs and were required to generate fairly specific outputs particularly during later stages of detailed planning. Moves tended to be conveyed in written format such as “logs, mission sheets, or overlays indicating forces to be committed, the objective, time-scale, and other factors.”²⁷²

Because the Jordan games were run late in the Project Sierra series, a range of procedures developed by the game design team improved the ability of the research team to learn from the games. First, the team had developed the flexibility to open up the actions available to participants to enable greater player choice at key junctures while still retaining the ability to impose political limits on actions and hide adversary intentions and actions for realistic play.²⁷³ Procedures also had to be created for the control team to be able to formalize key tasks like communications to and from the player teams to ensure that the information needed for the player to make decision and for control to determine the outcomes of player decisions was available without distorting game play by revealing information that would realistically be

²⁶⁹ Milton G. Weiner. "War Gaming Methodology: Sierra near East Series." (Santa Monica, CA: RAND Corporation, D-4926-PR, 1958). p 1

²⁷⁰ Ibid. p 10

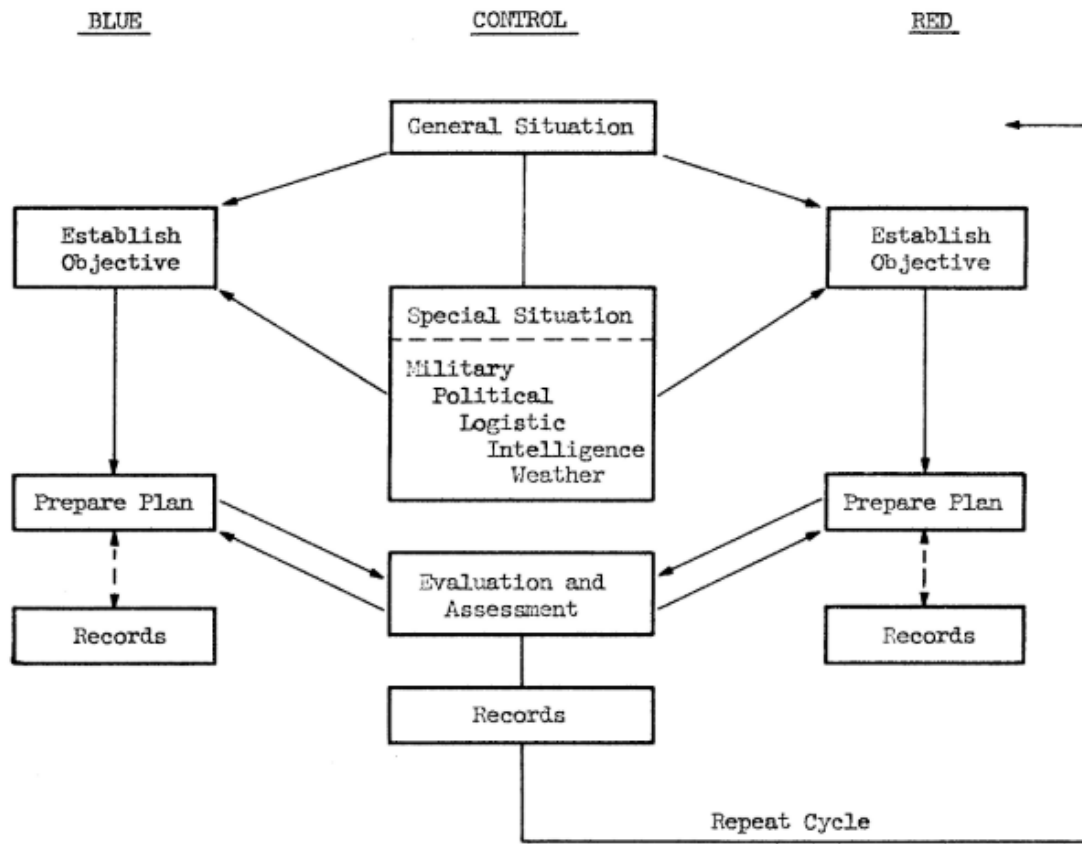
²⁷¹ Ibid. pp 13-14

²⁷² Ibid. p 12

²⁷³ Ibid. pp 30-36

hidden from key actors.²⁷⁴ These approaches allowed the analytical team to focus on examining key points in the game where alternative actions were considered so that these decision points could be further explored in later analysis and games.²⁷⁵

Figure 5.1. Process of Project Sierra Games



Source: Weiner, "War Gaming Methodology." p 56

Building to a Structured Seminar Game for System Exploration: U.S. Army/ RAND Gray Zone Wargame

Recent RAND gaming efforts have supported the Department of Defense's current exploration of "gray zone" tactics—that is:

ambiguous political, economic, informational, or military actions that primarily target domestic or international public opinion and are employed to advance an

²⁷⁴ Ibid. pp 36-39

²⁷⁵ Ibid. pp 43-47

nation's interests while still aiming to avoid retaliation, escalation, or third-party intervention.²⁷⁶

Following the Russian invasion of Ukraine there was a great deal of concern that Russia would be able to undermine the dominance of NATO in Europe through this suite of tactics. There was less consensus about what tactics should be considered within the suite or the exact challenge they pose to the United States. One effort for the Army to better understand the nature of these “gray zone” conflicts focused on their use by Russia in the Balkans to undermine European cohesion. The game featured three teams staffed by RAND researchers who specialized in relevant areas of defense and intelligence policy. A Russia team attempted to gain influence in the Balkans while also undermining NATO while two teams representing the United States and Europe had to work together to defend against Russian activity without inadvertently starting a war.

The design of the game described here was the output of two previous games, each design using the results of the previous game to better develop a model of what actions should be included in the “gray zone tactics” suite and how outcomes should be determined.²⁷⁷ In the final game's structured format, potential player actions were represented on cards as shown in Figure 5.2. Players placed cards on a timeline thus stipulating whether an action was intended to have short-term or long-term effects as well as where it was occurring. The outcomes of these actions was adjudicated using a series of probability distributions (referred to, using the commercial hobby gaming term, as a combat results table or CRT), and outcomes were displayed on a central board as shown in Figure 5.3. Results were presented at the level of individual countries as well as the broader balance of power.

The Russian team began by selecting three pivotal countries to focus on. This was then relayed to the blue and green team.²⁷⁸ In each turn of play, the three teams develop a strategy to achieve their assigned objectives and select a fixed number of short- and long-term action cards to play in each of the three pivotal countries.²⁷⁹ During the deliberations, the United States and European teams could coordinate with one another during planning to secure support for their approach.²⁸⁰ All teams then placed their action cards face down on the timeline. Each team explained their strategy, revealing any overt actions. Covert actions remained face down and were visible only to the control team. Each action was then adjudicated based on the defined probabilities of success and the new situation represented on the game board to inform the next round of planning.

²⁷⁶ Becca Wasser et al., "Gaming Grey Zone Tactics," (Santa Monica, CA: RAND Corporation, RR-2915-A, 2019). p 2

²⁷⁷ Ibid. pp 10-11

²⁷⁸ Ibid. p 32

²⁷⁹ Ibid. pp 32-33

²⁸⁰ Interview with Becca Wasser, Washington DC, January 2020.

Figure 5.2: Sample Action Cards from the Gray Zone Game



NOTE: N/A = not applicable.

Source: Becca Wasser et al., "Gaming Grey Zone Tactics," (Santa Monica, CA: RAND Corporation, RR-2915-A, 2019) p 31.

Figure 5.3: The Gray Zone Game Board



Source: Becca Wasser et al., "Gaming Grey Zone Tactics," (Santa Monica, CA: RAND Corporation, RR-2915-A, 2019) p 21.

Design Tradeoffs Related to the Game Environment

The primary purpose of the environment of a system exploration game is to focus player attention on the correct problem in a context that is seen as credible by players—that is, that “the particular event or situation *could* occur under the conditions specified.”²⁸¹ This is inherently tricky as often different mental models will frame an issue differently with different facets of the environment being relevant to decisionmaking. Picking the wrong scope for the game may cause analysts to miss key aspects of the problem, or in more extreme cases, cause participants to balk at participation because the problem is so mis-framed as to be unrecognizable. Since the design of the game environment plays such a pivotal role in shaping what players will contribute to the game, it can be useful to think of this process as setting the initial parameters of the collective model building exercise. In other words, the game development in general, and the game environment in particular, will shape the information that will be generated by the game as a whole.²⁸² Three challenges stand out in this process: the need to select the game environment, scaling the environment so players can engage with the problem of interest, and the need to build a credible environment early in the research process.

Challenges of Selecting the Game Environment

While selecting the correct environment for the game may seem obvious, often it is more difficult than it seems. Much of the cost of developing a systems exploration game is the time to research and assemble information about the environment. Often it is only possible to explore a single environment thus making selection highly influential on the results of the game. Because system exploration games are about gaining a better understanding of interaction among elements of a policy problem, when only one environment may be explored, it is generally best to select a clear example of the policy problem of interest.²⁸³ At the same time, if a particular environment is extensively studied, a game to explore the general mechanics of the policy system is not likely to add much value. This logic argues for selecting an environment that is a fairly clear example of a policy problem but has not been previously studied in great detail. In the case of the Gray Zone game, the concept of the “Gray Zone” was underdefined. Since much of the discussion driving what the term should mean was driven by discussions about Russian behavior in Eastern Europe, that was a natural environment to select for the game.²⁸⁴ However, at

²⁸¹ Milton G. Weiner, “War Gaming Methodology,” (Santa Monica, CA: RAND, RM-2413, 1959). pp 17-18

²⁸² Interview with Margaret McCown, Washington DC, July 2018.

²⁸³ In social science terminology, this can sometimes be referred to as selecting on the dependent variable—that is, you are picking a case specifically because it has the causal outcome of interest. While this can be seen as a shortcoming in other research approaches, in the case of systems exploration games that are likely to be run under an analyst logic that makes no move to generalize the results to a broad universe of cases, it is an appropriate analytic choice (for clarifying this point, I am indebted to Margaret McCown).

²⁸⁴ Wasser et al., “Gaming Grey Zone Tactics.” pp 1-3

the time the game was run, most previous work had focused on either potential Russian aggression in Ukraine, because of the events of 2014, or Russian actions in the Baltic—meaning Russian actions in the Balkans were understudied.²⁸⁵ This made the Balkans an environment where Russian “gray zone” tactics were likely to be seen but where the results of a game would add distinct value by looking at an understudied environment.

When multiple environments can be studied, the problem of selection is more analogous to that of case selection in other types of qualitative research. This requires carefully defining the characteristics that will circumscribe cases of the “type” and then intentionally selecting a strategy for prioritizing which of those options are examined.²⁸⁶ In the case of Project Sierra, the research team was clear to define the characteristics of a “limited war” environment that would further research, including existence of both U.S. and Soviet interests sufficient to justify meaningful intervention, but not so great as to engender general warfare involving the homelands of both countries. They also considered practical issues—since the U.S. Air Force was the sponsor, they looked for cases which would require a meaningful role for USAF in the U.S. intervention that would also engage a range of military, political, and economic leavers of national power. Within these constraints, a great deal of variation in the type and level of conflict was used, intentionally introducing a wide range of different scenarios.²⁸⁷ This system allowed the research to be clear about what characteristics of the environment defined the scope of “limited war scenarios” while still recognizing the great diversity of conflicts in the category by selecting diverse cases.

Finally, too often, sponsors are tempted to stipulate the environment of interest, and even generate the specific narrative scenario leading up to the start of the game, rather than allowing the research objects to dictate what environment is most helpful to the research question at hand. Often, this requires interrogating sponsors about why they think the proposed scenario is important to better understand what the other potential environments might be based on the factors and assumptions that drive the sponsors interest.²⁸⁸ In simple cases where the selected environment is reasonable, this discussion provides the information needed to transparently describe the factors that drove the selection of the environment to the consumer of game-generated analysis. In the more difficult case of the desired environment being a poor fit for the research objectives, this information can be used to explain to the sponsor the potential

²⁸⁵ Ibid. pp 16-17

²⁸⁶ Excellent references on case selection approaches include: Beach and Pedersen, *Causal Case Study Methods: Foundations and Guidelines for Comparing, Matching, and Tracing*.; George and Bennet, *Case Studies and Theory Development in the Social Sciences*.; Matthew Lange, *Comparative-Historical Methods* (Los Angeles, CA: SAGE, 2013).; and Robert K. Yin, *Case Study Research: Design and Methods*, 5th ed. (Thousand Oaks, CA: Sage, 2014).

²⁸⁷ Weiner, "War Gaming Methodology."pp 7-8

²⁸⁸ Downes-Martin, "Your Boss, Players and Sponsor: The Three Witches of War Gaming."

weaknesses of the approach and provide a (hopefully) compelling argument for alternative environments.

Challenges of Scaling the Game Environment

System exploration games are often scaled to explore problems that exist across many domains. As a result, the game environment must often include information drawn from many different areas. This can create challenges both for the design team and players to integrate the information in a manner ensuring enough specificity to generate concrete recommendations while not becoming overwhelmed by details. The design notes from the Project Sierra games noted that the need to consider political, economic, logistical, and intelligence inputs as well as military concerns greatly increased the number and variety of game inputs. This complicated planning.²⁸⁹ Researchers noted that because players had considerable latitude to set their own political-military objectives, the salience of any specific input to players could vary dramatically depending on its relevance to the specific objective selected.²⁹⁰ In other words, information that was ignored in one game might be key in a different iteration. The same issue challenged the designers of the Gray Zone game who eventually opted to narrow the geographic scope of play to better focus players.²⁹¹

In practice, this often leaves game designers with an unattractive choice—devote considerable pre-game effort to developing inputs about the environments that may never be relevant to players or face the need to generate specifics about the environment on the spot as players request information. Worse, both conditions can occur in the same game—failure to correctly anticipate player needs resulted in a great deal of prior information about the environment being generated only to face players making unexpected demands that required different information to be generated quickly.

One approach to ensure that necessary inputs are available is to develop the game in stages, so that feedback from outside the design can be incorporated. For example, in some of the Sierra games, decisions about red investments and posture included in the scenario were generated through “pre-gaming” by asking red players to postulate an initial red strategy then determining what changes would be needed to move from the current Soviet strategy to achieve the desired starting position. This information was then relayed to the blue team to allow them to make realistic counter investments, which in turn was relayed to the red team who was allowed to modify their plan.²⁹² Such an approach requests a substantial additional investment of time by the players in terms of longer commitment to the project and hours spent in game-related

²⁸⁹ Weiner. "War Gaming Methodology: Sierra near East Series." p 19

²⁹⁰ Ibid. p 20

²⁹¹ Wasser et al., "Gaming Grey Zone Tactics." p 11

²⁹² Weiner, "War Gaming Methodology." pp 12-15

activities. However, more resilient projections that better align with player's mental models are gained.²⁹³ As an alternative, the Gray Zone game used multiple games each with different players. Between each game the design team refined what actions were allowed and how they were articulated based on the play of the proceeding game. In general, this meant that the first games relied far more on play knowledge of the environment, while later games provided more structured information about each country so that players were all working with the same understanding of the threats and opportunities in play.²⁹⁴

Challenges of Building a Credible Environment

Designers then need to populate the game environment with information that will be credible both to players and later consumers of the game. This is most obviously challenging when the game must project far into the future, since more assumptions must be made about how “future history” will unfold. However, even shorter-term games can struggle to provide useful, credible information to players. Good design requires research on the part of the designer, not only of general subject matter resources but also on the potentially different perspectives the players. Designers of systems exploration games should expect to conduct a literature review and interview multiple subject matter experts as part of the game design process. However, designers of system exploration games should expect this process to yield many unknowns, even on basic factual questions. This uncertainty must be taken into account when designing game materials that present the environment to players.

Often, because systems exploration games occur fairly early in the research process before the team has developed a fully structured understanding of the system, it can be easier to use less structured approaches to communicate about the environment. For example, it's fairly common for games of this type to focus on a written or briefed scenario rather than developing a detailed game environment such as the counters and maps of a traditional board game. This more open form of communication leaves more space for players to intersperse their own understanding of key concerns but may lead to instances in which players project different understandings of the same factor into the game. This can be a particular problem if teams are separated into different physical spaces—often some mechanism such as “intelligence reporting” will be required to share information between groups so common information is updated.

However, where possible more structured tools like maps, boards, and counters can be invaluable to rapidly communicating the current state of play and generating shared situational awareness. The early Gray Zone games in which players supplied the majority of game information suffered from a lack of direction, in part because the scope of action was so broad players had difficulty making decisions and tracking outcomes.²⁹⁵ Later games built out a much

²⁹³ Ibid. pp 15-16

²⁹⁴ Wasser et al., "Gaming Grey Zone Tactics." p 22

²⁹⁵ Ibid. p 11

more specified environment at the cost of considerable research on the part of the design team both from the previous games and other research. This included the use of a board that displayed high level information about each country as well as country fact sheets that could provide more narrative information in a format that was still relatively consistent.²⁹⁶ While structuring information in this way is time consuming for the research team, it pays dividends in focusing players' attention.

Design Tradeoffs Related to the Game Actors

System exploration games depend deeply on identifying the actors in a game in general and the selection of players in particular. The participants of the game play a key role in contributing new information to build up the policy problem model that is the key game output.

Scoping What Actors are In Play Shapes System Exploration Games

Because of the strong role that players take in shaping how actors in system exploration games are depicted, it can be easy to gloss over the role game designers play in shaping what actors are included and how they are depicted. They must synthesize “existing knowledge of the nature and operative norms of the actors”²⁹⁷ to the extent possible so that the game can focus on generating new information rather than rehashing conventional wisdom.

Selecting the number of teams and what actors they represent requires determining which actors can make decisions that shape the policy problem and who's decisionmaking is of interest given the research question. Often the default is assumed to be two nations or alliances, each represented by a group of players in competition with one another,²⁹⁸ such as was used in Project Sierra. However, systems exploration games focused on multiple actors with both conflicting and complementary objectives are not uncommon, particularly in games exploring the different reaction of parts of a single government or alliance to a crisis. A game focused on alliance cohesion might opt to represent the members of the alliance separately, rather than as a block, to better examine dynamics between allies. The RAND Gray Zone game did this to some degree by separating the United States team from its European allies.²⁹⁹ However, the game designers acknowledged that it might have been preferable to further subdivide the European team to better capture these alliance dynamics.³⁰⁰ Similarly, a game examining the options available to the U.S. in response to an international crisis might opt to represent several key departments as separate teams to explore gaps and seams in preferences and capabilities. While events focused only on

²⁹⁶ Ibid. p 22

²⁹⁷ Jones, "On Free-Form Gaming." pp v-vi

²⁹⁸ Ibid. p 5

²⁹⁹ Wasser et al., "Gaming Grey Zone Tactics." pp 17-18

³⁰⁰ Ibid. p 40

different perspectives within a single government are sometimes considered to not have sufficient competition elements to be considered a game, the presence of competing objectives and tool preferences produces dramatic tension between these actors.³⁰¹

Another key choice is the specificity to which actors, particularly those within a team, are modeled. Designers should be clear about both the breadth of responsibility (said differently, how many issues are controlled by the actor) and level of control players can exercise. Both will shape what decisions the players can make. Design choices may range from offering very specific guidance about what individual or office each player is intended to depict to establishing broad teams representing countries or departments and allowing players to informally represent the interest of different components with broad guidance such as “consider the relevant actors”.

Generally, less-structured approaches are used in systems exploration games, again, with the goal of allowing players to add their own expertise and experience. However, even a broad frame must specify the range of decisionmakers appropriately or risk that the model emerging from the game will be missing major elements of the problem system. As a result, it is important to be mindful of the cues provided to participants about the scope of their role. These may include the range of participants invited, the taskings provided to players, and the elements of the environment that are highlighted. Designers must not only be conscious of these potential effects during design, but during analysis must carefully consider how the game’s scope may have produced a specific model.

Depicting Actors Not Assigned to Players

Actors that are not assigned to players are generally assumed to be in the domain of control. Designers have a range of options in how the “play” in the game. Which actors will prove to be most influential is often a key question of system exploration games. Considerable prior attention should be given to how contingencies in which a control-represented actor turns out to be critical to decision making should be managed. Approaches may range from careful documentation of in stride judgements by the control team so that their assumptions can be revisited in later analysis to reassigning players with appropriate experience to represent the role in the course game play.

A particular variant of this decision is whether players will have full authority to execute their strategy as they wish, or if the control team retains the ability to veto specific moves. In most political-military games, the ability to veto decisions is usually modeled by the control team retaining “national command authority” and having players represent lower echelons of decisionmaking. If players have national command authority decisionmaking capability, they are able to take far more aggressive actions than if the control team can serve as a check. Often retaining control over key decisions within the control team is important to ensuring players make decisions within the scope of player objectives. For example, in the Project Sierra games, where a primary difference between variants of the game was the degree of U.S. involvement,

³⁰¹ Interview with Margaret McCown, Washington DC, July 2018.

retaining authorization of key capabilities was an important tool to keep players within desired limits conforming to the research objective.³⁰²

Player Selection is Critical in System Exploration Games

Generally, in a systems exploration game, players should be recruited that have more experience than the designer. In the Project Sierra games, reports on the experience and position of the players stressed their expert judgement: “To function realistically, the players must have firsthand knowledge of military and political organization, procedure, and doctrine.”³⁰³ For the game results to be reliable the judgement represented in the planning of the players has a reasonable basis.”³⁰⁴

Diversity is also key--if players all have the same perspective there is less benefit to the interactions and debate between different players. Moreover, designers stressed that because of the range of issues that needed to be considered as part of planning, it would have been impossible to find a single player with all necessary skills. Instead, groups of players with experience in different military domains, as well as political, economic, logistical, and intelligence functions were recruited to build a suite of experts who could cover issues credibly.³⁰⁵ Vigorous debate provides a natural opportunity to capture not only what players believe but is one of the easiest ways to elicit different player perspectives. Good facilitation can assist, but players with diverse, deep experience will be able to ask questions and raise issues that even the best-prepared facilitator will not know to ask. For example, the Sierra games found that “it is not infrequently the differences in emphasis that arise between [player] points of view that serves to isolate a problem which can be studied intensively by the gaming process.”³⁰⁶

However, selection of players will generally require making tradeoffs between the credibility of the resulting model and the practical realities of gathering individuals whose time is in high demand. Politics, limited budgets, and set time-lines will almost always prevent the designer from assembling the ideal group. The choice is which players to prioritize and where to accept a player who less closely resembles true decisionmakers on the issue. Depending on the problem, different experience may be more or less important. A group solely of senior government personnel and another of academic experts with experience studying the conflict may be considered equally expert, yet their experience differs. The resulting models would be credible and useful to different applications. That built by interagency figures would include a solid understanding of organizational interests and resources as well as an understanding of the

³⁰² Weiner. "War Gaming Methodology: Sierra near East Series." p 25

³⁰³ "War Gaming Methodology." p 5

³⁰⁴ "War Gaming Methodology: Sierra near East Series." pp 19-20

³⁰⁵ Ibid. p 22

³⁰⁶ Ibid. p 23

preferences of government. institutions. Academic experts would produce a model that may benefit from structure provided by models from different academic disciplines and field and archival research and so might be more credible to stakeholders outside the government.

System exploration games may bring together different communities to synthesize understandings or understand key sources of divergence. However, cost, scale, and sensitivity pose barriers to mixing. A designer must consider which perspectives will produce information that is of most value to the sponsoring office. If the sponsor is new to her office, understanding from other parts of the bureaucracy might be valuable in bringing her up to speed quickly on the operations of that organization. A more experienced sponsor might derive greater benefit from voices outside the organization. Discussions with the sponsor about what game information would be valuable are a designer's best guide to making tradeoffs.

Player Engagement is Essential to Credible Results

Effective engagement requires that game participants perceive value in their participation. In the words of one longtime RAND game designer:

The utility of a game is critically dependent on the knowledge and experience brought to it by the players. This means that the designers/conductors of such games must focus on issues of interest to the participants and conduct game operations in a way that efficiently and productively uses the time they commit.³⁰⁷

System exploration games need to make the value of participation evident, so that they are willing to contribute their expertise. The more the different understandings of players are accounted for, the more the game can be tuned for productive dialog and deeper insights. In many cases, it may be helpful to interview potential players or staff from their organizations as part of pre-game research to ascertain their interest and concerns – and convey that these considerations are acknowledged by the design team.³⁰⁸ An alternative is to use players invested in the problem because they are, themselves, members of the research team as in Project Sierra.³⁰⁹ Other game designers argue that strong narrative elements of the game are key to hooking player attention.³¹⁰ Whichever tools are used, systems exploration games require thought to ensure player buy-in.

Common Limitations of Players in System Exploration Games

It must be recognized that there are important gaps between the experience of the player and the institutional role they are intended to represent no matter their expertise. Project Sierra game

³⁰⁷ Jones, "On Free-Form Gaming." p iii

³⁰⁸ Interview with Margaret McCown, Washington, D.C., July 2018.

³⁰⁹ Weiner. "War Gaming Methodology: Sierra near East Series." p 2

³¹⁰ McGrady, "Getting the Story Right About Wargaming."

analysts noted that, almost by definition, individual experience of players was narrower than their role both in terms of the multiple echelons they controlled in the game and the range of substantive issues discussed as a team that would normally be limited to siloed discussions among specialists.³¹¹ Such artificialities can be minimized with careful structuring of the actors involved, player selection (and particularly the use of multiple players with different experience), and role guidance. But the simplification of reality to make a game tractable makes them unlikely to be fully eliminated. Most games ask individuals or small teams to take on the role of institutions that are populated by thousands; some distortion is inevitable. While this is true of all games, because of the critical role of player inputs in system exploration games, the difference between player experience and the role they are assigned can call the credibility of game results into question.

Other practical problems arise from differentials in status. Bringing together players with unequal experience could see some defer to others or be ignored in group discussions due to less seniority. Teams could be set up to mirror these levels such as having a national-level strategic decisionmaking cell while more junior players manage planning for day-to-day military operations.³¹² Such an approach adds considerable complexity demanding more time, personnel, and thus cost to the project. Recruiting more experienced players may mean that they have worked across more levels of the organization, but that experience could well be dated and, as discussed earlier, senior players can be harder to recruit and retain. Put simply, a game design can make better and worse choices, but cannot claim to represent the whole of the governmental systems that are in play—simplifications will always be made³¹³ and must be accounted for in results.

Finally, the motivations of players, and the effects that these can have on game play, are also important to consider. Players with strong agendas and limited understanding of game design may try to change the game in stride, potentially skewing results.³¹⁴ In other cases, the same competitive instinct that makes games engaging can corrupt their usefulness. If players become more attached to “winning” the game at all costs, they may be more inclined to take advantage of artificialities of the game structure. While players can derail any game, system exploration games, with their greater flexibility for players to alter the environment and rules and potentially less ability for game staff to recognize distortions, are particularly vulnerable. As a result, a degree of trust in players by the designer is a consideration for the recruitment process.

³¹¹ Weiner, "War Gaming Methodology." pp 30 and 32-33, "War Gaming Methodology: Sierra near East Series." pp 24-25

³¹² Tucker Hughes and Josh Jones, "The Parts and the Whole: Linking Operational and Strategic Wargaming," *Phalanx* 51, no. 2 (2018).

³¹³ Jones, "On Free-Form Gaming." p 2

³¹⁴ Downes-Martin, "Your Boss, Players and Sponsor: The Three Witches of War Gaming." pp 34-35

Design Tradeoffs Related to the Game Rules

Much like the environment, the rules of a systems exploration game are often somewhat unstructured. If a designer is still trying to understand the problem, it is not likely that they can pre-identify all potential actions and their likely effects in enough detail to generate comprehensive rigid rules. As a result, free-form games,³¹⁵ matrix games,³¹⁶ and seminar-style techniques that allow a great deal of flexible interaction between different players, as well as players and adjudicators, are the norm for such games. More structured approaches can be successful as long as the game control team has a plan for revising game rules during play.

This openness, if not carefully guided by an attentive analytic eye, could mean that the information produced by a systems exploration game will be unfocused and lack utility. Designers benefit from being able to clearly articulate what information they need to be able to observe and document, and then designing interactions that insure these moments can be recorded. Put differently, in system exploration games: “playing teams are not rigidly constrained in addressing the problems presented—or in the form in which their recurrent moves are formulated—but the input requirements of any analytical model of fixed procedure for the analysis of interactions.”³¹⁷ The designer must then create rules that enable freedom for the players to consider different actions, including those not anticipated by the design team, while ensuring a plan is in place to capture the key insights in a structured way that can inform decisionmakers.

Use the Best Available Evidence to Build Initial Rules

Design teams should also look for opportunities to leverage more structured approaches to adjudication when there is empirical evidence available. Project Sierra game reports urged the use of pre-calculated values for factors such as fire power when data available from existing military texts, offered a “high confidence factor.”³¹⁸ Mathematical models were also sometimes used to work out the effectiveness of operations in-stride—in some cases using the computers

³¹⁵ Jones, “On Free-Form Gaming.”

³¹⁶ Matrix games refer to an approach to adjudication that depends on umpired arguments between competing actors. While several variations of the approach exist, a common form consists of an actor stating an action, the desired effect, and a rationale for why that effect is likely to occur. Actors who oppose the action are then free to offer an alternative narrative about the effects of the action and justification for their belief. An umpire then determines the outcome. A more detailed description of the approach and several examples can be found in: John Curry and Tim Price, *Matrix Games for Modern Wargaming: Developments in Professional and Educational Wargames* (History of Wargaming Project, 2014).

³¹⁷ Jones, “On Free-Form Gaming.” p iii

³¹⁸ John P. Evans. “Guide for Ground Force Adjudication in War Games.” (Santa Monica, CA: RAND Corporation, D-4765, 1957). p 3

that were becoming increasingly common in other areas of RAND work.³¹⁹ In contrast, emerging issues and planning factors that emphasized human judgement based on experience were adjudicated based on the judgement of umpires, in conjunction with player expertise.³²⁰

The Gray Zone game used an alternative approach of using probability tables prepared prior to game play for generating the results of actions. These rules were visible to players and the design team was willing to adjust the likelihood of outcomes based on player feedback.³²¹ This approach allowed debate to focus on rules about which players disagreed with the game design team rather than needing all decisions to be debated.

A third alternative sees decisions about which actions to allow made in stride. Here, the risk is that quick decisions by control can deviate from paths that best support research to those that seem interesting or convenient in the moment. It is helpful for the control team to have predetermined heuristics to judge which of several possible outcomes is selected. Some Project Sierra games were run under a set of rules designed to help identify key “decision-points” where multiple courses of action were available that substantially shaped the course of the conflict. In these games, researchers used the guides of relevancy (that is, selecting options that seemed more closely tied to the purpose of the game) and a desire to select options that generated additional game play to guide decision making.³²²

However, this third style of game substantially increases the work of the control team, since these decisions involve a great deal of judgement. The Sierra team found that these games demanded more time and a more experienced staff than games in which player choice was more constrained.³²³ In part, this is because players faced with uncertainty will attempt to extract more information from control in their interactions.³²⁴ Furthermore, it is generally necessary to leave some degree of decision-making authority in the hands of the game staff to ensure that the game stays within the parameters of the research question.

Adjudication Should Focus on Transparency

The tendency to use non-rigid approaches to adjudication creates an important burden on game analysts to develop data capture process to help document what rules were created by players and adjudicators in-stride, since they are such an important part of the model built during

³¹⁹ E. W. Paxson. "The Sierra Project -- a Study of Limited Wars." (Santa Monica, CA: RAND Corporation, B-41 (WITHDRAWN), 1958). p 11

³²⁰ Milton G. Weiner. "War Gaming: Two Methods Used in Sierra." (Santa Monica, CA: RAND Corporation, D-4332-PR, 1957). p 9

³²¹ Wasser et al., "Gaming Grey Zone Tactics." pp 31-32

³²² Weiner, "War Gaming Methodology." pp 60-63

³²³ Ibid. pp 22-24

³²⁴ "War Gaming Methodology: Sierra near East Series." p 39

the game.³²⁵ In the Project Sierra games, player decisions were collected in a fairly standard structure and formalized logs were kept of the status of military assets such as air fields, fuel, and ammunition, designed to ensure that necessary operational details were available to the control team during and after the game. However, records that tended to take a more narrative form were also kept for more subjective information, such as the intended effect of operations.³²⁶ The structure of the Gray Zone game was also designed to make player choices easy to observe and capture for analysis—game reporting could easily document what action cards teams played as well as those they considered but ultimately opted not to use.³²⁷ These practices help the research teams to understand not just what had occurred but why, aiding in-stride adjudication and post-game synthesis of findings to improve practices prior to the next game of the series. Without such practices, too often seminar-style approaches are left largely unanalyzed, rather than using the opportunity of the game to elicit implicit understandings and assumptions about who can (and cannot) produce what influences. While limited capture can be convenient in the moment, such lacunae in data collection ultimately limit the ability of the research team to fully articulate the model of the problem generated by game play.

Communicating rules effectively, particularly when there are complex interactions that are difficult to model satisfactorily, is also important. Figure 5.4. from the Sierra game series illustrates how even relatively straightforward operational missions such as an airstrike could require consideration of many planning factors by the control team, including issues pertaining to several different areas of expertise, and possibly requiring the use of both rigid and judgement-formed outcomes.³²⁸ While the number and nature of planning factors will vary considerably depending on the purpose of the game, often they are sufficient to overwhelm unaided human memory. Using mapping tools like this diagram can help the control team work together more effectively, since it allows members to understand both what types of decisions they may be called on to make and how it is likely to affect the work of other members of the team.

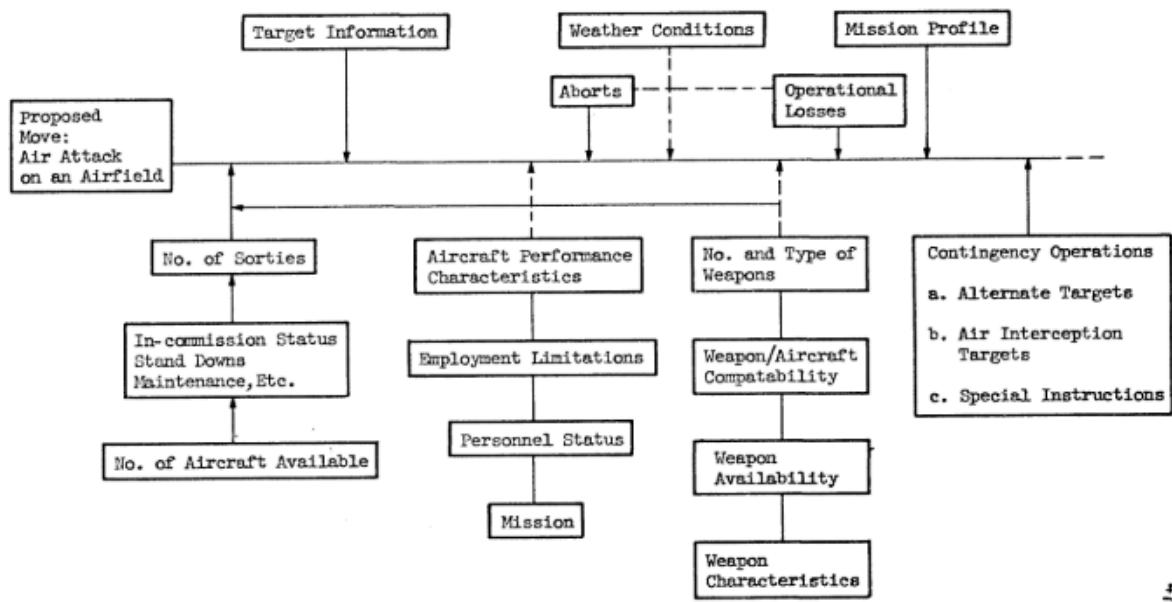
³²⁵ Of course, there are real limits in documentation practice, particularly when it comes to documenting player beliefs and mental processes. While recognizing these limits, social scientific research provides a wide range of tools for observation, documentation, and thoughtful analysis that attempts to account for these concerns. Games benefit to the extent they can credibly document decisionmaking processes and choices.

³²⁶ Weiner, "War Gaming Methodology." pp 37-38

³²⁷ Interview with Becca Wasser, Washington DC, January 2020.

³²⁸ Weiner, "War Gaming Methodology." pp 45-48

Figure 5.4: Factors in air strike adjudication



Source: Weiner, "War Gaming Methodology." p 47

Managing Time

Generally, system exploration games treat player planning sessions as if time has stopped and all teams make decisions simultaneously.³²⁹ This is largely a matter of convenience for the players and adjudicators. While there are gaming tools to allow real-time, no-turn play,³³⁰ they are often chaotic for the control team to manage given the wide aperture for player action, limiting the ability to capture systematically the model generated by players in the game. Thus, these designs can interfere with the analytical purpose of systems exploration games. Conversely, I-go/you-go games in which teams alternate turns generates considerable time in which players are not active. This can risk reducing player engagement, a key concern for system exploration games. Of course, that is not to say that these other approaches are never appropriate, simply that they create tradeoff that can be in tension with the objectives of system exploration, and thus are less frequently selected.

Another area of adjudication that requires particular care in system exploration games is the management of game pace in relation to clock time. Often the outcomes of actions in different domains require different times to manifest and have a different normal rhythm. For example, moderate air operations are often tied to a standard 72-hour Air Tasking Order (ATO) cycle,

³²⁹ Jones, "On Free-Form Gaming." p 8

³³⁰ Wong et al., "Next Generation Wargaming for the U.S. Marine Corps: Recommended Courses of Action." p 31

making 12-, 24-, and 72-hour cycles natural periods for game play—however such timing may not be a good fit with naval or ground operations. This disconnect is a realistic feature of the coordination of joint operations, but for game purposes it can be hard to select a timeframe without the appearance of “preferring” one domain’s time-scale over another. The problem is compounded when looking at issues beyond military conflict where longer and more varied time frames dominate. Gray Zone game designers wrestled with representing gray zone tactics designed to be long-term investments along with others manifest on very short order. Building both a long- and short-term track allowed the game to recognize two different time steps and the interplay between them without adding a great deal of complexity.³³¹

Challenges of Hidden or Incomplete Information

A final trade-off is the extent to which players representing different actors are aware of information regarding the objectives, capabilities, and actions of competitors. Providing all information to all players (sometimes referred to as an “open” game) can be attractive in system exploration games, because it allows more players to contribute expertise to refining the depiction of the environment and its meaning for game play. However, complete knowledge about the goals and tools available is highly unrealistic and limits the opportunities to study issues like signaling, perception, and communications, frequently of great interest to research teams.

Some Project Sierra games were played with open information to allow players to focus on “tactical operations and the military outcomes of the war based on the capabilities assumed for each side.”³³² Researchers noted that this style of game was useful for identifying factors, particularly related to military capabilities, that drove outcomes, and that visibility by both sides provided a useful check on assumptions built into the game context. However, these advantages came at the cost of contingency planning or consideration of the interactions between political and military decision-making.³³³ This approach proved useful to explore variations of play from previous games (which players were frequently involved with, and thus already had knowledge about what choices were made), and thus could afford to narrowly focus planning within the boundaries of observed game play in the interest of speed.³³⁴ However, in games first exploring a problem, “closed” information was seen as important for achieving game objectives, and thus more information was kept hidden from opposing player. The Gray Zone game’s greater stylization made this problem somewhat easier for the game team, since action cards were all played upside down, and teams opted which actions to brief as part of their plan and which to

³³¹ Wasser et al., “Gaming Grey Zone Tactics.” Pp 18=19

³³² Weiner, “War Gaming Methodology.” p 20

³³³ Ibid. pp 20-21

³³⁴ “War Gaming: Two Methods Used in Sierra.” p 10

treat as covert, the players had more of the responsibility for managing the revelation of information.³³⁵

Conclusions

System exploration games produce information about a policy problem by eliciting and synthesizing the views of players about the structure of the system. Depending on the topic of the game, the level of detail and extent to which players are asked to project into the future may vary, however, generally the goal is a useful model of the problem to help focus future areas of study. The importance of expert mental models to the desired information from the game means that the game's actors, and more specifically the players used to represent them, take on outsized importance in system exploration games. The game's environment and rules are generally shaped by the need to incorporate new information from players. This creates a requirement for relatively flexible game elements which can capture player inputs, as well as a system to capture and communicate any changes made in response to player inputs.

³³⁵ Wasser et al., "Gaming Grey Zone Tactics." pp 36-37

Chapter 6: Designing Games for Alternative Conditions

Chapter 4 presented the second archetype of alternative conditions—that is, games designed to produce information that helps researchers, sponsors, and consumers better understand the nature of problems by highlighting the impact of alternative conditions on decision making. Since these games require a decision about what alternative conditions are likely to change decisions in interesting ways, they tend to come after initial research framing the policy problem; perhaps through a systems exploration game, perhaps through another form of research. Since they produce a more refined understanding of the problem, these games represent mature research products that can be shared with outside stakeholders to influence decisionmakers. Alternative conditions games are likely to appeal to positivists, since the fundamental logic of controlled comparison that underpins the approach aligns well with positivist philosophical claims. Analysts and critical realists are likely to see such an approach as fussy and best and generalizing with limited grounds at worst, and so are unlikely to advocate for this style of game design.

The design of alternative conditions games rests fundamentally on the need to create comparisons that allow the game analysis to highlight similarities and differences between games which are conducted under different conditions. Specifically, in order to produce information about how different conditions impact decisionmaking processes and the choices that result from them, these games are run multiple times with different factors worked into the design to collect data about similarities and differences between player debate and decisions.³³⁶ The strength of findings will depend to a large degree on the extent to which the analyst can argue that variations beyond that of the key factor does not offer an alternative explanation for any variation that occurs in the outcome. Put differently, analysts need to demonstrate that variation in other potential factors has been held constant or *controlled* by the research design. Most positivists hold that such analysis can be helpful in illustrating the mechanism or process that connects cause and effect, but that the credibility of such claims rests on careful analysis to eliminate potential alternative explanations and careful thought about the conditions that will be similar enough that we can expect to see the same process unfold. Both of these issues are somewhat complicated by the nature of games.

³³⁶ Because games provide the flexibility for designers to make many choices about the set-up of the game but have limited ability to guarantee variation in player decision, alternative condition games often follow the pattern of a “most similar” comparative design. In a most-similar design, cases are selected to be as similar as possible except for variation in causal input of interest which is then used to explain any differences in the outcome variable. For a detailed description of this research design from the case study literatures, see: George and Bennet, *Case Studies and Theory Development in the Social Sciences.*, p 81

The synthetic environment of the game gives a researcher a great deal of influence over the initial conditions of the game compared to observation of the real world. Designers are free to develop convenient alternative futures, highlighting some actors while abstracting other, and to develop formally documented processes to shape player behavior, even while considering national level security concerns that are generally beyond the ability of individuals, let alone researchers, to affect. In this way, games can be seen as a “laboratory” to conduct experimental research that would not be possible in the real world. However, researchers on gaming have long cautioned that the nature of games makes this metaphor imperfect for two reasons. First, the very artificiality of the game raises questions about how well we can generalize results of the game to real world conflicts. Second, the agency of the human players during the game creates a major source of uncontrolled variation which prevents games from ever being repeated or replicated.³³⁷

Instead, the more apt analogy for the design of these comparative games seems to be approaches to structured comparison from the case study literature.³³⁸ Rather than resting on replication, case studies seek to trace patterns in rich data over time, generating a deeper understanding from 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 (sometimes referred to as within-case variation) and create clear comparisons between multiple cases (known as cross-case or between-case variation). These patterns of similarity and difference then form the evidence to argue for an explanation of the causal relationship or mechanism for how a factor is impacting an outcome of interest.³³⁹

Applying this competitive model to games does not remove the twin concerns of artificiality and player non-comparability, but the case study literature does point to an approach to managing the impact of these realities on the credibility of findings. First, on the issue of translating findings from the synthetic world of the game to the real world, researchers should be prepared to discuss ways in which the representation of the game world may have deviated, and those deviations’ potential effects on findings. 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. An argument may be made that the synthetic environment may make individuals less risk adverse because they

³³⁷ For a particularly clear articulation of the shortfalls of games as experiments, see: Parson, “What Can You Learn from a Game?,” pp 237-240.

³³⁸ Excellent sources on case study research design, see: George and Bennet, *Case Studies and Theory Development in the Social Sciences.*, Lange, *Comparative-Historical Methods.*; and Beach and Pedersen, *Causal Case Study Methods: Foundations and Guidelines for Comparing, Matching, and Tracing.*

³³⁹ 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.

recognize costs are artificial, enabling more aggressive behavior. This type of non-representative behavior will pose more problems for findings dependent on aggressive play than findings that leaders look for off ramps. Similarly, the potential problems of non-comparable players and group dynamics should be approached head on and is discussed in some detail in this chapter's discussion of actors in alternative conditions games. In short, attempts to anticipate and address potential shortcomings clearly will do much to mitigate these risks.

Given these considerations, generally, design of an alternative conditions game starts with the identification of the key variable(s) of interest to be varied between games. The environment of the game might be varied 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 in an instigating event) or broad (selecting different environments in which a crisis over water could occur).³⁴⁰ Similarly, the actors may be varied either by changing the identity, objectives, and resources available in the starting conditions, or through careful selection of the players representing each role. An important example is changing the capabilities available to different forces at the start of the game. Varying the actions available to players supports investigation of the effect of processes (such as communication protocols or deliberative protocols). Regardless of where the key variation is located, the goal of the designer is to vary that factor, and only that factor, between plays of the game.

Specific considerations for the depiction of the environment, actors, and rules follow below. These tradeoffs are illustrated with two example games. The first is a RAND game from the 1960s that studied the impact of different strategic and budgetary conditions on force structure, posture, and use. The second is a recent RAND effort designed by the author that considered how the analytic inputs to force structure decisions shaped player choices. I use both games as illustrative examples of how designers have navigated the tradeoffs and potential pitfalls of designing games to study decisionmaking under alternative conditions.

Overview of Example Alternative Condition Games

The use of comparison to illuminate the impact of a specific factor on player decision making is hardly a new type of game but based on available records and committee consensus they are less common than their system exploration counterparts. In part, this is likely because the approach requires executing multiple games, and thus are more expensive to conduct than approaches that can use only a single game to good effect. It is also possible that the dominance of a rigid form of positivism in defense analysis (discussed in detail in Chapter 3) has made this approach somewhat less common. Since alternative condition games do not align well with other philosophical traditions, these games are not likely to be used by practitioners working with those perspectives. At the same time, positivists coming from a more engineering approach,

³⁴⁰ Bartels, McCown, and Wilkie, "Designing Peace and Conflict Exercises: Level of Analysis, Scenario, and Role Specification."

rather than from the social sciences, are less likely to be comfortable with the types of qualifiers needed to overcome the weakness of games as data generating tools. It is perhaps unsurprising that many of the examples of this style of game design come from designers with social science backgrounds.³⁴¹

The two examples detailed below look at a common problem—decisions about how to best invest in future forces. However, the two games look at the causal impact of different contextual factors. SAFE considers the impact of different U.S. and Soviet strategies and budget levels. The later force structure game instead looks at how presenting different analysis of the strengths and weaknesses of different force structure elements shapes decisionmaking. As a result of these differences, the two research teams took somewhat different approaches to designing the environment, actors, and rules of the game, while sharing a common concern that the different iterations of the game generate credible companions.

US Air Force/RAND Strategy and Force Evaluation (SAFE) Games

In 1962, RAND ran six iterations of a manual force posture planning game that looked at ten years of strategic force structure investment and posturing decisions. The goal of the games was to “explore the extent to which alternative sets of strategic objectives would lead to distinguishable general war force postures”³⁴²—in other words, did different statements about the overarching strategy of the U.S. or Soviet Union, and different budgets, actually cash out into different choices about what forces were bought, where they were postured, or how they were used? In addition to studying variation within a single country, the research team was interested in the interaction between different U.S. and Soviet strategies and budgets.³⁴³ Figure 6.1 shows the six combinations of strategy and budget that were selected for play. By allowing both blue and red to make repeated decisions about strategic investments based on limited information about the choices of the other side, game play provided a realistic middle ground between force posture analysis processes that assume either a static adversary inventory, or a reaction made with perfect knowledge of what selections had been made.³⁴⁴ As such, the game was an exploration of the impact of alternative strategies and budget limits on U.S. and Soviet decision about force structure, posture, and use of their strategic arsenals.

³⁴¹ For some other example that draw on social science comparative research design traditions, see: *ibid.*; Hank J. Brightman and Melissa K. Dewey, "Trends in Modern War Gaming," *Naval War College Review* 67, no. 1 (2014).; Jacquelyn G. Schneider, "Cyber Attacks on Critical Infrastructure: Insights from War Gaming," *War on the Rocks*, July 26 2017.; and Erik Lin-Greenberg, "(War)Game of Drones: Remote Warfighting Technology and Escalation Control Evidence from Wargames." SSRN, (2019).

³⁴² Thomas A. Brown and Edwin W. Paxson, "A Retrospective Look at Some Strategy and Force Evaluation Games," (Santa Monica, CA: RAND Corporation, R-1619, 1975). p 2

³⁴³ Harvey A. Averch and Sorrel Wildhorn, "Risk, Ambiguity and Force Structure: An Analysis of General-War Forces and Strategic Objectives in Cases C and D (of Six Case Studies)," (Santa Monica: RAND Corporation, RM-3511-PR, 1963). p iii.

³⁴⁴ Brown and Paxson, "A Retrospective Look at Some Strategy and Force Evaluation Games." p 47

Figure 6.1: Strategies represented in plays of the SAFE games

RED (SU) Policy		Emphasize Defense, Given Easy Deterrence of BLUE	Build a Reliable Deterrent; Achieve Force Parity with BLUE	Build a Reliable Deterrent Plus a Disarming Capability	Achieve Force Superiority over BLUE
		RED Budget ^a	BLUE Budget ^b		
BLUE (U.S.) Policy		1.5	Low ^c	1.5	3.0
Peacetime Deterrence ^d plus Wartime Deterrence	Low ^c		D		C
Peacetime Deterrence plus Disarming Capability	3.0	E			
Peacetime Deterrence plus Wartime Deterrence plus Disarming Capability	1.0	F			
	1.5	A		B	

^a10-year total game budget relative to ten times the estimated annual "strategic" budget for the Soviet Union, averaged over FY-1961-1962.

^b10-year total game budget relative to ten times the corresponding DoD budget sector for FY-1963.

^cDetermined in the course of play.

^dAs used in this Memorandum, the term deterrence connotes specifically a retaliatory capability, not necessarily limited to counter-population.

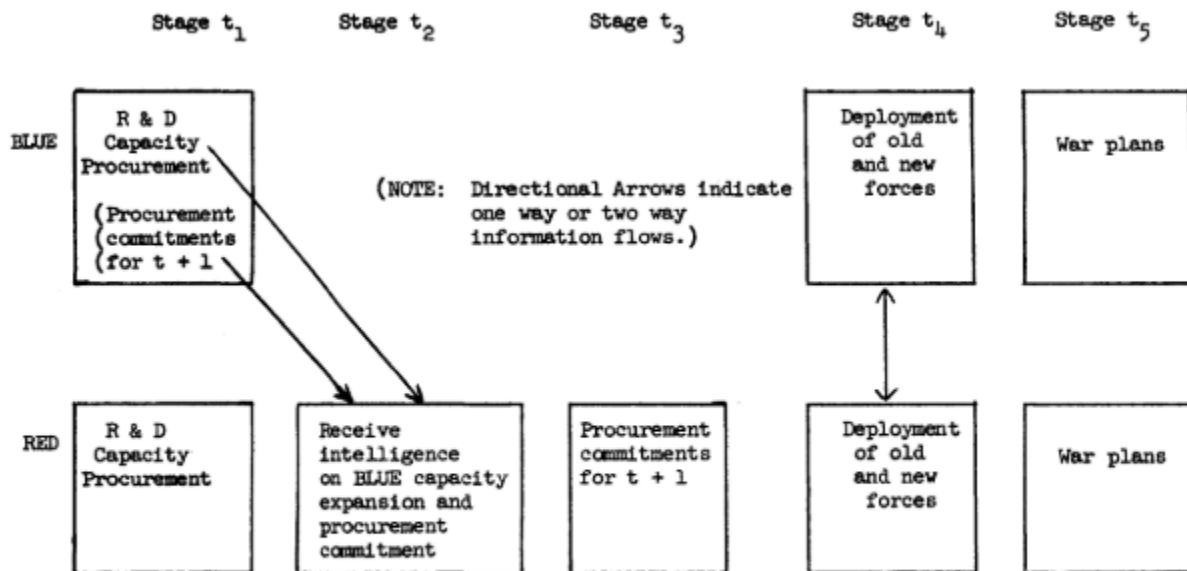
Source: Harvey A. Averch and Sorrel Wildhorn, "Risk, Ambiguity and Force Structure: An Analysis of General-War Forces and Strategic Objectives in Cases C and D (of Six Case Studies)," (Santa Monica: RAND Corporation, RM-3511-PR, 1963). p v.

The core of SAFE was a series of investment rounds, in which the blue and red team each received strategic guidance, a budget, and a portfolio of capabilities they could consider for investment over a 10-year period. They also received an initial force posture, and intelligence on the other team's previous actions. Play progressed in five moves, each representing two years of investment, in which players split available funds between R&D, procurement costs, and operations and maintenance costs for various strategic systems. Players also posture forces on a simplified map organized by "zone."³⁴⁵ This enabled careful tracking of the inventory of each side, which could also be systematically compared to the results of other runs to highlight

³⁴⁵ Olaf Helmer-Hirschberg and Robert E. Bickner, *How to Play Safe--Book of Rules of the Strategy and Force Evaluation Game* (Santa Monica, CA: RAND Corporation, 1961).

similarities and differences in player choices.³⁴⁶ Players were also asked to assemble “war plans” for a specific contingency each turn that could be compared to one other by the control team to establish a basic understanding of the posture’s capabilities, and thus the risk, however this activity represented a small portion of the total exercise time.³⁴⁷ The outcomes of this move focused on the strategic damages that could be inflicted by the selected force, measured in total casualties for easy comparison between sides and cases.³⁴⁸ This process is illustrated in Figure 6.2.

Figure 6.2: Process of SAFE game play



Source: Harvey A. Averch and Sorrel Wildhorn, "Risk, Ambiguity and Force Structure: An Analysis of General-War Forces and Strategic Objectives in Cases C and D (of Six Case Studies)," (Santa Monica: RAND Corporation, RM-3511-PR, 1963). p 30.

RAND Force Structure Decision Analysis Game

Research institutions are frequently generating new tools that researchers hope will enable leaders to make better decisions, but they rarely have evidence beyond their convictions that the tool will actually positively change behavior compared to traditional approaches. To gather evidence about how robust decisionmaking analysis³⁴⁹ might lead to different force structure

³⁴⁶ J. A. Davis and S. C. Silivski. "Estimated Total Obligational Authority of the Force Structures Generated by the Safe/Acws Games." (Santa Monica, CA: RAND Corporation, D-10948-PR, 1963).

³⁴⁷ Brown and Paxson, "A Retrospective Look at Some Strategy and Force Evaluation Games." p 1-2.

³⁴⁸ Averch and Wildhorn, "Risk, Ambiguity and Force Structure: An Analysis of General-War Forces and Strategic Objectives in Cases C and D (of Six Case Studies)." p vii

³⁴⁹ Robust decision making (RDM) is a decision support tool developed at RAND. The goal is to separate identification of decision that perform well under a wide range of conditions (that is, that is *robust*) rather than over-

decisions and decision processes than traditional scenario analysis, a team of RAND analysts (including the author) designed a game that would simulate a senior DoD decisionmaking body in 2017. This game would provide an opportunity for researchers to closely observe how analysis was leveraged by a group of stakeholders as they debated competing force structure options and came to a group decision.³⁵⁰

Game play featured a single team, in which each player was asked to roleplay a senior DoD decisionmaker who typically sits on the Deputy's Management Action Group (DMAG), including the civilian service secretaries, key leaders from the Office of the Secretary of Defense, and chiefs of the military staffs. The group of players was presented with a decision brief presenting previously completed analysis of the strengths and weaknesses of three different force structure options. Players were asked to debate the merits of the options and then offer a recommended option to the chair of the group (played by a member of the control team who acted as facilitator of the discussion). In the second move, players were confronted with a scenario that was designed to be stressing to the selected force structure—that is, regardless of what decision players made, they were given a future in which they had made the “wrong” choice. Players then discussed how they assessed their decision, and how they might have made a better choice to understand how they processed the analysis in light of an “incorrect” decision.

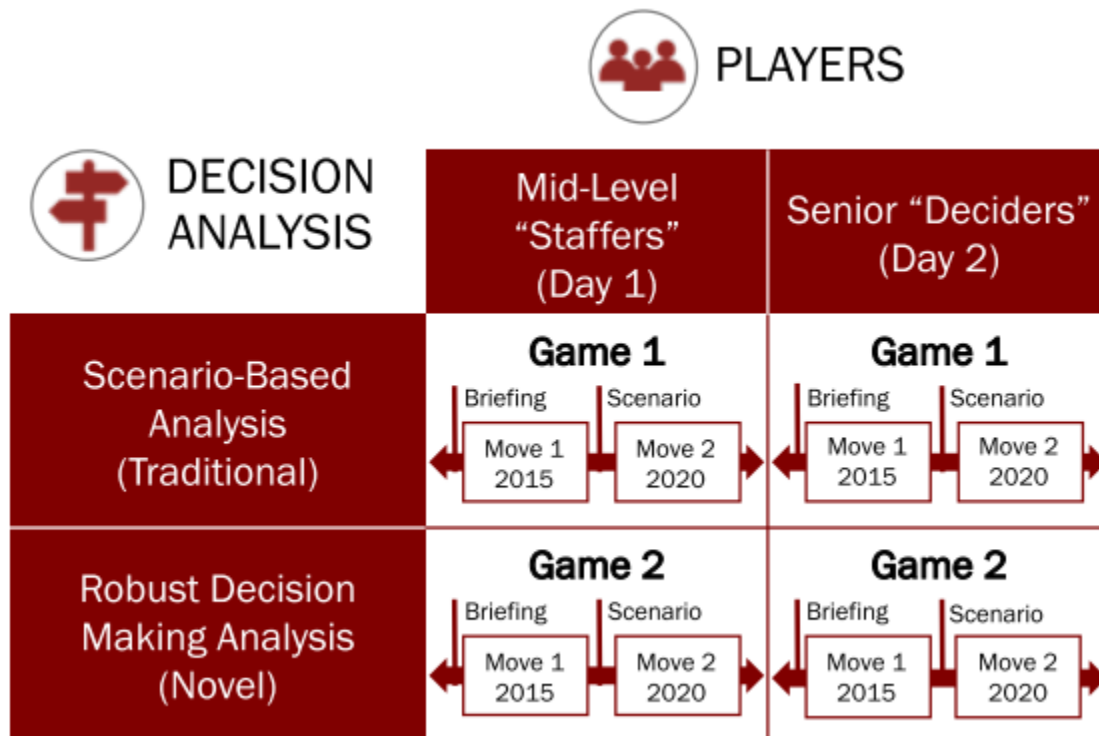
To explore the impact of two different analyses we repeated this game twice with the same group of participants—first with traditional scenario planning analysis, and second with the novel RDM force structure analysis. As is discussed in more detail below, the research team was concerned that differences in player experience might shape their responses to decision analysis tools, so we ran this process twice, once with mid-level players and once with more experienced individuals, resulting in a total of four games. The setup of these four games is shown in Figure 6.3.³⁵¹

optimizing on a very small number of selected scenarios. See: Robert J. Lempert et al., "Making Good Decisions without Predictions: Robust Decision Making for Planning under Deep Uncertainty," (Santa Monica, CA: RAND Corporation, RB-9701, 2013).

³⁵⁰ Bartels et al., "Do Differing Analyses Change the Decision?: Using a Game to Assess Whether Differing Analytic Approaches Improve Decisionmaking." pp 1-2.

³⁵¹ Ibid. pp 7-10

Figure 6.3: Comparative format for the Force Structure Analysis game



Source: Elizabeth M. Bartels et al., "Do Differing Analyses Change the Decision?: Using a Game to Assess Whether Differing Analytic Approaches Improve Decisionmaking," (Santa Monica, CA: RAND Corporation, 2019) p 8.

Design Tradeoffs Related to the Game Environment

The environment of alternative conditions games is critical for establishing a strong foundation for comparison. Since the designer has a great deal of latitude in how they opt to depict the policy problem, much can be done to ensure that the environment serves as a consistent context for the variation of interest to play out. However, two issues often challenge designers. First is the need to balance explicit world building—which has the benefit of transparent accountability that factors were consistent between runs of the game—with the need to ensure game time forces player debate and decisions, rather than introducing players to a detailed alternative world. Second, often the variation of interest will produce second order effects that realistically should change the game environment. Deciding how many of these changes must be manifested to be credible, while ensuring that they do not constitute an alternative explanation for any differences in decisionmaking, can be a tricky endeavor.

Challenges of Explicit World Building

One tradeoff designers face is between how much of the game environment to make explicit to players. Unlike in system exploration games, where the primary purpose of the environment is

to ensure all players are focused on the policy problem of interest while leaving a great deal of space for players to inject their own understanding of the problem, alternative conditions games usually have somewhat more explicitly specified environments. Without more explicit guidance, there is a strong possibility that different players will have different understanding or make different assumptions, creating problems for comparison. As a result, often more is done to explicitly define the environment of an alternative conditions game. The Decision Analysis game provided consistent guidance to all players about the “decision context” that included information about the budget, strategic priorities, and perceived global threats modeled on the types of read ahead materials that might be available to real world decision makers.³⁵² However, even with this explicit guidance, players still made different assumptions about the global environment than were intended by the design team, highlight the limits of such an approach.³⁵³ In part, this shortcoming highlights a common concern—balancing the desire for an explicit depiction with the need to make the most of valuable player time, and thus minimize the time spent on describing the world to players.

Challenges of Second Order Effects and Change over Time

Design decisions about how the environment is represented also require carefully thinking through 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 rules will necessitate making other changes to ensure narrative plausibility. One must consider both what variation is demanded by plausibility and what variation threatens to complicate the comparative narrative. If the additional variation is required for coherence could provide an alternative explanation, this should be addressed explicitly in analysis.

Other times, the need to support comparison forces the design team to minimize realist change over time. In the SAFE games, researchers noted that a major artificiality of game play was that the strategic guidance provided to players was not changed over the course of the ten-year period of play. Comparison to real world policy shifts over the actual ten-year period simulated in game play, which featured substantial shifts to both US and Soviet policy, illustrate how unrealistic this assumption was from the perspective of predicting the future.³⁵⁴ However, from the perspective of generating useful comparisons of the impact of the alternative strategic environments, maintaining consistent guidance over the period of game play was incredibly helpful. Because the game was originally designed to generate information about whether different strategies would produce different force postures, a game in which the guidance was changed at some point would have created a requirement to compare postures within a single

³⁵² Ibid. pp 32-35.

³⁵³ Ibid. p 23.

³⁵⁴ Brown and Paxson, "A Retrospective Look at Some Strategy and Force Evaluation Games." p 35.

run. Designers worried that these considerations would make game results more difficult to interpret, limiting the usefulness of analysis. As a result, they opted to create a more artificial, but more controlled environment to enhance the comparative analysis, at some cost to generating credible realism.

Design Tradeoffs Related to the Game Actors

Regardless of what factor is the basis for the alternative conditions, actors are a particular concern in alternative condition games because the variation among human players and the interactions between them make full controls between game impossible. Player learn from play, and their moods and relationships change, as often as not for reasons that have nothing to do with the game, and thus are outside the control of the designer. What is more, whereas the game designer can take substantial steps to make the game environment and rules transparent, enabling clear comparison by game analysts and consumers of the game report, the difficulties of fully understanding human decisions compromise transparency. As a result, the designer can mitigate variation generated by players, but cannot credibly claim to have removed it. Common strategies are either to have the same group of players play several games or to attempt to recruit different, demographically comparable groups of participants—both approaches are imperfect.³⁵⁵ Beyond the issues created by players, there are also some broader concerns about how the actors who's decisions drive the game are structured that also bares careful consideration by the designer of an alternative conditions game.

Challenges of Players

If the designer ops to use the same group of players in multiple games, it is necessary to account for learning between the rounds of play, as well as other changes that may change player disposition, attitudes, and relationships. If the same group of players is presented with the same decision context twice, we would expect that their subsequent decision would in part be influenced by the discussions, decisions, and outcomes of first round of play. In the RAND Force Structure Decision Analysis game, we were careful consider the order in which players saw to two types of analysis. We opted to give players the status quo analysis first, since that would most closely align to familiar models. The new analysis was used in the second round, so that if it did (as hoped!) materially change the way players thought about the problem, it would not bias the comparison.³⁵⁶ This design was appropriate in a context where one treatment represented the status quo approach and the other was novel. However, for other types of comparisons, different experimental designs might be more appropriate.

³⁵⁵ Jones, "On Free-Form Gaming." pp 2-3

³⁵⁶ Bartels et al., "Do Differing Analyses Change the Decision?: Using a Game to Assess Whether Differing Analytic Approaches Improve Decisionmaking." p

There is also the challenge of maintaining engagement and interest in multiple rounds of truly identical play. Doing so often requires that additional variations be introduced. Again, turning to the Decision Analysis game as an example, we prepared two structurally equivalent scenarios, so that even if the players picked the same force structure option, they could be equally stressed without repetition.³⁵⁷ In this case, because the previous RDM analysis had clearly defined the characteristics of a scenario that would be stressing to any particular force structure, it was possible to generate credibly equivalent scenario. Lacking such a pre-defined structure, this effort would have required substantial additional work to accomplish.

The second option is to try to recruit multiple groups of reasonably comparable players. This can be practically attractive, because it allows games to be run in parallel at the same time. However, for the comparison to be credible, the designer must be able to define what characteristics of the players are salient to decisionmaking and defend why the two groups are similar on these dimensions. This can be quite challenging, given the range of experience that may shape decisionmaking, and the limited knowledge about player background that game designers may have available. Each of the SAFE iterations utilized players drawn from the associated project teams, and most players participated in at least 2-3 games.³⁵⁸ This gave designers an unusual level of knowledge of their players, their priors, and their relationships to help inform placement on a team. When players are recruited from outside the organization, as is far more common today, this level of insight will be difficult to achieve.

Regardless of which option is selected, it is unlikely that difference in player experience, attitudes, and beliefs will not play a role in shaping decisionmaking. 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. Tools such as process tracing or analysis of player discourse³⁵⁹ can be particularly powerful here. In the Decision Analysis game, we were concerned that different seniority might impact how players consumed information, so we opted to stratify players into two groups to enable us to conduct a secondary comparison of mid-level and senior players. As it turned out, player seniority had a larger effect on game discussion—senior players more faithfully mimicked bureaucratic behaviors, and thus the change in modes of analysis had less influence on their decisionmaking than more junior audiences.³⁶⁰ However, it may not always be possible to anticipate potential confounding effects, or to take this degree of mitigating action even when they are anticipated. Thus game data capture protocol should

³⁵⁷ Ibid. p 10

³⁵⁸ Brown and Paxson, "A Retrospective Look at Some Strategy and Force Evaluation Games." p 52

³⁵⁹ For a recent example of this technique, see: John Derosa and Lauren Kinney, "Narrative Analysis of Wargaming" (paper presented at the Connections Wargaming Conference US, Washington, DC, 2018 of Conference).

³⁶⁰ Bartels et al., "Do Differing Analyses Change the Decision?: Using a Game to Assess Whether Differing Analytic Approaches Improve Decisionmaking." pp 20-21

include capturing information that may help identify differences between player behavior across games and game analysis should transparently consider how these differences might drive game results.

Challenges of Representing Actors

Beyond the problem of finding a way to make players comparable, there is also the question of how to make the representation of actors comparable. Given the need for game play to represent decisions that actually occur in large organizations, games almost always simplify complexity—but this creates space for different players to take on different interpretations of the role they are being asked to assume. One option, which was used in the Decision Analysis game, is to provide players with specific “role guidance” laying out the specific bureaucratic equities that they are to represent.³⁶¹ This approach has the advantage of baselining players to at least some degree. However, such an approach may not be as effective in a game in which more than one individual represents a role, or in which roles are not differentiated.

In part, the problem is once again that a simpler representation of game actors makes controlled comparison easier, but also risks over-simplifying or otherwise distorting the nature of the problem. In the case of the Decision Analysis game this was less of a concern, because the membership of the DMAG is established, and could be prioritized without much controversy. On the other hand, SAFE faces the problem of treating only two of the many actors of the global security environment. Retrospective analysis noted that the need to manage two teams was due to the limitation of the small team of human adjudicators. They noted that the game’s realism would have been improved by relaxing this constraint, and allowing the representation of other actors, including Europe and China.³⁶² However, it seems likely that such a design choice would have been made at the expense of the ability to conduct clean comparisons of the strategic interactions between decisionmaking on both sides.

Design Tradeoffs Related to the Game Rules

Like the game environment, designers have a great deal of control over game rules to enforce consistency but must think through how these choices interact with desired variation, player differences, and other considerations that drive the credibility of game results. One key concern is the degree to which rules can be formalized—formalization tends to make for easier comparison. However, since players may still make decisions that are outside the bounds of the rules, game procedures need to be flexible and transparent enough to account for changes in game play. A similar point also holds for adjudication procedures—there is sometimes a tension

³⁶¹ Ibid. pp 8-9 and 28-31

³⁶² Brown and Paxson, "A Retrospective Look at Some Strategy and Force Evaluation Games." p 51.

between developing a more nuanced set of rules, and ease of comparability that designers must juggle.

Challenges of Formalized Rules

A parallel concern to the issue of how explicit to make the world building of the game environment is how much to depend on formally stipulated rules versus allowing players to take actions that might not have been anticipated by the game team. To take one extreme, if players can take any action that occurs to them and player perceptions about what actions are available may differ considerably between different groups of players there is a great risk of variation. As a first order assumption, it is safer to assume that allowing players free rein will generate games played under different rules than that players will share the same understanding of the problem. An exception to this is if the players are drawn from a population that has been engaged on the problem for an extended period. The SAFE game used players who were part of the research team,³⁶³ thus minimizing variation between players mental models to at least some extent. At the other extreme is to fully specify all game rules in advance. This has the advantage of ensuring more comparable play, but if the game is built on a flawed understanding of the problem such a choice may reduce the credibility of results for sponsors and consumers.

If players are also involved in adjudication (as for example, in a matrix game) this concern also extends to the projected outcomes of actions. Furthermore, since it is not always possible to fully tease out plays' decision processes, players may enforce limits that they never articulate, creating "hidden" game rules. If on the other hand, available actions and adjudication are highly formalized, it is much easier to ensure transparent consistency between rounds of play, but the game risks becoming either too abstract to be useful or too cumbersome to be manageable. As a result of these concerns, more structure approaches to adjudication tend to be preferred.

The choice of how structured a game system to use is often dictated by the existing knowledge about the problem. Better understood problems generally have more substantial conceptual models available to form the basis for rule development, while emerging problems may have uncertain or contested models. In this later class of cases, 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. For this reason, alternative conditions games conducted very early in the research process are not likely to produce very strong findings.

Challenges to Comparison from Player Decisionmaking

Beyond the formalization of the rules of play, to ensure that player choices are clearly comparable, alternative conditions games benefit from insisting on structured player decisions. This requires considerable work by the research team to formulate clear options, that none the less do not overly constrain player choice. These games also benefit from clear data collection

³⁶³ Ibid. p 52

templates that minimize the need to reformulate information in post-game analysis to support comparisons. This task is generally relatively straightforward when it comes to player decisions, though the difference in the granularity between SAFE's detailed force structure moves³⁶⁴ verses the more aggregate decisions made in the more recent RAND force structure game³⁶⁵ illustrates how similar design decisions might still results in different levels of complexity in the actual game. In the case of a game with more granular decisions, highly structured recording tools can be helpful, like the one from SAFE shown in Figure 6.4. However, the need to capture relatively unconstrained decisionmaking processes poses more challenges, since a structured capture approach will, by its nature, exclude a great deal of information. In the RAND force structure game, we opted to adopt a modified form of qualitative coding to produce comparable analysis of player discussion,³⁶⁶ which was relatively effective, but very time consuming. An alternative option would have been to constrain play discussion through the use of more structured rules—in this case we worried that such a choice would artificially constrain debate on key topics of our research, but for a game with a different purpose such a design might prove useful. Another alternative would be to survey players about their own decisionmaking processes using the same instrument, again allowing for comparison between runs of the game. While there are concerns about the credibility of such self-reported reflections on decisionmaking, existing psychometric research may be able to offer more credible survey instruments.³⁶⁷ However, regardless of how much the design team structures player choices in advance, it is always possible that players may develop unexpected, but valid, decisions during the game. In the SAFE games, efforts were made to minimize changes between iterations of the game, but some changes were necessitated by player decisions. Perhaps the largest was in play D, in which play was cut short after 3 turns because the players' focus had turned to arms control. While such decisions were consistent with the guidance provided to both sides, and thus consisted of "legal" play, the control team assessed that the teams were unlikely to return to building new posture, and thus were no longer contributing useful additional information to the game.³⁶⁸ The created a challenge for comparison between play D and other plays of the game that produces more turns of data, that had to be explicitly discussed in comparative analysis.³⁶⁹

³⁶⁴ See: Averch and Wildhorn, "Risk, Ambiguity and Force Structure: An Analysis of General-War Forces and Strategic Objectives in Cases C and D (of Six Case Studies)." pp 5-10

³⁶⁵ Bartels et al., "Do Differing Analyses Change the Decision?: Using a Game to Assess Whether Differing Analytic Approaches Improve Decisionmaking." pp 32-35

³⁶⁶ Ibid. pp 46-47

³⁶⁷ I am indebted to Christopher Nelson and Andrew Parker for their thoughts on the application of existing literatures on decisionmaking to game design and data collection










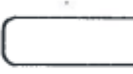
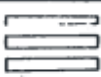




³⁶⁸ Brown and Paxson, "A Retrospective Look at Some Strategy and Force Evaluation Games." p 31

³⁶⁹ Averch and Wildhorn, "Risk, Ambiguity and Force Structure: An Analysis of General-War Forces and Strategic Objectives in Cases C and D (of Six Case Studies)." p 5

Figure 6.4: Sample of Menu of Investment options in SAFE

Initially available procurement items

R+D

Nuclear Systems				Improvement of Nuclear Features	
ITEM	Symbol	Ref. Performance			
1	M-BOMBERS		E, G ¹ , A, A ¹ =  in 64-67 in 67-70 g+p ¹ = 18 (see Reference A) Range: 4000 (unref), 5100 (ref) Low-alt penalty: 500 per zone CEP: 1 Load: 12, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100		
Prod: 200 Oper: 80+10g+5g ¹ +40a+20a ¹	Unit 45				
2	H-BOMBERS		E, G ¹ , A, A ¹ =  in 64-67 in 67-70 g+p ¹ +2a+2a ¹ = 18 (see Reference A) Range: 7500 (unref), 5100 (ref) Low-alt penalty: 500 per zone CEP: 1 Load: 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100		
Prod: 600 Oper: 150+5g+5g ¹ +40a+20a ¹	Unit 45				
3	SSM-BOMBERS		E, G ¹ , A, A ¹ =  in 64-67 in 67-70 g+p ¹ +2a+2a ¹ = 18 (see Reference A) Range: 5000 (unref), 7000 (ref); CEP: 1 Speed: 1. Last Speed; 2. Launch point; 3. CEP; 4. Load: 12, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100	103	R+D on SUPERSONIC A/C BOOSTS
Prod: 1000 Oper: 60+5g+5g ¹ +80a+40a ¹	Unit 45				In Items 3, 4 incl decoys R: 30 per 10% D: 50
4	SSH-BOMBERS		E, G ¹ , A, A ¹ =  in 64-67 in 67-70 g+p ¹ +2a+2a ¹ = 18 (see Reference A) Range: 7000 (unref), 9000 (ref); CEP: 1 Load: 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100		
Prod: 3000 Oper: 300+10g+5g ¹ +60a+30a ¹	Unit 45				
5	TANKERS		1 unit of Tankers supports one unit of either Item 1 or 2 or 3 or 4 at a level of a = 9 or two such units at a level of a = 0		
Prod: 200 Oper: 60	Unit 45				
11	POWER BASE COMPLEX		The 3 bases of a complex must be uniform with regard to occupancy, shelters, and point defenses. Hardness: 7 psi Max base capacity: 2 Wings of B-5s, or 1 Wing of other B ₅ , plus Tankers		
Prod: 80 Oper: 30	Unit 1				
12	SHELTERS FOR H-BOMBERS		1 unit = 6 shelters (= 2 shelters for each base of a complex) Hardness: 100 300 1000 * Constraint: Maximum of 2 units per complex		
Prod: 12 36 72 *	Unit 6				
13	SHELTERS FOR SSM-BOMBERS		1 unit = 6 shelters (= 2 shelters for each base of a complex) Hardness: 100 300 1000 * Constraint: Maximum of 2 units per complex		
Prod: 6 12 24 *	Unit 6				
14	SHELTERS FOR SSH-BOMBERS		1 unit = 6 shelters (= 2 shelters for each base of a complex) Hardness: 100 300 1000 * Constraint: Maximum of 2 units per complex		
Prod: 2 12 36 *	Unit 6				
21	I-ZONE ASMs		Range: 1 some Abort: 20% Yield: 2 (see Reference A) CEP: 1 (see Reference A) Constraint: Uniform usage by all A/C of same type and same alert status	121	R+D on DGR CEP FOR ASMs
Prod: 50 Oper: 15	Unit 45				In Items 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100
25	ARMED RECCE EQUIPMENT		Every ship or a single base the installation is for 2 Unmanned: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 Manned: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 Load: 12, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100	125	R+D on DGR ARMED RECCE
Prod: 150 Oper: 30	Unit 125				In Item 25 use Imp row R: 10 per 10% D: 30

Items 1, 2, 3, 4

Reference A

g = No. of A/C on permanent ground Alert
 a = No. of A/C on permanent airborne Alert
 E, G¹, A, A¹ should be multiples of 3
 Airborne Alert A/C assured to be 1000 miles (equivalent of 4 zones) from enemy penetration areas at time of outbreak

g¹ = No. of additional A/C on ground Alert during tension only
 a¹ = No. of additional A/C on airborne Alert during tension only and on ground Alert otherwise

R+D CELL

*Hex one corresponding pair of numbers and cross out the others.
Any similarity of the above performance and what figures to the actual figures is purely coincidental.

For example, the SAFE team was particularly interested in tacit communication between the two sides—that is, the signals that were sent by “decisions on R&D, procurement, and unilateral deactivations of operational systems.”³⁷⁰ However, the research team noted that game rules would “bias and/or constrain tacit communication in some respects. Hence, we must re-examine any conclusions that are suggested by analysis.”³⁷¹ As a result, the team was careful to analyze the ambiguity of signals available (particularly to the blue team about red’s intentions), and how the game rules might have impacted how these signals were communicated.³⁷² The team also theorized about how *different* game rules regarding the flow of information might have changed blue’s decisions as a way to add additional caveats to the analysis of the game.³⁷³ The detailed nature of this analysis makes resulting claims about how Blue’s understanding of different red objectives shaped game play more credible.

Challenges to Comparison from Adjudication

Beyond the structure of player choices is how the outcomes of those choices are adjudicated. One specific issue worth considering is the inclusion of random chance in game rules. Games often use chance as a convenient means of introducing uncertainty into a game, particularly when modeling the aggregate result of many sub-actions.³⁷⁴ However, insertion of chance means that while it is possible to be transparent about the probabilities, it is less likely that outcomes will be consistent over a relatively small number of games. As a result, more deterministic adjudication may be preferable. In the SAFE game, researchers opted not to represent process delays or attenuation of direction due to bureaucratic dynamics—in other words players were assumed to have perfect control over the department’s activities.³⁷⁵ While this is a simplification that gave player more control than their real world counterparts, it ensured a greater degree of comparative leverage since designers did not have to account for differences due to random chance.

A somewhat related point is the role of human error in adjudication. Even with a formalized set of rules, adjudication often takes place under a fair degree of time pressure with many distractions. As a result, it is not surprising that many adjudicators will confess they make errors in the moment, which could complicate comparison. To some extent, this can be mitigated by careful record keeping and oversight, so errors are noticed and corrected. It may also be helpful to computerize complex, fully formalized adjudication processes. Retrospective consideration of

³⁷⁰ Ibid. p 3

³⁷¹ Ibid. p 3

³⁷² Ibid. pp 27-52

³⁷³ Ibid. pp 53-63

³⁷⁴ Greg Costikyan, *Uncertainty in Games* (Cambridge, MA: MIT Press, 2013).

³⁷⁵ Brown and Paxson, "A Retrospective Look at Some Strategy and Force Evaluation Games." p 2

the SAFE game design noted that computerized adjudication aides, which were not available at the time of play, would have been helpful in ensuring that outcomes were produced consistently with less human effort.³⁷⁶

The team was careful to document the process for developing key game inputs. Cost-effectiveness was generated by first calculating how many of each type of weapon would be needed to target different broad types of targets, assuming perfect weapons reliability and no functional defenses against incoming warheads. The costs of the weapons system was then multiplied by the requirement they generated. While this process was predicated on many assumptions, it had the benefit of being transparent. In fact, the team occasionally opted to provide “an assessment aid...with which the reader may crank into the graphs any reliability or reliability penetration probability combination he chooses.” On a similar logic, the SAFE game opted to randomize the costs of “far out” systems, to represent systemic uncertainty, but provided the distributions to allow for concrete analysis.³⁷⁷

Conclusions

Alternative conditions games are designed to produce information about how different condition shape decisionmaking. These games depend on comparison between multiple runs in order to establish patterns of similarities and differences. Because of this, much of design is dedicated to insuring that variation outside of the factor of interest is minimized as much as possible. This is made particularly complicated by the role of groups of human players, introducing inevitable variation that is not always transparent, but can be compounded by other design choices. To the extent possible, designers should work to minimize the difference between runs of the game. Where uncontrolled variation does emerge, researchers should be careful to transparently note the potential problem and discuss how it might affect the credibility of results.

³⁷⁶ Ibid. p 50

³⁷⁷ Thomas A. Brown, "Elementary Cost-Effectiveness Computations Based on the Blue Menu of the Safe Game," (Santa Monica, CA: RAND, D-10717, 1962). p 1

Chapter 7: Designing Games for Innovation

The third type of game presented in Chapter 4 are games designed to generate innovative solutions to policy problems. These games are designed to produce candidate solutions which can then be subjected to further research and analysis, and thus innovation games tend to occur fairly early in research processes. Relatedly, these games are usually designed to inform internal audiences such as the sponsor and research team. Innovation games are most associated with critical realist approaches to research—the goal is to generate candidate ideas that make the most sense in the game but will need to be tested out before one can be confident how they will transfer to the real world. Analyticists may also build innovation games as models to “play” with or “prototype” potential ideas,³⁷⁸ but they may find the barriers to transferring their ideas into the real world to be too onerous to make the costs of a game worthwhile. Positivists are likely to see games as a very expensive approach to hypothesis generation that can be accomplished other ways for less cost.

The nature of innovation in national security spaces must fundamentally shape both the goals and structure of games designed for innovation. Like any other activity to generate new ideas, this requires that the designer create a space that fosters collaboration of different types of expertise in a space suitable for low cost experimentation with new ideas.³⁷⁹ In policy settings, where problems tend to be complex and high stakes by nature, it can be difficult to create the space need for experimentation without being accused of generating ideas that never have a chance of working, so careful framing of the environment that describes the problem is critical. Finally, innovative solutions will also tend to be more evolutionary than revolutionary in character³⁸⁰—but this should not be taken to mean that the change they offer is insubstantial or does not offer real benefits. Taken together, this suggests that for innovation to occur, games need to create a space that brings together diverse perspectives to tackle a concrete policy problem in a specific context, with the goal of making adaptive changes.

As a result of the nature of national security innovation, policy analysis game designers face a difficult challenge where they must design a game that loosens some status quo constraints to create space for new solutions to emerge but must be careful to retain constraints needed to ensure realistic outcomes. The game must present a problem that is hard enough to be useful, while not being too difficult to feel productive. Players should have access to a good

³⁷⁸ Michael Schrage, *Serious Play: How the World's Best Companies Simulate to Innovate* (Boston, MA: Harvard Business School Press, 2000).

³⁷⁹ Ibid. pp xiv-xvii

³⁸⁰ Theo Farrell, "Military Adaptation in War," in *Military Adaptation in Afghanistan*, ed. Theo Farrell, Frans Osinga, and James A. Russell (Palo Alto, CA: Stanford University Press, 2013). p 7

understanding of both the nature of the problem and the currently understood solutions, since new candidate solutions are likely to evolve from existing understandings. At the same time, to encourage collaboration and motivate players, the innovation game designer needs to make solving the policy problem at hand enticing enough to motivate player engagement and creativity. As a result, often innovation games are framed as a puzzle or competition, since many individuals are naturally motivated by these frames.³⁸¹

Specialists in innovation gaming also argue that innovation gaming cannot only be about the initial idea—appealing ideas must be implemented to ensure they are not half baked, and ideas that at first do not seem promising can be usefully refined through implementation.³⁸² Games can be particularly good at this task, because they bring together players with different experiences who can “walk-through” how a strategy, policy, or concept might be put into practice bringing their own experience to bare. Adjudication processes can then serve as a “second set of eyes” to further refine ideas and raise potential sticking points for player consideration.

This chapter illustrates these imparities translate into design tradeoffs using two games: Persistent Hobgoblin, which sought to develop new concepts of operations against a near peer competitor, and OCEANS 17, which focused on developing new processes and procedures for sharing forces between elements of the U.S. military. After describing the general purpose and design of each game, the chapter then describes some design tradeoffs that are common in innovation games related to the environment, actors, and rules of a game. These tradeoffs are illustrated with examples of how the two game designers opted to confront these challenges as examples to help make the more theoretical points concrete to the reader.

Overview of Example Innovation Games

The two games described below are intended to highlight some of the range of innovation games. Often the focus is placed on innovation in terms of warfighting, and Persistent Hobgoblin offers a fairly classic example of two-sided operation conflict with the goal being the development of a “theory of success”³⁸³ about how best to execute military operations against a near-peer adversary. However, innovation games can also be used to study a wide range of other processes and policies in national security spaces. OCEANS17 is an example of a game that seeks to uncover innovative solutions to what is fundamentally a bureaucratic problem: operating across the division between geographic regions controlled by different organizations within the U.S. military. The differences in the nature of the problem to be solved highlight some of the different decision designers can make to achieve the shared goal of innovation.

³⁸¹ For thoughtful discussion about the role of competition in innovations games, I am grateful to Graham Longley-Brown (interview Washington DC, August 2018) and Philip Pournelle (interviewed Washington, DC, March 2019).

³⁸² Interview with Philip Pournelle, Washington, DC, March 2019.

³⁸³ The term “theory of success” is from Compton. “Analytical Gaming.”

Innovation in Warfighting: OSD(P)/ CNA Persistent Hobgoblin Series

Persistent Hobgoblin was a series of wargames run in the mid-2010's as part of the early efforts to implement the Deputy Secretary of Defense's memo to revitalize wargaming for innovation in the department. Sponsored by the Under Secretary of Defense for Policy (USD(P)) Strategy and Force Development and designed by CNA, the game was designed to fill a gap in the process of developing new warfighting capabilities. In the words of one of the architects of the project: "there were lots of PowerPoint charts about capabilities, but there wasn't a lot out there on what the actual operational concepts were to turn those capabilities into actual warfighting capability... without the organizational constructs [emerging technologies] don't matter."³⁸⁴ The goal of the game was to generate fleshed out organization constructs and operational concepts based on a "cookbook" of capabilities that would then feed into later games.³⁸⁵ The set of concepts generated by players was documented in what became known as a "playbook" that described how the future capabilities were used in game play.³⁸⁶ This playbook then became a starting point for refinement and adaptation in later games that could build on strengths and mitigate weaknesses of previous concepts as a spur to creativity.³⁸⁷

Games in the Persistent Hobgoblin series were two-sided game between the U.S. (blue) and a near peer competitor (red) engaged in theater warfare. The games forced on a series of four tactical vignettes in the context of defense planning scenarios, with the primary goal of documenting 1) assumptions about the projected technical specifics of blue capabilities and 2) player concepts for how to employ those capabilities to solve specific key operational problems identified by past analysis as being critical challenges. Both teams were small, handpicked, and featured operators with expertise on specific warfighting domains. Each vignette featured a series of tactical engagements that were adjudicated using a semi-rigid process.³⁸⁸

The project was originally envisioned as a five-game series, though only the first two are discussed in the following analysis. The first game focused on developing the "cookbook" detailing the capabilities and embedding them in organizational constructs and warfighting concepts. After the first game, Deputy Secretary of Defense Robert Work suggested capturing the concepts into the "playbook," which could be referenced in the later plays of the game. In

³⁸⁴ Interview with Jacob Heim, Arlington, VA, July 2018.

³⁸⁵ The notion of a "cookbook" was adopted from the Halsey Alpha games run at the U.S. Naval War College. Interview with Jacob Heim, Arlington, VA, July 2018.

³⁸⁶ Ibid. The idea of a "playbook" was adopted from the 20XX games, run by CSBA in the 1990s which also considered the role of emerging technology in future concepts. One lesson learned from that effort is that game players were not always able to easily move from the specifications of a new technology, so it was helpful to have a "playbook" of how the technology had been used in past games. More details about the 20XX games can be found in: Michael Vickers and Robert Martinage, "Future Warfare 20xx Wargame Series: Lessons Learned Report," (Washington, DC: Center for Strategic and Budgetary Assessments, 2001).

³⁸⁷ Interview with Jacob Heim, Arlington, VA, July 2018.

³⁸⁸ Ibid

addition to the generation of the playbook, the other major change between the first and second run of the game was the addition of a more substantial red team to stress the blue concepts. While the tactical vignettes and blue team were relatively unchanged between plays, the red team was increased in size. They were also provided the opportunity to modify the red force composition, in order to better mirror the types of real-world adaptation as a result of intelligence we would expect to see from a near peer competitor. As a result, despite the operational context remaining unchanged, blue faced a more stressing fight in the second game. While not discussed in detail below, the third game of the series introduces a new blue team, the fourth expanded the tactical vignettes to include limited actions prior to and after the engagements represented in the original vignettes, and the fifth examined competitive technology development in peacetime.³⁸⁹

Innovation in Processes and Procedures: U.S. AFRICOM/ RAND OCEANS 17 Table Top Exercise

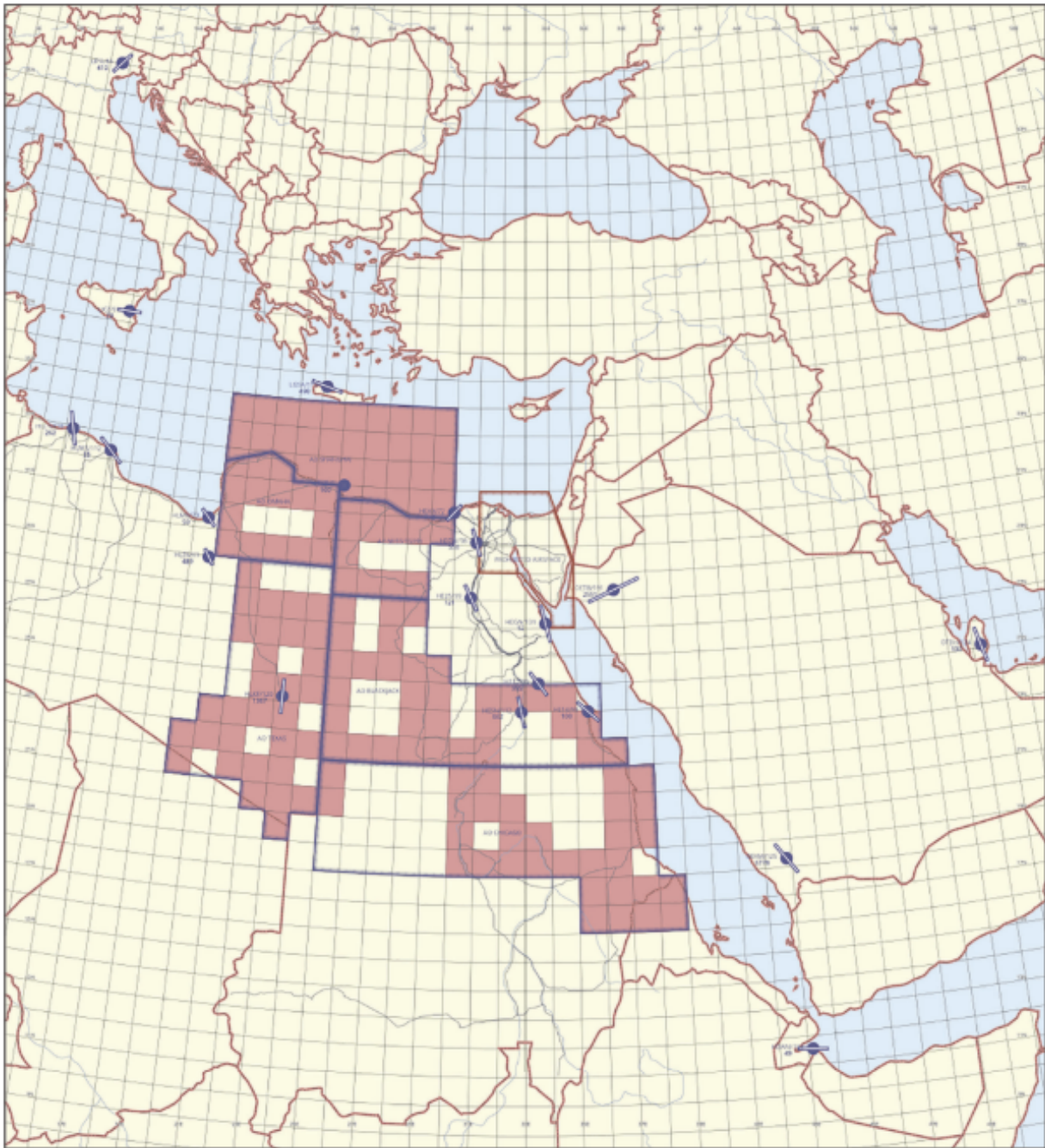
OCEANS 17 was a 2017 game designed by the author which had the goal of highlighting new processes and procedures that could allow military aircraft to operate more effectively in cases where a single threat actor operated across the boundaries between the geographic combatant commands (GCCs) area of responsibility. For the game, we posited a terrorist organization operating across the Egyptian-Libyan boarder, including a naval capacity in the Mediterranean islands, as shown in Figure 7.1. Since Egypt is in U.S. Central Command (CENTCOM), Libya is in U.S. Africa Command (AFRICOM), and the Mediterranean is in European Command (EUCOM) the geographic location of the threat required the three commands to work together to provide air assets and basing for optimal coverage. Since such cross-command operations are currently managed on an *ad hoc* bases, the problem posed was designed to spark a discussion about new approaches to managing more routine sharing of assets between the combatant commands.³⁹⁰

Given the nature of the problem, all participants, who generally had operational experience in the region, represented U.S. military officers tasked to recommend solutions to manage the command and control (C2) of air craft supporting trans-regional operations. Players were divided into three teams, each focused on a specific mission—personnel recovery (PR), intelligence reconnaissance, and surveillance (ISR), and close air support (CAS). These missions are all characterized by the use of a small number of specialized aircraft, accentuating the need to develop means of sharing limited aircraft between the commands to gain operational efficiency. In the first move of the game, players were confronted with the operational problem, and

³⁸⁹ Ibid.

³⁹⁰ Elizabeth M. Bartels et al., "Oceans 17 Tabletop Exercise: Findings and Recommendations," (Santa Monica, CA: RAND Corporation, RR-2521-OSD, 2019). p 1

Figure 7.1: Game board for OCEANS 17, Showing the Threat Actor in Red and Available U.S. Air Bases in Violet



Source: Elizabeth M. Bartels et al., "Oceans 17 Tabletop Exercise: Findings and Recommendations," (Santa Monica, CA: RAND Corporation, RR-2521-OSD, 2019) p 2.

challenged to develop a plan for operations in which aircraft were assigned to and based within a single command. In addition to the map shown above that illustrated the area of operation of the threat group and available bases, players also had range-finding aids to illustrate the potential area of coverage of different basing arrangements.

The first move of the game asked players to try to optimize coverage while maintaining the GCC boundaries, meaning that air assets had to base and operate in only one GCC. The solution players developed was captured as a baseline by which to measure the difference between the status quo and the solutions the players developed to allow transregional operations in the later moves. The second and third moves of the game asked players to develop ways of organizing and operating that would allow them to increase the efficiency of their coverage—that is cover more area with the same number of assets—by coming up with new procedures to operate across GCC boundaries. Proposed solutions included a range of options from putting new rules in place at the beginning of the operation to establish clearer expectations about how assets would be managed in the event of a time sensitive incident to standing up a small command and control organizations that could support specific missions.³⁹¹ The alternative concepts developed by the players were then “stress tested” in the fourth move with a series of random events demanding air coverage, which players were asked to describe how their chosen solution would respond to the operational demand signal.³⁹²

Design Tradeoffs Related to the Game Environment

Since innovation games are fundamentally about posing new solutions to policy problems, the environment of these games plays a critical role in framing the problem in a way that motivates players to solve it. At the same time, the way the game environment is designed is critical in shaping the credibility of the game’s results—if the game is seen as representing a fantasy, stakeholders will not buy into the solutions posed by players. Many security problems become abstract when removed from a specific context, so for innovation to be motivated, and for the results to be credible, games must grounded in specifics.³⁹³

First, the game environment clearly frames the problem in a way that motivates player engagement and problem-solving. A compelling narrative, and particularly the nature of the central conflict of the game, can help motivate players, both by tapping into their competitive spirit and sense of mission. In the case of competition with near peer competitors, the stakes can seem obvious when treated abstractly—of course a major war with a formidable adversary is an important national security challenge. However, in practice, players can be skeptical about the plausibility of any specific scenario on a wide range of ground. With lower-stakes problem sets,

³⁹¹ Ibid. pp 5-6

³⁹² Ibid. pp 1-3

³⁹³ Interview with Philip Pournelle, Washington, D.C., March 2019.

like the terrorist organization in OCEANS 17, this challenge can be compounded. For that game, considerable effort was taken in advance of the game to ensure that the available aircraft and bases could not supply adequate operational coverage for the problems presented so that players would be motivated to solve the problem.³⁹⁴ In effect the research team had to conduct a detailed analysis of the geography in order to ensure that we set up the game to be sufficiently stressing, without making it a puzzle that was artificially contrived.

While altering some aspects of the game's environment may be helpful to open up space for innovation, as a general rule many aspects are outside the control of any of the game's stakeholders and must be preserved to ensure credibility. They are thus somewhat less likely to be a promising avenue for innovation. Instead, the environment is more likely to be a source of constraints that players must work within. In the case of OCEANS 17, the game design team worked hard to ensure that the limitations of available bases and aircraft were accurately represented—for example, each base made available to players had a limit on the number and type of aircraft that could be located there, based on real world data about the facility. Similarly, the ranges of the different aircraft were depicted to the scale of the game board, so players could understand the area different assets could realistically cover.³⁹⁵ These restrictions on player options meant that players were forced to work within real world constraints that would affect actual operational staffs when developing potential solutions.

Finally, solutions to national security problems are deeply contextual. This means that innovation games tend to suffer if an environment is too abstract, since the abstraction likely lacks a realistic set of limitations that make the problem hard in the first place. It also means that designers must be careful before making claims that an innovation that works in one context will be portable into another. The first two Persistent Hobgoblin games looked at a set of four different vignettes, to cover a range of different environments. Later games added additional contexts, since there was no guarantee that the solutions developed in one environment would transfer to others.³⁹⁶ Alternatively, in OCEANS 17, we attempted to characterize the aspects of the problem and associate what types of solutions tended to align with the different variations of the problem—in other words to characterize which solutions we thought might be more appropriate in different settings.³⁹⁷ However, fundamentally, this issue highlights that innovation games are rarely satisfying as standalone efforts—instead they work best to tee up additional stages of research that will better characterize candidate solutions and the environments where they might be effective.

³⁹⁴ Bartels et al., "Oceans 17 Tabletop Exercise: Findings and Recommendations." pp 3-4.

³⁹⁵ Ibid. p 1.

³⁹⁶ Interview with Jacob Heim, Arlington, VA, July 2018.

³⁹⁷ Bartels et al., "Oceans 17 Tabletop Exercise: Findings and Recommendations." pp 3-6

Design Tradeoffs Related to the Game Actors

Competition is a key component in this archetype. To motivate players to discover new ways of acting, the game must motivate them with a problem. Most often, this comes in the form of seeking to achieve an objective against a thinking, reacting adversary.³⁹⁸ However depending on the problem, it could also be found in semi-cooperative relationships in which the different motivations between offices, departments, or partner countries provide tension as players strive to overcome an environmental challenge such as a natural disaster or disease. Persistent Hobgoblin was of this former type, whereas OCEANS 17 took the form of the later, where the threat was treated as part of the game construct so that it could be fully manipulated by the control team in the interest of setting up the very particular challenge we wanted blue players to focus on.³⁹⁹ In part, this motivation is psychological—games tap into competitive instincts that can fuel creativity. Competitive dynamics also aid the task of innovation game analysis, by providing a “test” of the new ideas. Competitive perspectives incentivize players to highlight potential flaws in suggested ideas and to manifest unintended consequences. This allows the “dominant strategy” to emerge from a set of potential ideas.⁴⁰⁰ As a result, when a red team is used in an innovation game, its composition is critical to the success of the game, even when the nominal focus of research is on uncovering blue strategies.

Challenges of Blue

Like other types of games for analysis, the selection of blue team players is an important tool for generating vigorous debate, new ideas, and stakeholder buy-in. In particular, it is worth considering where potential players fall within existing organizational hierarchies when recruiting. Players who are very senior or respected will lend credibility to game results, however since their ideas are already listened to, they may not be in a position to add new solutions to the conversation. In contrast, inexperienced and junior players may welcome the chance to inject new ideas which they might not be able to share through other channels, but the ideas they generate may not be credible as a result of that inexperience. One option is to aim for a “goldie locks” position in between the two extremes. In OCEANS 17 we requested players with several years’ experience in relevant operational roles, but specifically asked for mid-level officers, rather than senior officers to provide a forum for new voices. Other mitigation options include using highly credible adjudication approaches to compensate for less experienced players or teaming less experienced players with more experienced mentors who can provide additional perspective.

³⁹⁸ Interview with Philip Pournelle, Washington, D.C., March 2019.

³⁹⁹ Bartels et al., “Oceans 17 Tabletop Exercise: Findings and Recommendations.” p 1

⁴⁰⁰ Interview with Philip Pournelle, Washington, DC, March 2019.

Another option is to draw from communities that are already considering future operations, and thus are not bound by the status quo. In Persistent Hobgoblin, players were individually selected operators from the joint force with the goal of getting “the smartest person on [a particular] domain”⁴⁰¹—this approach ensured strong expertise and diverse bureaucratic representation that enhanced buy-in from service and other organizational stakeholders, while still gathering critical thinkers willing to generate new ideas. However, this option may be less successful in cases where game sponsors have less influence—“give me your best” only works if you are in a position to be the most important priority facing the office providing the player. While a game designer can work with sponsors to communicate the importance of the game to stakeholder offices, realistically recruiting is often as much an issue of competition priorities that are outside the designers control.

When games are run more than once, there is also a consideration about the relative benefits of having the same players participate. On one hand, having players re-engage with a problem, particularly one they failed to resolve previously, can be a means of motivating new ideas and approaches.⁴⁰² In Persistent Hobgoblin largely the same blue team lead headed up the first two games, with the second iteration providing an opportunity to refine plans against a more challenging red.⁴⁰³ On the other hand, there are benefits to drawing on a wider range of participants, who bring diversity of perspective and fresh eye on the problem. The Persistent Hobgoblin playbook represents one way to split the difference—the ideas of previous teams are readily available to new players to build on, but additional voices can be added to the game injecting new approaches as well.

Challenges of Red

Selection of red players for innovation games can be tricky. On one hand, generally the analytic focus of innovation games is on uncovering new blue strategies. As a result, it can be tempting to pay less attention to the selection of red players. However, to the extent that a game design requires the motivation of competing against an adversary or that competitive tensions are seen a key to uncovering a “theory of success” the selection of red players is, in fact critical. In the initial Hobgoblin game, the red team was somewhat smaller than the blue team, since it was not the focus of play. However, later games increased the size and expertise of the red team, including allowing red to adapt their procurements based on the new capabilities blue was bringing on line so that the red actor represented a more formidable challenge. This ensured that the ideas put forward were properly stress tested in the game, before determining whether to commit additional resources to studying them.

⁴⁰¹ Interview with Jacob Heim, Arlington, VA, July 2018.

⁴⁰² Interview with Philip Pournelle, Washington, DC, March 2019.

⁴⁰³ Interview with Jacob Heim, Arlington, VA, July 2018.

One common tool used to discuss the nature of a red team is the “Caffrey Triangle” that argues that red team can have one of three primary objectives—win the conflict, mimic red doctrine, or to support the white team by teeing up specific challenges of discussion for blue.⁴⁰⁴ Innovation game practitioners tend to be looking for both a red team that will stress and challenge blue by trying to win, but also will credibly mimic red decisionmaking processes to ensure that the proposed strategy is useful in the particular context represented by the game. Finding players who can achieve both of these goals is not trivial—for any given adversary the number of true experts who can emulate red decisionmaking credibly who are also able to conduct competitive play is not large. Adding in other common restrictions on recruiting players, including considerations of sensitivity and classification, can narrow the pool even further. This limited population of red players creates a particular problem for innovation games, because it means that there may not be as much diversity of perspective available, and thus fewer tools to generate new ideas on the part of red. This can result in a red team that is not sufficiently competitive simply because its approach can be anticipated by blue players from past experience with red player analysts.

Design Tradeoffs Related to the Game Rules

Game rules that govern how teams make decisions and what happens as a result can be a fruitful area for a game design to increase the freedom available to players to try new approaches. In contrast to the game environment, which often represents aspects of the policy problem that are outside of the control of the key actors, game rules around decisionmaking generally represent bureaucratic processes and procedures, or what decisions are considered valid moves, and thus could be fruitful areas to recommend changes. As a result, the rules of innovation games are often more unconstrained than alternative conditions and evaluation games. However, it is important to note that in any national security issue, the ease of changing procedures varies dramatically, so an important caveat to analysis of the game should be a sense of the other consequences of any change—in particular, whether game play uncovered unanticipated second order effects of the new processes and procedures.

At the same time, game rules must be careful to still implement credible limitations. This is particularly important when considering the consequences of actions, particularly in two sided games when the outcomes of actions are likely to serve as an initial screen by players about whether an approach is worth pursuing. This requirement often requires that designers find a balance between openness that will allow players to try new ideas, and a transparent adjudication process that gives those ideas a fair shape. In the case of OCEANS 17, players were able to conduct adjudication in the last stage of the game themselves—they could see on the map what

⁴⁰⁴ Caffrey, *On Wargaming: How Wargames Have Shaped History and How They May Shape the Future*. pp 322-323

areas of the game map were covered by their proposed basing solution at what level of coverage, and then see whether that met the standard called for by the event card.⁴⁰⁵ In contrast, Persistent Hobgoblin used a semi-rigid adjudication structure to resolve combat in order to have enough detail to force players to argue about the correct value for parameters for the different technical capabilities as a means of eliciting the details need to populate a “cookbook” with technical specifications for the different systems in play. Again, this device gave players a fair amount of latitude in what actions they took, but force players to explicitly discuss why they should be able to have the effects they wanted, providing traceability of the results.

Conclusions

Innovation games are designed to produce information about potential solutions to a policy problem. Generally, the goal is to develop fleshed out candidate solutions that are promising in the specific context of the game to help focus future areas of study. Designers are challenged to frame a useful problem for players to solve that is difficult enough that status quo solutions are insufficient, but not so difficult that no options are available to players. This requires determining which aspects of the game environment and rules need to constrain player action, and which current restrictions can be relaxed in order to allow players to make decisions that are not available to them today. Finally, designers must consider how to build a collaborative space that evokes players sense of competition, either with a rival team or the problem.

⁴⁰⁵ Bartels et al., "Oceans 17 Tabletop Exercise: Findings and Recommendations." p 1

Chapter 8: Designing Games for Evaluation

The fourth archetypic of information generated by a game for policy analysis described in Chapter 4 is evaluation. These games face an inherently difficult task: provide information about the outcomes of a proposed solution, whether concept, capability, strategy, or plan, with enough fidelity that the plan can be judged. The fidelity to which these outcomes can be determined is debated. Games of this type need to enable players to make decisions they think will solve a policy problem and then generate outcomes that are plausible enough to be judged in a way that is seen as credible by individuals outside of the game team *despite* the artificialities inherent in games. The standards by which the outcome is judged can vary, but common options are comparison (that is, which of several options produces the best outcome), sufficiency (that is, whether the projected outcomes of a proposed solution meets some pre-established minimum standard), or expert assessment of the “goodness” of the projected outcomes based on heuristics or other tacit standards. As discussed in Chapter 4, the inherent difficulty of this task has long made observers of all philosophical persuasions skeptical of these types of games. However, since evaluation games are frequently run by national security organizations, it is worth devoting attention to how they can be designed to meet their objectives, and what limitations should be noted in the analysis.

To date, game designers have attempted to issue a range of different recommendations about how best to treat the problem of evaluation, each of which is applicable under some, but not all, philosophical approaches to research. One particularly elegant approach comes from Tom Moaut, who argues that the evaluation offered by games takes the form of “COA falsification”⁴⁰⁶—that is, a game can point out why a solution is likely to fail, but if the game shows a plan is successful, it should not count as strong evidence that a plan will work. This framing may be particularly attractive to researchers trained in positivism, since it is consistent with the dominant Popperian approach of nullifying hypotheses. A second tack, often taken by critical realists, is to argue that the game produces a “theory of success” that comes out of the competitive dynamics of the game.⁴⁰⁷ Games can generate additional evidence that aligns with the theory, and thus can add credibility to its claims, but a game cannot be used to adjudicate whether the findings of the game will translate into real world environments—the theory remains nothing more than a “best guess.” Finally, for analysts who treat games as a type of model, the key is to be clear about the limits of games (such as a limited ability to know if a game run represents the “central tendency” of a model the way you might with a computerized model), but recognize that these limits, such as incomplete real world data about future environments and systems, will be shared with many

⁴⁰⁶ Tom Moaut, personal communication, November 2019.

⁴⁰⁷ Compton. “Analytical Gaming.”

other approaches. For these researchers, the claim is not that games are perfect, but that for some problems they are better than the other tools of evaluation available, and thus are still appropriate to use.⁴⁰⁸ Regardless of approach, all of these arguments are fundamentally a plea for nuance and caution in using a game as strong evidence in favor of a particular solution or strategy—a game may be the best evidence available, but to claim it “proves” or “validates” a plan is to overpromise what can be delivered.

Bearing these limitations in mind, the chief concern in designing an evaluation game is building a space that will allow players to implement the plan, and then see the outcome of the choices so that the plan can be assessed. The attention of the game designer is generally focused on the rules of the game, and rules that govern the projection of outcomes (that is, the adjudication system) in particular, since these are core to the credibility. Because evaluation games fall later in a program of research, decisions about what actors and aspects of the environment are most important are often relatively straightforward. However, choices about how to best represent them to enable credible evaluation are key. In particular, often there is a tension between the desire for simplicity in how these aspects are represented, which allows for clearer observation and measurement, and complexity of depiction that can be seen as a more credible approximation of real-world environments and actors.

This chapter considers two different games designed to evaluate decisions to illustrate these tradeoffs. The first, SCUD Hunt, was designed to examine the ability of different communications systems to improve shared situational awareness. The game took an experimental approach to design, using a fairly simple game in order to generate highly credible findings about the differences in behavior in a fairly abstracted environment. The second game was designed to investigate alternative approaches to security force assistance (SFA). The game’s design leveraged past research on SFA to generate heuristic guides to suggest the plausible likelihood of operational success. Because the literature on SFA suggests success is driven by multiple factors, a more complex game design was necessary to capture the policy issue of interest, but came at the cost of clarity of findings compared to SCUD Hunt.

Overview of Example Evaluation Games

In this case, I have opted to use two game which both explore activities outside the bounds of kinetic force-on-force conflict. Since these defense areas are less common in gaming, they make for an interesting exploration of design principles without falling back on arguments about specific operational theaters, current concepts, or standard modes of adjudication that are commonly debated in the field but often cannot be replicated well here due to sensitivities.⁴⁰⁹

⁴⁰⁸ For clarifying this point, I am indebted to David Shlapak.

⁴⁰⁹ One recent and publicly accessible example of this kind of debate can be found in the discussion of RAND’s Baltic Wargames in War on the Rocks. See: David A. Shlapak and Michael W. Johnson, "Outnumbered, Outranged, and Outgunned: How Russia Defeats Nato," *War on the Rocks*, April 21 2016; Michael Kofman, "Fixing Nato

Instead, the two examples used here look at two, quite different areas of military operations—communication between echelons of force and advising partner militaries to improve their capabilities. Both games sought to build out our understanding of what works and why by designing games that supported evaluation of different player approaches.

Experimental Design: DARPA/CNA ScudHunt

The first example is an experimental approach to research design used in CNA and ThoughtLink Inc.'s SCUD Hunt game, designed for DARPA.⁴¹⁰ The game was designed to examine which of six possible capability packages for collaboration most improved shared situational awareness across a distributed team using a computer interface.⁴¹¹ The research design featured six conditions, each consisting of one of three types of communications capabilities and one of two shared visualization tools. The three capability levels consisted of no communications, internet text chat, and a telephone conference call. The two types of shared visualization played were no shared information, and a visualization of all search outcomes from the across the team displayed at the end of the turn.⁴¹²

Each team consisted of four player all working together to meet the common objective of correctly locating the hidden SCUDs.⁴¹³ Teams were informed they represented asset managers from the military of a state, and the environment they were seeking was the home territory of an adversary which was threatening to attack a friendly state.⁴¹⁴ The game featured a highly abstracted environment, shown in Figure 8.1. Players confronted a 5x5 grid of featureless squares and with little information about the adversary other than their hostile intentions toward a friendly state, and that intelligence stated that there were concealed SCUD missile launchers that they were tasked to locate.⁴¹⁵ Each player had full control of one or more types of assets, which they were free to deploy to search the game board as the player saw fit (that is, individual players were not bound to follow the decisions of the team if they opted not to).⁴¹⁶ Players played

Deterrence in the East Or: How I Learned to Stop Worrying and Love Nato's Crushing Defeat by Russia," *ibid.* and Karl Mueller et al., "In Defense of a Wargame: Bolstering Deterrence on Nato's Eastern Flank," *ibid.* The RAND Baltic Wargames are surveyed in Chapter 9.

⁴¹⁰ Peter Perla et al., "Gaming and Shared Situational Awareness," (Alexandria, VA: Center for Naval Analysis, 2000). and Perla, Markowitz, and Weuve, "Game-Based Experimentation for Research in Command and Control and Shared Situational Awareness."

⁴¹¹ More broadly, the study was intended as a proof of concept for an experimental game decision, particularly in studying situational awareness. Perla et al., "Gaming and Shared Situational Awareness." pp 2-4

⁴¹² *Ibid.* p 32

⁴¹³ *Ibid.* p 2

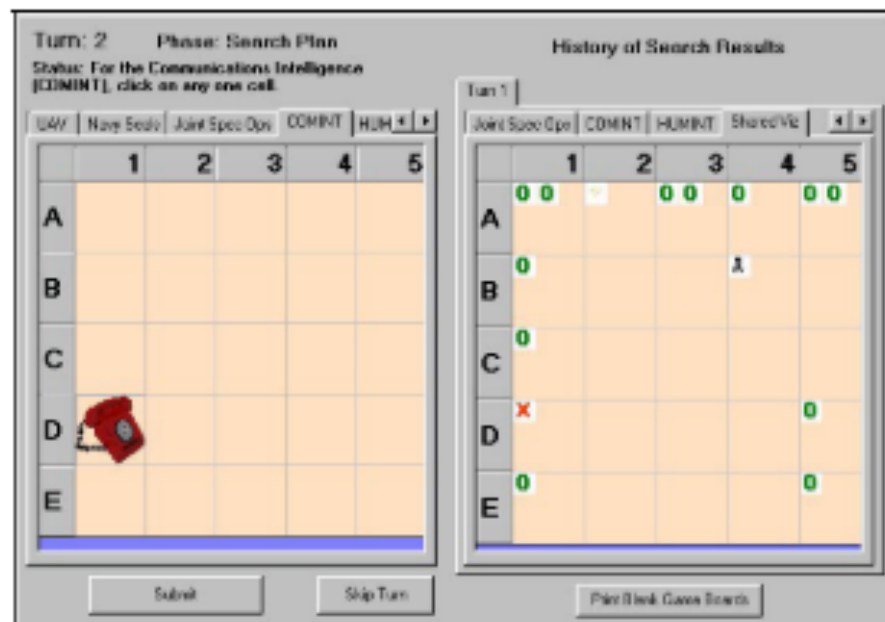
⁴¹⁴ *Ibid.* p 28

⁴¹⁵ *Ibid.* p 2

⁴¹⁶ *Ibid.* p 29

in physical isolation from one another, and in most games players had not met and did not have any information about the background of their fellow players.⁴¹⁷

Figure 8.1: SCUD Hunt Game Board



Source: Anthony H. Dekker. "Revisiting "Scudhunt" and the Human Dimension of Ncw: Some Thoughts." (Canberra, Australia: Defence Systems Analysis Division, DSTO, Australian Department of Defence, Undated) p 2.

Rules of the game were similarly simple. At the start of the game, the location of the target SCUDs was randomly selected. SCUDs could not move during the course of the game, so the players' task was to determine the original location. Somewhat more detailed rules governed the capabilities of player assets, which included availability, the probability the sensor would be detected and destroyed by adversary forces, and the probability of returning information on a given grid square when sent to search it.⁴¹⁸ After players allocated their sensors, they were provided an update that included whether the asset had survived, and if so, what information it was reporting. This information was generated based on 1) whether the SCUD was in the space (that is, ground truth) and 2) the capabilities of the system selected by the player. This information included a range of confidences, including potentially false information.⁴¹⁹ Players could opt to share information about the search capabilities of their assets, results of searches, search plans, and recommended target with their team members as they saw fit using

⁴¹⁷ Ibid. p 38

⁴¹⁸ Ibid. p 28-29

⁴¹⁹ Ibid. p 29

the tools provided by the communication capability package being tested in that run of the game.⁴²⁰

The team found a statistically significant improvement in shared situational awareness from having the ability to communicate, but none differentiating the type communication system. Having a visualization system also improved the teams' shared situational awareness. Finally, the effect of having both communication and visualization was greater than the sum of each individual systems benefit.⁴²¹ Later re-analysis of the results also suggested substantial differences between different teams and argued that the ability of the groups to coordinate was fundamentally unequal for a range of reasons.⁴²²

Comparative Case: AFRICOM/RAND Security Force Assistance Game

A second approach can be found in RAND's 2017 Security Force Assistance Game, designed to support AFRICOM. The goal of the game was to compare and contrast the effects of three different strategies of investment in potential partner forces in Libya.⁴²³ The game allowed three teams, each tasked with implementing a specific strategy, to make decisions in parallel to one another, allowing easy comparison of the operational decisions and potential outcomes in order to assess the comparative benefits and drawbacks of each plan.⁴²⁴

All teams were provided with the same starting scenario, which depicted a plausible, near-future state of Libya. At a more tactical level, players' view of the environment was provided in the form of detailed information about the interests and current capabilities of a large number of different armed groups within Libya. Each team consisted of multiple players with experience as planners and analysts supporting U.S. operations in Africa, who were placed in the role of combatant command planners. Players were tasked to work together to develop a multi-year plan of investment that would operationalize their assigned strategy. Game play was focus on how to select groups, and what types of training they would receive, in order to achieve the objectives laid out in the strategy. Players were given a fixed budget to allocate for each move, and asked to make a series of decisions:

1. Which organizations were they willing to invest in?
2. Which units within those organizations would receive training?
3. What capabilities would be targeted for training?

⁴²⁰ Ibid. p 2

⁴²¹ Ibid. p 35-36

⁴²² Anthony H. Dekker. "Revisiting "Scudhunt" and the Human Dimension of New: Some Thoughts." (Canberra, Australia: Defence Systems Analysis Division, DSTO, Australian Department of Defence, Undated).

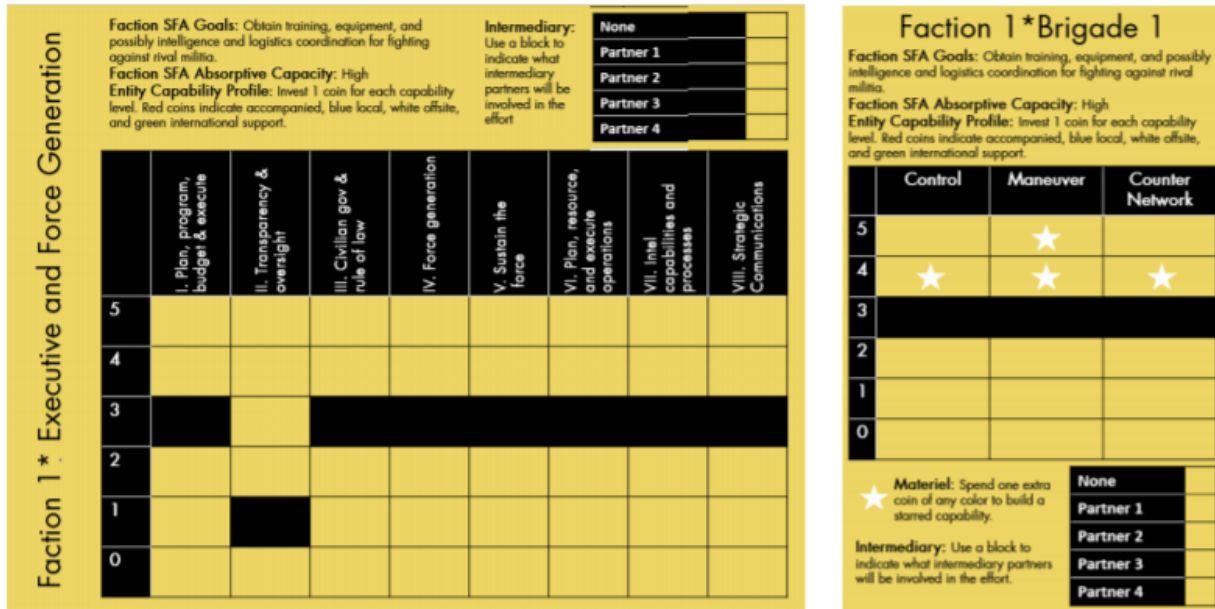
⁴²³ Elizabeth M. Bartels et al., "Conceptual Design for a Multiplayer Security Force Assistance Strategy Game," (Santa Monica, CA: RAND Corporation, RR-2850, 2019). p 1. Reporting on the findings of the game is not publicly available.

⁴²⁴ Ibid. p 3.

4. How would training be provided?

To represent these choices, players were given a fixed number of wooden coins they could place on paper cards (shown in figure 8.2) to indicate what investments they wanted to make. Choices about how the training was provided were shown by the color of the wooden coin and additional markings made on the card.⁴²⁵

Figure 8.2. Sample cards to show investment in potential partner forces



Source: Bartels et al, "Conceptual Design for a Multiplayer Security Force Assistance Strategy Game," p 8.

The success of these investments was then determined using a two-step adjudication system. First, the outcome of each investment was determined using a probability table based on the historical probability of the success of investments. Player decisions were used to select what probability table was used, representing the structure advantages and detractions of key choices. A die was then used to select a specific outcome, representing the many factors outside U.S. control that shape success. Second, subject matter experts on the team examined how the success of investments changed the relative capabilities of the different groups and projected how these changes to the balance of power would likely affect the political, security, and economic trajectories of the conflict.⁴²⁶ The adjudication strategy allowed players to compare the outcomes achieved by the different strategies on multiple dimensions and discuss the relative costs and

⁴²⁵ Ibid. pp 6-8

⁴²⁶ Ibid. pp 8-13

benefits of the different approaches. Teams are able to discuss the relative outcomes, and thus assess the comparative performance of the strategies.

Design Tradeoffs Related to the Game Environment

The environment of most evaluation games is driven by two primary considerations: the information the players need about the policy context in order to implement the course of action (that is, input to the play's decisions) and information needed to inform the outcomes of player actions (that is, input to adjudication). In both cases, there is a tension between the desire to include more detailed renderings of the environment in the belief that this will represent a more "realistic" version of the environment, and a desire to simplify the environment to make for more traceable, comparable results. In other words, two different factors contributing to the credibility of analysis tend to cut against each other in the design of the environment and must be carefully balanced.

One commonly adopted approach, particularly for games focused on kinetic conflict, is to use existing computerized modeling and simulation programs to supply the game environment. The advantage of this approach is that the game gains the credibility of the modeling system, and the environment can be assured of linking directly to the adjudication rules since they leverage the same software. However, there are several drawbacks to this approach. First, even well-established modeling and simulation efforts are not universally credible. For example, research conducted by CAPE in the 2010s revealed concerns that the complexity of major modeling and simulation software made appropriate validation possible, raising concerns about the results generated.⁴²⁷ Second, existing computer models and simulations may not have the flexibility needed to include future systems and environments properly. Some major models only include forces that are in the program of record in order to ensure that high quality data is available about performance. While designers may be able to develop a work-around, at some point these adjustments compromise ability to rely on existing validation work, thus careful documentation of the modifications must be included. This solution is also unsatisfying when considering a range of issues that are not handled within the existing simulation. Since simulations usually focus on the technical capabilities of systems, "soft" factors like command and control or will to fight are often not well represented.⁴²⁸ Military actions other than kinetic conflict may not have any appropriate modeling systems available at all. Finally, many manual game designers raise concerns that involving computer systems dilutes the focus on human debate and decisionmaking, mitigating the advantage of running a game at all. This problem is particularly acute when using models that do not have well-developed user interfaces, in which the volume of

⁴²⁷ Interview with Igor Mikolic-Torreira, Arlington, VA, October 2018.

⁴²⁸ Ben Connable et al., "Will to Fight: Returning to the Human Fundamentals of War," (Santa Monica, CA: RAND Corporation, RB-10040-A, 2019).

information about the environment provided to the player can be overwhelming. As a result of these limitations, other means of representing the environment in games are often needed.

However, the tradeoff of moving away from a computerized system is quickly apparent—computers are simply better than humans at accurately tracking and displaying a highly complex environment than manual displays. A designer opting to a manual environment will be forced to limit what is represented. On one hand, this need not mean that the environment is inherently simple—as with other types of games, players contribute their own understanding of the environment to flesh out a richer vision that can be as or more complex than what a computer can display. However, the downside is that this complexity is resident in the mind of the participants and (as discussed in detail in Chapter 5) can be difficult to capture and communicate to those not directly involved. Because the goal of evaluation games is to generate information that is credible to those who are not present in the game, failure to develop such a description that is persuasive to outsiders can be fatal to the utility of the game. As a result, while unspecified environments can be used, they pose risks if not carefully managed.

Regardless of whether the game environment is computer or manual, the common issue of balancing sufficient complexity (to make for credible decisions and outcomes) with sufficient simplicity (for clear tracing and communication) is evident. For a designer to make such decisions about what aspects of the world to include, they must have a well-developed model of what factors about the environment will influence both player decisions and the projected outcomes. The model informs the minimum necessary set of factors, which still may be quite extensive depending on the nature of the policy problem under study. The game must then allow analysts to trace interactions between these environmental factors, player decisions, and the outcomes projected by the rules of the game with sufficient clarity to 1) understand player choices and 2) judge the projected outcomes of those decisions. Finally, this process must be explainable to others outside the game in order to develop the needed external credibility. The resulting degree of complexity will be different for each game.

One extreme of this is to develop a simple, highly abstract environment that allows players and analysts to focus on the essential decision. This approach may be particularly attractive in a positivist model, where the ability to make consistent observations is critical to generating credible information from the game. The SCUD Hunt team designed a very simple environment for the game. This decision was driven by the observation that previous studies on shared situational awareness using games had great difficulty exercising control and focusing the game on the research objectives of interest. The team opted to not try to represent the true complexity of the problem of finding assets (such as the details of terrain that might impact the effectiveness of a given ISR asset), in order to gain the ability to clearly see how players updated their mental models about the environment, and how these updates drove decisionmaking.⁴²⁹ Because the research was most interested in how team communication was working, rather than, say, the

⁴²⁹ Perla et al., "Gaming and Shared Situational Awareness." p 6

optimal mix of ISR assets to complete the task, stripping out environmental details clarified the ability to observe the role of shared situational awareness on decisionmaking, thus furthering the research objectives of the game.

The SFA game's design opted for a more complicated depiction of the environment, in part because the underlying philosophy of science of analysis created different demands on the game. The model of security force assistance developed by the game design team generated key feedback about the impact of changes in forces caused by the provision of assistance that allowed players to consider the advantages and drawbacks of their decisions. As a result, more information about the environment was needed for players to make decisions and see the outcomes of their choices than in SCUD Hunt. Instead of a generalized ally and adversary country, the SFA game featured a specific country, with a projected near-future environment. While the scenario was designed to be communicated quickly to players, it featured updates on key issues including the political and military balance of power, economic status of the oil sector, and relations with key regional powers—all of which were understood by the model of the game to be key indicators of strategic success. Updates on similar topics were provided to players after each move, allowing players to debate the strengths and weaknesses of their selected strategy.⁴³⁰ A simpler description of the environment would not have allowed players to receive understandable feedback, where as a more complex environment would have made it difficult to trace performance turn over turn.

Design Tradeoffs Related to the Game Actors

Actors in evaluation games can be overshadowed by design concerns related to the rules and, to a lesser extent, the environment. Yet how actors are represented can still play a role in the credibility of the game's results. First, selecting which actors are represented by human players, versus those that are abstracted away or handled only in adjudication can influence the realism of results. Second the selection of players can simplify or complicate analysis. Designer decisions in both areas can add credibility to game findings.

When selecting which actors are represented by players versus which are embedded in the game design, evaluation games always start with considering the organization that will implement the potential solutions, as they will be the focus of the game. This actor will always be represented by human players, and is the focus of game play. From there, the designer should consider 1) key actors who support the solutions and 2) actors who can block the solution. While not all these roles need be played by a human player, they will need to be considered in the design. SCUD Hunt was focused on the process of developing shared situational awareness of a static problem set, so only the blue actor needed to be represented by human players because the interactions between team members, and between team members and a static operational

⁴³⁰ Bartels et al., "Conceptual Design for a Multiplayer Security Force Assistance Strategy Game." p 13

challenge, were sufficient to answer the research question. Similarly, the SFA game was able to only use human player to represent the U.S. because the focus of the evaluation was on U.S. decisionmaking—as a result, static preferences from other actors were sufficient to provide an initial assessment of strategy.

In both cases, the behavior of other key actors—the adversary in the case of SCUD Hunt and the potential recipients of assistance in the case of SFA—could be prescribed. This had the advantage of creating a consistent challenge confronting players, making iterations of the game more comparable for evaluation. In the SFA game, treating the preferences of each potential recipient of assistance as static meant that teams faced equivalent challenges: while outcomes of the same investment might differ because of the role of probability in determining outcomes, the probability of success was the same, and thus no team had a structural advantage over another.⁴³¹ This control is not always necessary—there are other possible methods of evaluation other than companion—but can be a helpful design since it is often easier to assert that a plan’s outcome is “better” or “worst” rather than “good enough.”

In other cases, judgement of the solution depends on how other actors react to counter the solution. In the case of SCUD Hunt, if the problem set was changed to finding mobile missile launchers in which the pattern of deployment was changing in response to blue’s search solution, then a reacting red would be essential to evaluating the strategy.⁴³² Similarly, if the analytic question had turned from the costs and benefits of pursuing different US strategies to thinking about how investments in Libya partners could contribute to great power competition, additional teams of live players would have been needed to allow for the adversary’s responses to be included fully in the model.⁴³³ Kinetic conflict nearly always requires both a red and blue team be played by human players, so evaluation games considering combat are almost always too sided.

The player selected to represent actors can also be important, though generally they are not seen as critical as in the other three types of game. Because of the high degree of structure within the environment and rules, and the fact that the basic course of action under evaluation is already defined, evaluation games may be able to draw on less-expert players compared with other game types. However, familiarity with standard practices in the relevant domains is still key to ensure that players have basic credibility. In the SFA game, fairly clear guidance was provided to delineate three potential courses of actions, so that players without extensive experience in security force assistance could still implement the plans.⁴³⁴

However, as always, the extent to which designers ought to be willing to sacrifice experience for other considerations will be determined by what types of players are seen as credible by the

⁴³¹ Ibid. p 7

⁴³² Perla et al., "Gaming and Shared Situational Awareness." p 47

⁴³³ Bartels et al., "Conceptual Design for a Multiplayer Security Force Assistance Strategy Game." pp 23-24

⁴³⁴ Ibid. pp 6-7

consumer of analysis. For example, sponsors of operational military games almost always want to see military staff experience from all relevant services among the players. In cases where new concepts are being tested, it may also increase the game's credibility to have individuals who have advocated for the approach playing, so that they can assure others that the proposed course of action was played out as they designed it.⁴³⁵ There also can be concerns if the ability of players to implement the course of action correctly varies considerably. Later analysis of the SCUD Hunt game data by a different team suggested that the differences between teams was actually larger than the size of the effect of different technologies, potentially because of highly inexperienced players in some games, and the selection of team leaders who were particularly bad at core tasks in others.⁴³⁶ In some cases such concerns are not too impactful—in the case of SCUD Hunt, the experimental design of the game was strong enough that systematic differences between teams could be measured analytically as a check to make sure they did not eliminate the core findings. However, in other cases, differences in player ability can be difficult to parse out from the effectiveness of the strategy—that is it can be difficult to know if a strategy is genuinely better, or simply implemented more skillfully by clever players.⁴³⁷ This may be a concern when a great deal of discretion is left to the players. A useful rule of thumb is the more choices players have, the more salient the identity of players is likely to be to analysis.

Design Tradeoffs Related to the Game Rules

The rules of an evaluation game are critical to the credibility of findings as the rules will determine how players opt to implement to proposed policy solution and how the results of those player decision will be projected. As a rule, evaluation games tend to feature more formalized rules than other types of games because the need to produce credible projections of player actions is assisted by documentation. However, when these rules become too complicated, concerns emerge about unpredictable interactions generating erroneous results. This same lack of traceability can make complex rules more difficult to implement consistently and in a way where they can be audited by outside stakeholders. As a result, without relying on existing established models complicated game rules can be difficult to make credible, whereas simpler rules can be more transparent, and thus more credible.

The qualities of a “credible” rules in general, and adjudication model in particular, are not one size-fits-all—instead, the designer needs to consider what will be credible to the specific audiences of the evaluation the game is designed to produce. As with the environment, one particularly important tradeoff to consider is that between complexity and transparency. There is

⁴³⁵ Interview with Phil Pournelle, Washington, DC, March 2019.

⁴³⁶ Dekker. "Revisiting "Scudhunt" and the Human Dimension of New: Some Thoughts."

⁴³⁷ W. Phillips Davison. "A Summary of Experimental Research on "Political Gaming"." (Santa Monica, CA: RAND Corporation, D-5695-RC, 1958). p 6

long-standing literature within defense modeling arguing that as models gain complexity, it becomes all too easy for modeling assumptions to interact in ways that produce chaotic results;⁴³⁸ as rules feature more elements and interactions, it becomes more difficult to trace why a particular result is generated, and thus to assess that the result is credible. As a result, for an audience that values the ability to explain *why* a game's adjudication rules produced the result that it did, it is generally preferable to use a simpler model that allows the designer to explain the logic of how player actions produced the outcome in question. However, the very existence of literature on the negative effects of complexity in defense modeling points to the entrenchment of an opposing view arguing that more complex models are better simulations of real-world phenomena. To convince stakeholders holding this perspective of the usefulness of a game's results, it may be necessary to demonstrate the ability of the games rules to replicate a detailed vision of the environment.

In some cases, it is possible to use existing models to help structure the game but where accepted rules are not available more work on the part of designers will be needed. For example, attrition warfare using existing systems have been extensively modeled—as a result while different gamers might make somewhat different choices in rulesets, the majority of force-on-force rule sets will be readily recognizable to designers across gaming organizations. When gaming on less common topics, consensus models may be less readily available. The RAND SFA game developed an original rule set for projecting the results of SFA training by summarizing trends from the historical case study literature on factors that contribute to the success.⁴³⁹ By documenting the trends found in the case study literature, the designers could offer a clear pedigree for the adjudication model, bolstering credibility. Furthermore, by documenting these rules, they are available for other teams to build on and improve as the state of knowledge about SFA improves over time.

It is also necessary to demonstrate that the rules were executed in a consistent matter. While this may sound trivial, anyone who has observed a large game with many staff members and players interacting knows how easy it is for new processes and interpretations to emerge under pressure to accommodate unexpected, but reasonable player requests and produce many results quickly. This requirement tends to push evaluation games towards the use of explicit rules that are legible to all players, adjudicators and consumers of the game results who may wish to inspect the game processes. This requires a fairly established understanding of what actions players are likely to take, and a procedure to rapidly add any *ad hoc* decisions to the core rules so they can be implement consistently going forward.

⁴³⁸ For two well-known critiques of this type, see: Davis, "The Base of Sand Problem : A White Paper on the State of Military Combat Modeling."; J.A. Dewar, J.J. Gillogly, and M.L. Juncosa, "Non-Monotonicity, Chaos, and Combat Models," (Santa Monica, CA: RAND Corporation, 1991).

⁴³⁹ Bartels et al., "Conceptual Design for a Multiplayer Security Force Assistance Strategy Game." pp 15-22

Again, simplicity is one approach to developing internal consistency. The SCUD Hunt team noted that the simplicity of the rules helped focus attention on the outcomes of interest, which was the teams shared situational awareness, rather than the stated outcome of the game: finding the scuds. A more complicated game might have made it difficult to measure the latter, leaving researchers to attempt to disentangle how much of a team's ability to meet the objective was actual ability to develop a common understanding, and how much of it was the luck or skill of the players in finding hidden targets.⁴⁴⁰ While the SFA game rules were a degree more complicated to account for the wider range of player decisions about how assistance was provided—and the interactions of those decisions on the likelihood of success—the rules of the game were still simple enough to be presented to players on a single page of paper. These simple rules are easier to understand and monitor.

More complex games face greater challenges, since the rules can no longer be trivially inspected. As noted in the discussion of game environments, the two approaches to complex topics are to rely on players to flesh out the actual complexity of the situation by self-monitoring their own actions and helping to project outcomes or using computerized models to implement rigid rules. The former is very difficult to ensure consistency, while the latter faces all the same barrier of computerized approaches discussed elsewhere in the chapter. As a result, it is advantageous to be as simple as possible in designing rules, and when that is not possible, to be prepared to manage additional questions about the credibility of the ruleset.

Another issue that needs to be considered is the role of chance in adjudication rules. Most games use random chance as a means of representing factors that are either not well understood enough or too contingent to be modeled well. In the SFA game, this included highly idiosyncratic factors like the preference of low-level commanders that historical case studies reveal to be important in determining success but highly opaque to outside analysts in advance. The random role of the die in the adjudication process then represents the factors that are unknown by U.S. decisionmakers and outside of their control which are critical to credibly modeling the phenomenon. This type of experience of uncertainty is often seen as key to the game.⁴⁴¹ However, analytically the use of random chance can create a problem for evaluation, because random results that generate extreme outcomes can make a plan look very good or very bad. In computerized modeling and simulation, these extreme outcomes are managed by running the same scenario many times so that the central tendency of the system is evident. However, since the difference in players and player interactions mean that games cannot be repeated in this way, that approach is not an option.

To avoid this problem sometimes designers opt to minimize the role of random chance in adjudication results. This has the advantage of providing more consistent feedback on the plan—

⁴⁴⁰ Perla et al., "Gaming and Shared Situational Awareness." pp 6 and 27

⁴⁴¹ For an excellent discussion of this issues of uncertainty and how it is represented in games, see: Costikyan, *Uncertainty in Games*.

extreme results are smoothed away. However, the cost is that the factors the randomization represents are removed from game play. If those factors are critical to how the solution preforms, removing them from the game risks the credibility of results. One particularly common example of this choice, and its potential costs, is the issue of will to fight. In most games, subordinate units will always act as ordered, regardless of the reality that soldiers and units are not automatons. In reality, hesitation, shirking, retreat, and desertion are all seen on the battlefield.⁴⁴² These behaviors can be modeled as a probability that the unit will perform actions as ordered, but this is rarely done in games. Other issues like the reliability of communication and detection of adversary units are other key issues that designers often opt to treat as given except in specialized context, but may be worth integrating more fully more often.

Conclusions

Evaluation games are designed in order to produce information to judge a potential solution or solutions to a policy problem. The criteria for judgement can vary, to include comparison between different courses of action, preset standards of sufficiency, or expert judgement on less defined criteria. The synthetic nature of games makes it impossible for games to single handedly “prove” that a solution will work, but they can offer potential pitfalls or modest evidence in support of a course of action as long as they are modestly interpreted within a particular philosophical system of claims.

Regardless of standard of judgement or philosophy undermining the claim, evaluation games need to product credible outcomes resulting from player decisions. This requirement puts considerable focus on the credibility of the adjudication rules. Key considerations are generating outcomes that are traceable to individuals who are not directly involved in the game in order to make sure results can be communicated clearly. This argues for relatively simple causal models, in which it is possible to clearly observe decisions, and present clearly understandable arguments about why outcomes occur. However, this pressure for simplicity cuts against a general belief that reflecting more of the complexity of the real world in a model will make for more realistic results. As a result, designers must balance considerations of complexity and simplicity based on what will be persuasive to the specific target audience of the game, and the state of credible knowledge about the topic of the game.

⁴⁴² Connable et al., "Will to Fight: Returning to the Human Fundimentals of War."

Chapter 9: Trends in RAND Corporation National Security Policy Analysis Gaming: 1948 to 2019

While Chapters 5-8 presented select RAND games in order to illustrate the four archetypes and design tradeoffs inherent in each, this chapter recontextualizes those games within the broader scope of RAND gaming efforts. In doing so, this chapter offers an alternative lens to the main thrust of this monograph's argument—rather than focusing on the enduring diversity of gaming, it focuses on how game design has evolved (or failed to evolve) over time, using RAND's practice as a concrete example of trends.

This survey of RAND gaming output reveals two key trends. The first is that the majority of RAND's gaming until very recently was focused on games to better understand policy problems—that is games that most closely resembles the system exploration archetype and, to a lesser extent, alternative conditions type. This contrasts with notable accounts of early games, particularly those run at the Naval War College, that tend to stress the application of games to operational innovation.⁴⁴³ However, this focus is consistent with RAND's historical position as a research institution seeking to understand emerging challenges. It is not until the 1990s, when RAND began providing substantial support to military-led games that innovation and evaluation games become a substantial focus of the organizations gaming work. This observation reinforces a core point of this monograph—that the design of games is tied to the purpose for which games are conducted. Ergo, we should expect organizations with different purposes to conduct different kinds of games. Thus, the historical patterns of RAND in terms of game type should not be expected to generalize to other organizations with different missions.

Second, the use of games has followed a cyclical pattern of boom and bust. It seems that games are popular when seeking to understand a new operating environment—that is in period of intense geo-strategic and technological change. Over time, as new issues become better understood, they are more tractable to other related techniques like standalone scenario development and computerized modeling and simulation, and gaming sees a period of less use. When new challenges that are not well suited to modeling and simulation again come to the fore, gaming sees renewed use. These trends are more likely to be shared across the broader national security analysis gaming establishment, and thus trends in relative amount of gaming are relatively more likely to be common across gaming organizations.

The following account of RAND gaming is, perforce, shaped by available materials. In some cases, there are known lacunae in the records. For example, more recent periods are less well represented in the archival evidence base. In part, this is due to changing standards of

⁴⁴³ Most notably the games run at the U.S. Naval War college, see: John M. Lillard, *Playing War: Wargaming and U.S. Naval Preparations for World War II* (Lincoln, NE: Potomac Books, 2016).

documentation and archiving—put simply, the rise of email has greatly reduced the preservation of internal working papers that make up some of the most useful resources on the details of game design. More recent materials are also less likely to be eligible for declassification and so further limits the sample. As a result, earlier periods of RAND's history are far more accessible than the last 25 to 30 years. In some cases, these gaps can be mitigated by interviewing researchers who were active at the time, but due to the size and atomization of RAND's workforce, there is little reason to expect that these perspectives are necessarily representative of RAND's total production. As a result, this account focuses more attention on earlier periods of RAND gaming and treats more modern efforts more superficially.

Gaming for Nuclear Comprehension at RAND: 1948-1958

The first documented RAND game was played in the fall of 1948 by a small team of RAND researcher, with the aim of playing out a global "cold war" at the strategic level, rather than engaging in operational or tactical warfare.⁴⁴⁴ These games considered both political and military activity which included both U.S. and Soviet atomic bombs.⁴⁴⁵ This effort grew into a series of games through 1949, designed:

to test in action the rules formulated for playing such a game. The sheer mechanics of conducting a war game at the grand strategic level presents vexing problems. We have been trying to work our way through these by 1) inventing rules which seem a priori, to be adequate; 2) trying them out in an actual – but rough and ready – game; 3) revisiting the rules in the light of that experience; 4) trying to revised rules, etc.⁴⁴⁶

A blue team representing the US and a red team representing the USSR were both composed of RAND different researchers, providing a range of different perspectives and experiences to further build out the rules.⁴⁴⁷ In the final form of the rules, time progressed in a series of set steps in which the umpire relayed information about world events since the last move, players could make additional requests for intelligence, propose actions, and make decisions to advance their strategy. Each side took a move by writing out their strategic plans and moves on a chart. The umpires would then declare whether the proposed move was allowed and if so whether the projected short-term consequences were adjudicated to have occurred, and what information the other team would learn about the opponent's actions. The opposing team was then allowed to respond.⁴⁴⁸ In other words, in the first years after its founding, RAND was playing games with

⁴⁴⁴ "War Games," *RANDom News*, October 1 1948. pp 3-4.

⁴⁴⁵ "War Games," *RANDom News*, October 29 1948. p 2.

⁴⁴⁶ "Research Note: War Games," *RANDom News*, January 21 1949. p 2.

⁴⁴⁷ "War Game," *RANDom News*, February 18 1949. p 3.

⁴⁴⁸ Olaf Helmer-Hirschberg. "War Game Rules (Fourth Version)." (Santa Monica, CA: RAND Corporation, D-446, 1949).

political, economic, and military stakes using umpired play in a structured format to better understand the new problem of atomic warfare.

Shortly thereafter, RAND also began to conduct more tactical and operational simulations of conflict, consciously following in the tradition of historical Kriegsspiels' rigid rules.⁴⁴⁹ These games, primarily associated with Alexander Mood, used many of the mechanics we now think of as core to operational board games—some in long use, such as the use of two separate maps depicting red and blue knowledge of the situation to enable representation of the fog of war, and some more novel such as the use of hexagonal grids to allow for more flexible representation of the direction of movement.⁴⁵⁰ The rules for these games were substantially more rigid than the first “Cold War” game. For example, processes to determine lines of sight, ability to fire on and kill enemy forces, and ability to capture enemy resources are all laid out in game rules.⁴⁵¹ Both the air and ground games as first played did not include bombing and thus exclude the possibility of exploring the emerging atomic tools' impact of warfare—however discussion of these games makes it clear that they were seen as a critical first step in developing games that might be suitable to analysis, in which:

The solution to a given problem usually means finding a sound strategy. The game representing the problem must be easily playable and must be played numerous times by the same players so that they can develop a knowledge of the structure of the game and a feel for good strategy. A game that is to be replayed many times needs a fixed set of rules so that experience gained in one play is valid in other plays. If a complete set of written rules is needed, then the game cannot represent a detailed global war... The game should include whatever context is needed for a proper treatment of the problem at hand, but no more. Further, those aspects which are retained in the game must be severely simplified and combined into easily manipulable factors in the interest of having a playable and understandable game.⁴⁵²

In other words, this more rigid approach sought to build up the understanding of a problem sufficient to be able to evaluate the “goodness” of strategies.

These first games show RANDs initial interest in working across the typology of games from system exploration to evaluation. However, as can be seen from these very first exploration games, RAND researchers realized a great deal more work was needed to understand the emerging shape of warfare in the atomic, and then nuclear, age before a sufficiently credible model to support evaluation games could be developed. Thus, the early years of RAND gaming were focused on system exploration games, that worked to build credible models of these new

⁴⁴⁹ John F. Nash and Robert M. Thrall, "Some War Games," (Santa Monica, CA: RAND, D-1379, 1952).

⁴⁵⁰ Ibid. p 1-1A

⁴⁵¹ Ibid. p 2-5 and p 9-11

⁴⁵² Alexander McFarlane Mood, "War Gaming as a Technique of Analysis," (Santa Monica, CA: RAND Corporation, P-899, 1954). p 4-5

policy problem. In this context games, were generally played by members of RAND research teams to provide:

mental furniture to give focus... [to teams that are] floundering and lacked focus or integration... After [game play], the project participants... had a common vocabulary which enabled them to communicate with one another with a degree of clarity and precision which had been impossible before.⁴⁵³

Key research topics that leveraged games included: force structure, posture, and planning; force employment in general wars; and the prospects for limited war between nuclear armed adversaries.

RAND's games in support of the U.S. Air Force complemented resurgent interest in gaming across the military. These efforts included the founding of a gaming capability in the Pentagon for the joint staff which focused also focused on system exploration games, implementation of computer-assisted games beginning with the Navy, and expansion of Army and Marine Corps gaming as a means of training a resource-constrained force.⁴⁵⁴ While RAND's work is often treated as emblematic of the era,⁴⁵⁵ it was part of a larger surge of gaming across a defense and national security community struggling to understand the consequences of the atomic and nuclear revolutions and shifting balances of power.

Early Force Structure, Posture, and Planning Games

Early games used to study the addition of atomic and nuclear weapons to high level force structure, posture, and planning considerations tended to take the form of relatively simple investment games with rigid rule sets. These tools were designed to illuminate the linkages between economic performance, force size and composition, and the resulting force's ability to damage the adversary economy in the event of a major war. While warfighting strategy was treated to a limited extent in these games, it generally took a back seat to efforts to understand the value of strategic investments in the emerging platforms—in other words, these games focused on building and understanding of the relationships between economic and military power, rather than on innovating new systems or evaluating the merits of any specific force.

The Planning War Games (1952) followed the model of the earlier Mood games, featuring pre-set rules intended to mimic “good military judgement”⁴⁵⁶ regarding the ability of an economy and population to support a military force structure and their operation in the European theater, as shown in Figure 9.1. The goal of play was “a rough classification of strategies”⁴⁵⁷ to

⁴⁵³ Brown and Paxson, "A Retrospective Look at Some Strategy and Force Evaluation Games." p 5

⁴⁵⁴ Caffrey, *On Wargaming: How Wargames Have Shaped History and How They May Shape the Future*. pp 74-

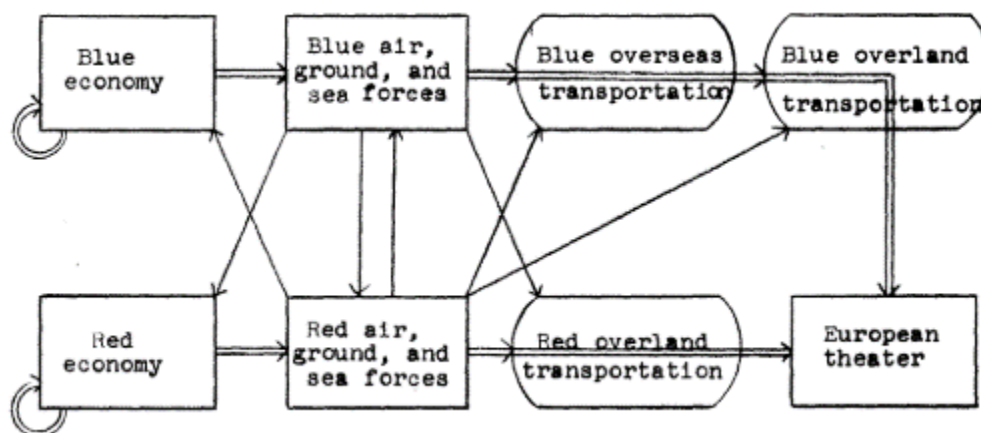
⁴⁵⁵ See for example: Wilson, *The Bomb and the Computer*.

⁴⁵⁶ Mood Alexander McFarlane Mood and Melvin P. Peisakoff. "A Planning Factor War Game." (Santa Monica, CA: RAND Corporation, D-1382-PR, 1952)., p 1

⁴⁵⁷ Ibid. p 2

balance investment in different forces, including atomic weapons, then observe their high-level ability to perform in a war. Perhaps not surprisingly given the focus on strategic bombing to degrade economic production in World War II, much of the game was focused on understanding the potential impact of strategic bombing on the ability of the economy to support the chosen force posture, rather than direct force-on-force attrition.

Figure 9.1: Planning and Aggregated War Game Process



Source: "Rules for an Aggregated War Game." (Santa Monica, CA: RAND Corporation, RM-1046-2, 1954) p 2

This work was continued the following two years with the Aggregated War Games series to study projected planning challenges for 1956 and 1960. The Aggregated War Games refined and elaborated the rules from the planning game (in fact, the length of the rule book nearly doubled). Particular depth was added to the theater air and base attack rules—allowing greater consideration of force on force attrition warfare--and new capabilities, like hydrogen bombs.⁴⁵⁸ However, the goal of play for the overall research agenda remained the same—to discover broad categories of strategy for planning to support a force through a major war in which nuclear weapons were available. Authors stressed that this iterative work was still in early stages, and the rule set too immature to be used as “firm” planning factors or for evaluation of a specific strategy without further development beyond what could be achieved in these early games.⁴⁵⁹

During this period, a gaming approach was also used to support games with a more educational bent. For example, the 1953 STRAW games (shown in Figure 9.2) were intended to educate players about the interactions between strategic targeting and the economy.⁴⁶⁰ Later efforts such as the Strategic War Planning (SWAP) game shared the educational objectives and rigid rules of the STRAW game but leveraged increasingly sophisticated understanding of the

⁴⁵⁸ "Rules for an Aggregated War Game." (Santa Monica, CA: RAND Corporation, RM-1046-2, 1954).

⁴⁵⁹ Ibid. p i

⁴⁶⁰ Marian Centers et al. "Rules for Straw." (Santa Monica, CA: RAND Corporation, D-1955-PR, 1953).

major aspects of air planning to create more robust rules to govern procurement, placement, and warfighting.⁴⁶¹ Game procedures were designed to be intuitive for players—for example, allocation of investments was managed by placing poker chips on a menu of potential purchases, removing the need for accounting, as shown in Figure 9.3.⁴⁶² Another interesting design choice was that the game played through five years of planning but then “rewound” the clock by selecting two of the years to play through global warfighting, giving the players a sense of the relative differences of the conduct of strategic war with different tool sets.⁴⁶³ Designers noted that rigid rules of the game meant that the game was more limited in scope, but that the advantage came in being able to compare plays of the game directly to one another. While the purpose of these games was not policy analysis, the design features developed as part of these efforts would be used to good effect in later RAND games, particularly SAFE.⁴⁶⁴

Figure 9.2: Play of STRAW with RAND Board of Trustees in 1953



Source: RAND Archive

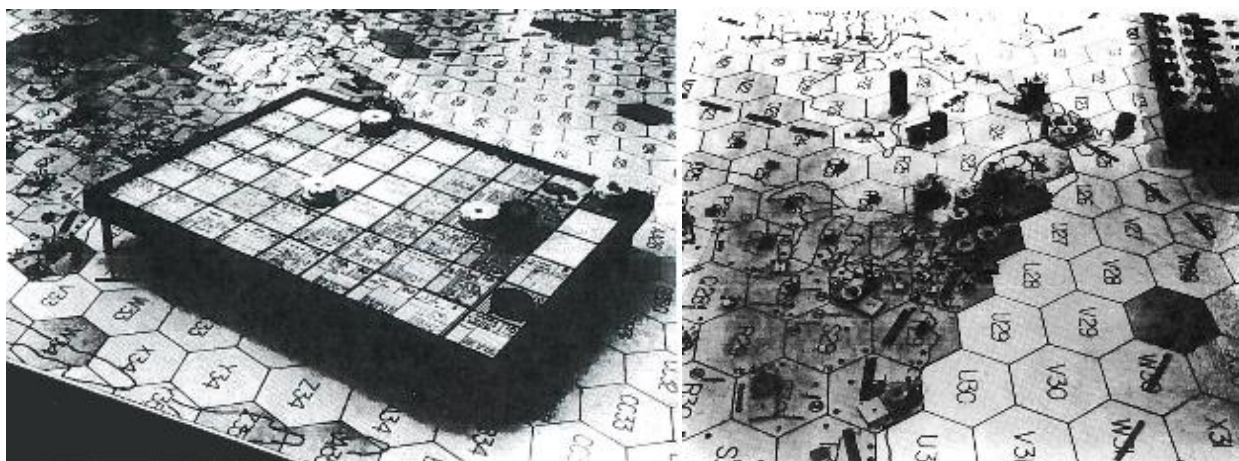
⁴⁶¹ Olaf Helmer-Hirschberg and Lloyd S. Shapley, "Brief Description of the Swap Game," (Santa Monica, CA: RAND, RM-2058-PR, 1957).

⁴⁶² Helmer-Hirschberg, "Strategic Gaming," p 8

⁴⁶³ Helmer-Hirschberg and Shapley, "Brief Description of the Swap Game." pp 5-6

⁴⁶⁴ Helmer-Hirschberg, "Strategic Gaming," p 15

Figure 9.3. Details of SWAP game boards for procurement and warfighting



Source: Olaf Helmer-Hirschberg and Lloyd S. Shapley, "Brief Description of the Swap Game," (Santa Monica, CA: RAND, RM-2058-PR, 1957) pp 9 and 10

Managing Crises and Fighting General Wars in Games

A second category of games covers a set of global problems similar to those in the planning games, but whereas crisis and conflict in planning games was intended only to stress test the force structure and posture, this second category of games focused attention on the details of how the conflict might unfold under the shadow of nuclear weapons. While some of these games expanded on the rigid rule systems used in the planning games, work in this vein also saw the development of less structured game formats. These games opted to return to the tradition of umpired Kriegspiel, limiting the use of formalized rules in order to game the flexibility to examine unanticipated dynamics. Other efforts attempted to split the difference between the two approaches, by designing games that leveraged both rigid rules and umpire judgement, depending on the issue at hand. The alignment between rigid rules for force-on-force problems and umpired games to examine political-military issues established a pattern that continues to define the majority of games to this day, but this period of gaming also illustrates how both approaches can be useful for systems exploration.

Rigid Rules for Air Combat Games

Using the model of Mood's earlier gaming work that employed rigid rules, a series of games was developed to look at "tactical" problems of theater warfare—what would now be considered operational level wargaming. In contrast to the strategic games that looked at global allocations, these games focused on the specifics of a particular theater. Both air and ground games were designed in this mode, with the goal of eventually integrating them to allow for the examination of joint operations, however the air rules are better documented in the available archival

resources.⁴⁶⁵ The focus of the air game was air superiority, which focused on calculating the number of aircraft and airfields available in the theater over time to generate interdiction and close air support for ground forces.⁴⁶⁶ Players made decisions and tracked resources on a “status board,” consisting of a map and detailed chart listing logs for bases, aircraft, and munitions shown in Figure 9.4. Players and game researchers used pins as indicators of current status, as well as intelligence on adversary forces.⁴⁶⁷ Because the rules were fixed, results tables could be pre-generated to ease playability.⁴⁶⁸ As a result, after players developed their plans, outcomes were generated using a “randomizing cylinder” which could be used to select a particular outcome from a pre-generated distribution, allowing stochastic results to reflect emerging research.⁴⁶⁹

Umpired Political-Military Games

In the mid-1950s, researchers were also exploring less structured formats. The best documented example can be seen in the Cold War games which focused far more on political decisionmaking⁴⁷⁰ and adopted an umpired approach to gaming to handle a wider, less predictable range of behavior. While the new series shared with the planning games a focus on war in Western Europe,⁴⁷¹ its designers hoped to produce a game without the “simplifying assumptions and special restrictions”⁴⁷² by developing a game in which researchers with relevant specializations represented the U.S., U.S.S.R, other countries, and phenomena outside the control of governments (through the excellently titled “committee on nature”) could take any action as they saw fit based on their understanding of capabilities and intentions of the actors.⁴⁷³ The goal was to minimize the number of rules and constraints and “allow expert opinion (referees, Nature Committee) to develop a corpus of rulings during play of the game itself”⁴⁷⁴ aided by discussion

⁴⁶⁵ Leo P. Holliday and Arnold S. Mengel. "A Tactical Air Superiority War Game." (Santa Monica, CA: RAND Corporation, D-2931, 1955). p 3

⁴⁶⁶ Ibid. p 4

⁴⁶⁷ Ibid. p 7-8

⁴⁶⁸ Ibid. p 6

⁴⁶⁹ Ibid. p 22

⁴⁷⁰ At the time, RAND researcher believed gaming focused on political decisionmaking was a novel innovation. It was not until several years into the effort that researchers discovered the Japanese Total War Research Institute had previously conducted games of this type. See: Davison. "A Summary of Experimental Research on "Political Gaming"." p 2

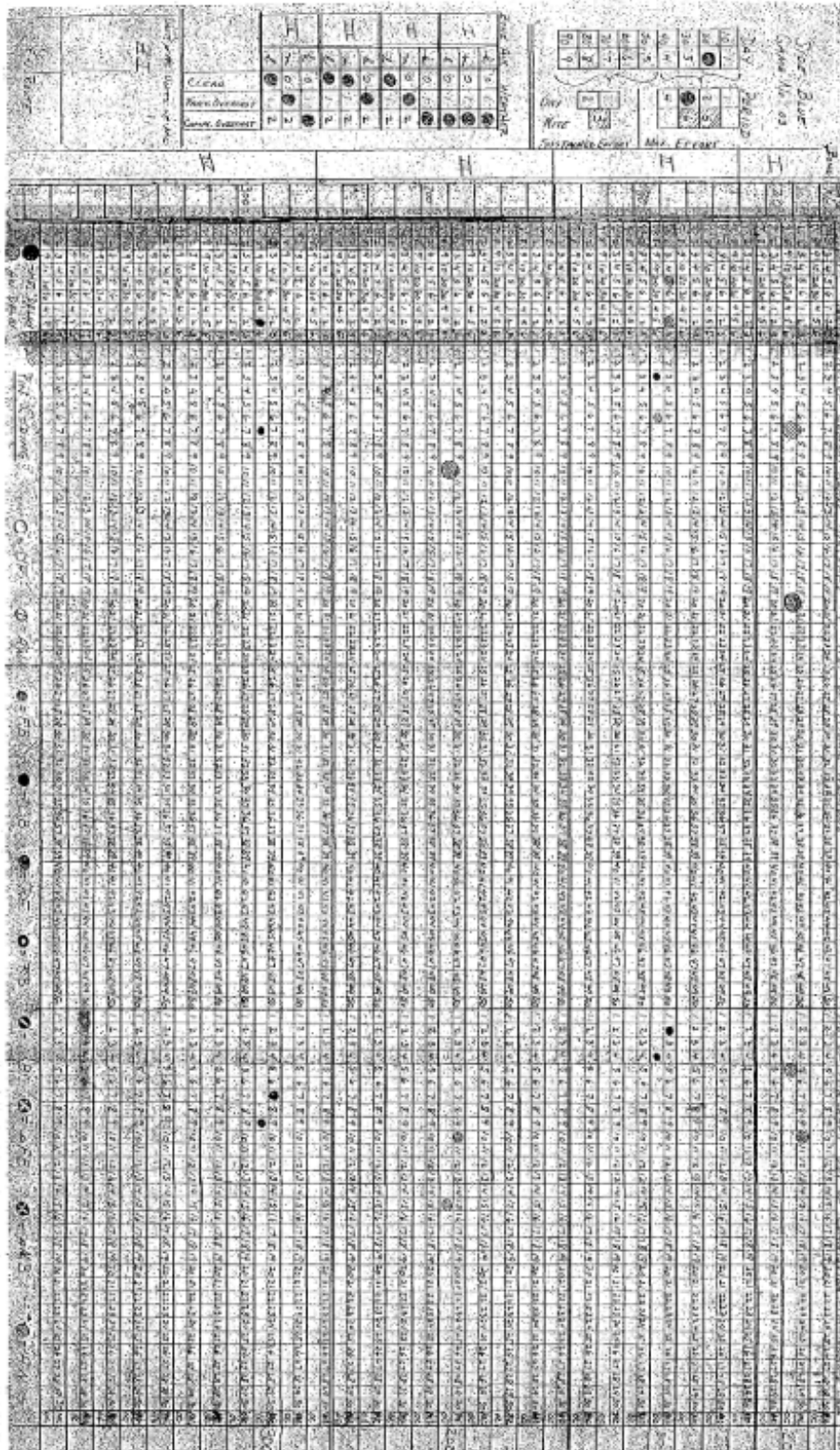
⁴⁷¹ Herbert Goldhamer. "Summary of Cold-War Game Activities in the Social Science Division." (Santa Monica, CA: RAND Corporation, D-2850, 1955). p 6

⁴⁷² "Toward a Cold War Game." (Santa Monica, CA: RAND Corporation, D-2603, 1954). p 1

⁴⁷³ Ibid. pp 2-4

⁴⁷⁴ Ibid. p 6

Figure 9.4: Status Board for Tactical Air War Game



Source: Holliday and Mengel, "A tactical air superiority war game" p 10

held before the game to scope the exercise.⁴⁷⁵ The games were also designed to put considerable focus on ways in which incorrect and incomplete information informed planning.⁴⁷⁶

As with the 1948 RAND “Cold War” games, players spoke (in early runs) or wrote out (in later iterations⁴⁷⁷) their move, the reasons for it, and the expected consequences of the action.⁴⁷⁸ Outcomes were not determined by pre-set rules, but rather determined in stride by the judgement of the umpires (and to a lesser extent, the Committee on Nature).⁴⁷⁹ While the designers recognized this would make comparative evaluation of strategies more difficult than the rigid rules used in the in the Planning War Games, they hoped it would provide for a more realistic range of decisions to be played in the game. The researchers’ belief was that by synthesizing the knowledge of a group of experts and identifying key factors influencing behavior on both sides, the game could potentially support forecasting.⁴⁸⁰ They assumed that while early play would be dominated by conflicts of opinion between experts, over time they would converge on a shared, credible model and more attention could be paid to outcomes.⁴⁸¹ Put in the terms of this monograph’s framework, over time the game would elicit and synthesize the mental models of expert participants to conduct system exploration.

By the end of the series, researchers determined that the scope of the games was simply too broad to generate useful insights in a cost-effective manner.⁴⁸² However the game was deemed useful in so far as it helped better define and prioritize issues for further analysis.⁴⁸³ Researchers also found the innovation of developing a scenario that projected the start of play several years into the future to be a helpful innovation⁴⁸⁴—as we shall see, focus on future scenario-building became a hallmark of later RAND umpired games.

⁴⁷⁵ "Summary of Cold-War Game Activities in the Social Science Division." p 3

⁴⁷⁶ Herbert Goldhamer and Hans Speier, "Some Observations on Political Gaming," (Santa Monica, CA: RAND Corporation, P-1679, 1959). p 11

⁴⁷⁷ 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 7

⁴⁷⁸ "Summary of Cold-War Game Activities in the Social Science Division." p c

⁴⁷⁹ "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 6

⁴⁸⁰ "Summary of Cold-War Game Activities in the Social Science Division." pp 7-10, "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 2

⁴⁸¹ "Toward a Cold War Game." p 8

⁴⁸² Joseph M. Goldsen. "The Political Exercise: An Assessment of the Fourth Round." (Santa Monica, CA: RAND Corporation, D-3640-RC, 1956). pp 57-58

⁴⁸³ Goldhamer and Speier, "Some Observations on Political Gaming." p 16

⁴⁸⁴ The notion of starting the game in the future was implemented in the fourth run, in order to try to minimize the extent to which game play could be overtaken by current events over the game period of several weeks. Ibid. p 8

A Middle Ground for Operational Games

Something of a middle ground between the narrowly specified rules of the Tactical Air Game on one hand, and the broad scope and wide-open strategic play of the Cold War Games on the other was also in evidence in early games. For example, a 1953 theater campaign game featured a red and blue team whose actions were guided by general rules. Like the Cold War Games, this approach depended on workshops and analysis conducted in advance to set ground rules, and to provide teams a chance to work out details of initial plans without slowing down game play.⁴⁸⁵ These early values could then be used as a basis for relatively rapid calculation to support adjudication during play, which would then follow the model of the rigid Tactical Air Game.⁴⁸⁶ In effect, the advanced work of players and game staff in workshops prior to the game was responsible for adjudication.

The resulting focus of play was the integration and execution of air, ground, and sea operations in order to explore the conduct of theater campaign plans. The game paid particular attention to timing and sequencing, using an electronic clock to track “accelerated war time” across all rooms that could be used by players to stop “time” across all groups as needed—for example to adjudicate the success of a critical atomic attack.⁴⁸⁷ This operational approach attempted to balance the rigid and open rule approaches but existing documentation does not provide a strong sense of how researchers rated the relative utility of this middle approach.

Gaming the Prospects of Limited War: Project Sierra

The third topic of early RAND system exploration games focused on limited war—that is, regional wars that may or may not involve nuclear exchange but did not involve attacks on the U.S. (or Soviet) homeland. As the boundaries of the Cold War solidified with the emergence of a meaningful Soviet nuclear arsenal, concerns emerged that war might consist of a limited incursion, rather than all-out general wars that early research had posited.⁴⁸⁸ RAND analysts were quick to recognize that the need to coordinate between warfare in the different domains, as well as political and economic dimensions of the conflict, required a research approach that could look at the interactions of the many parts of the problem.⁴⁸⁹ Teams also highlight the role of ambiguity and perception in the dynamics of limited war, making wargames with human decision makers an attractive tool.⁴⁹⁰ The largest of these efforts was the Sierra Project, which used dozens of manual games to explore joint warfighting in small scale conflicts against

⁴⁸⁵ Mood, "War Gaming as a Technique of Analysis." pp 2-3

⁴⁸⁶ Ibid. p 5

⁴⁸⁷ Ibid. pp 8-9

⁴⁸⁸ Harvey A. DeWeerd. "Nato Limited War Crises: Some Research Guidelines." (Santa Monica, CA: RAND Corporation, D-12201, 1964).

⁴⁸⁹ Weiner, "War Gaming Methodology." p iii

⁴⁹⁰ DeWeerd. "Nato Limited War Crises: Some Research Guidelines." pp 2-6

communist adversaries in Asia from 1954-1958. As the first of the major efforts to study limited war using games, the game designs were unusually well documented as an explicit deliverable of the project.

Rather than focusing on the Western European context, Project Sierra opted to examine a sizable number of different environments in which limited wars might occur. The early games focused on conflict in Southeast Asia before moving on to the Far East (including Korea and Taiwan)⁴⁹¹ and Near East (including Jordan and Israel).⁴⁹² Game play focused not only on air, ground, and maritime combat dynamics, but also political, economic, logistics, and intelligence factors prior to and during conflict to identify key trends and patterns across operations.⁴⁹³

Most games featured two teams of players, including military officers to supplement the operational understanding of the RAND staff, assigned to represent the “red” communist forces and “blue” U.S. and allied forces.⁴⁹⁴ The two teams generally worked through planning separately,⁴⁹⁵ with adjudication using a rule set supplemented by an umpire.⁴⁹⁶ Players first developed the political-military objective, which informed a general plan for the use of military forces. Once control approved these overarching strategic and operational choices, more tactical details of missions could be worked out by the players for their area of specialization.⁴⁹⁷

Over the four years of the project, designers experimented with a range of approaches to gaming to help achieve varying research objectives. For example, some games restricted the information available to players about adversary intentions and capabilities.⁴⁹⁸ Other games varied the extent to which the control team deferred to player recommendations regarding political decisions. Some games constrained choices (generally to study what could be achieved given a set of limited tools) while later games in a series tended to allow more flexibility in what actions players could take.⁴⁹⁹ Adjudication approaches also varied. Researchers urged the use of pre-calculated values where good data was available, often drawing on advances made in other

⁴⁹¹ By 1958 some 20 games had been played in the series, covering Thailand, Burma, Indochina, China, Formosa (modern Taiwan), and Korea. See: Paxson. "The Sierra Project -- a Study of Limited Wars." p 14

⁴⁹² Weiner. "War Gaming Methodology: Sierra near East Series."

⁴⁹³ Ibid. p 2

⁴⁹⁴ Paxson. "The Sierra Project -- a Study of Limited Wars." p 4. While two sided games made up the majority of Project Sierra games, some used a three-sided game approach. See: Weiner, "War Gaming Methodology." p 19.

⁴⁹⁵ In some games, teams worked together in a single room. This was generally done only for follow-on variant games, in which players were already familiar with the contours of the other side's intentions and capabilities, and the game focused explicitly on the outcomes of alternative choices at key nodes. In these games, the advantage of faster game play was seen as worth the tradeoff in studying the impact of hidden information on player choices, since those issues had been explored in previous games. See: "War Gaming: Two Methods Used in Sierra." p 13

⁴⁹⁶ Ibid. pp 9-10

⁴⁹⁷ "War Gaming Methodology." pp 55-57

⁴⁹⁸ Ibid. p 19

⁴⁹⁹ Ibid. pp 60-63

RAND work.⁵⁰⁰ In contrast, emerging issues and planning factors that emphasized human judgement based on experience were adjudicated based on the judgement of umpires in conjunction with player expertise.⁵⁰¹ As a result, the Sierra games featured design elements used in earlier rigid and umpired game designs. A more detailed consideration of the design elements of the Project Sierra Jordan game series is included in Chapter 5.

More innovative than the game format itself was the use of a series of games to explore a single issue. First, the project was able to look at limited wars in multiple settings, to better understand similarities and differences between these localized conflicts. Within each setting, multiple games—known as a series—were run in which the strength of the two sides varied.⁵⁰² The general principle was to give a major advantage to red in the first game by forcing the U.S. team to operate under a great many constraints, including on nuclear use, that were removed in later iterations of the game to explore different combinations of red and blue authorities.⁵⁰³ “Variant” games were also played that started at a critical decision point from a previous game. In effect, the variant game became a counterfactual—what would have happened had control selected the other possible decisions?⁵⁰⁴ By varying both the setting of the conflict, and the weapons available, the team was able to generate patterns about how access to and use of atomic and nuclear weapons impacted to conflict to better understand how these weapons systems might reshape the nature of limited warfare.⁵⁰⁵

Gaming and Gaining a Deeper Understanding: 1959-1970

Over the 1960s, RAND’s gaming efforts expanded beyond initial efforts at system exploration to begin to refine understandings of problem. One major leg of research continued to explore political-military aspects of crisis and conflict to continue to build out an understanding of how these systems would operate in a wide range of crisis. A second line of research focused on considering force structure decisionmaking under alternative conditions—that is, to deepen and refine the researchers understanding of the problem. Finally, games were used build out models, which could serve as the basis for computerized modeling and simulation efforts.

Continuation of the Tradition of Political-Military Crisis Games for Systems Exploration

Efforts to study political-military decisionmaking continued to use and refine the umpired game designs developed in the previous decade to further explore the maturing political and

⁵⁰⁰ Paxson. "The Sierra Project -- a Study of Limited Wars." p 11

⁵⁰¹ Weiner. "War Gaming: Two Methods Used in Sierra." p 9

⁵⁰² "War Gaming Methodology." pp 72-73

⁵⁰³ Paxson. "The Sierra Project -- a Study of Limited Wars." p 7

⁵⁰⁴ Weiner, "War Gaming Methodology." pp 70-71

⁵⁰⁵ Paxson. "The Sierra Project -- a Study of Limited Wars." pp 19-20

military systems for managing crises with nuclear weapons. Two substantial lines of effort are worthy of note. The first is the continuation of the exploration of limited war, and specifically efforts to build a better understanding of the political-military context of these limited conflicts. The second line focuses more narrowly on lines of communication and control between political and military decision makers in a crisis. Both efforts used games to further flesh out researcher's understanding of the policy problems of the day by eliciting how players understood the problem that confounded them, the available options, and the effects of their own and adversaries decisionmaking on the problem system.

Following the Project Sierra work, two additional projects, Back Stop and Red Wood, took up the question of limited war using games. Project Back Stop, for which more documentation is accessible, focused on the political-military context of limited war.⁵⁰⁶ In many ways, these efforts continued in the model of the Project Sierra games. They often used both scenario development and gaming to build out a narrative of both the political road to war and military trajectory of the fighting. The initial games focused on Iran.⁵⁰⁷ The team was quite explicit that the outcome of the war was, to a large extent, predetermined by the restrictions placed on players. Instead, the team focused on the problems that arose as a result of player decisions.⁵⁰⁸ These problems then became the subject of "collateral studies" that could dig into specific issues highlighted in the game to develop more actionable insights than the general knowledge offered by the context development efforts.⁵⁰⁹ This process is shown in Figure 9.5. As a result, over time the focus appears to have shifted from scenario-centric game designs to the use of standalone scenarios.⁵¹⁰

Efforts in the early 1960s to study crisis decision making—and particularly the ability of national political leadership to communicate direction to forces during a rapidly escalating crisis⁵¹¹--returned to manual games that provided open-ended decision-making options for players to consider. The most notable of these efforts featured a series of games, each focused on a different European crisis. Players took on the role of two groups of opposing national-level decisionmakers and were provided with a detailed scenario of a European crisis. After reviewing the scenario, each team met to determine their plan to respond to the situation. Both teams then met and presented their plans, including contingencies. To support the goal of not limiting player options, little in the way of structure for these reports was provided. On the basis of both teams'

⁵⁰⁶ Harvey A. DeWeerd, T. E. Greene, and F. M. Sallagar. "A Report on the Rand Limited War Program: A Project Back Stop Briefing." (Santa Monica, CA: RAND Corporation, D-6354-PR, 1958).

⁵⁰⁷ Ibid. p 8

⁵⁰⁸ Ibid. p 13

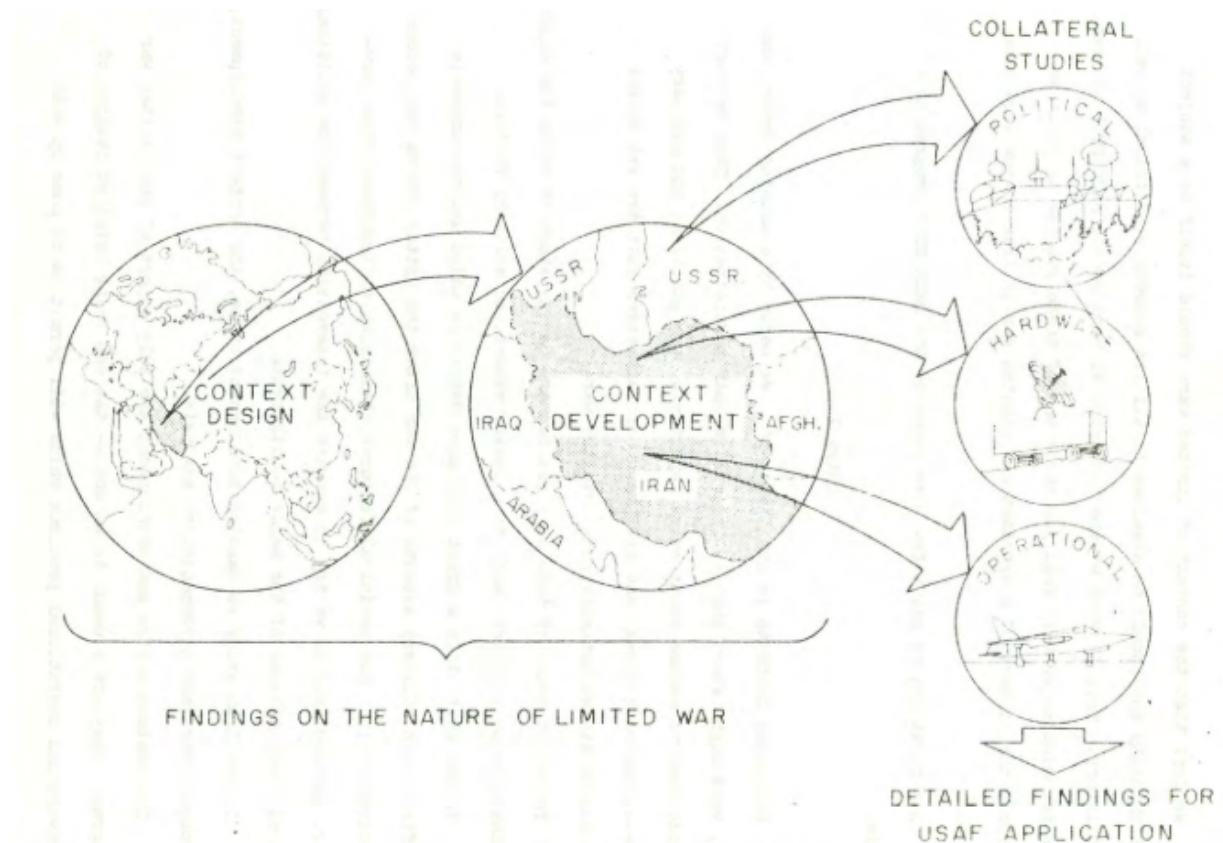
⁵⁰⁹ Ibid. pp 4-7

⁵¹⁰ DeWeerd, "A Contextual Approach to Scenario Construction."

⁵¹¹ Harvey A. Averch and Marvin M. Lavin, "Dilemmas in the Politico-Military Conduct of Escalating Crises," (Santa Monica, CA: RAND Corporation, P-3205, 1965). p 1

decisions, the control team would then use expert judgement to determine what actions had occurred. Then the teams divided again to consider their response to the new situation.⁵¹²

Figure 9.5: Project Back Stop Study Process



Source: Harvey A. DeWeerd, T. E. Greene, and F. M. Sallagar. "A Report on the Rand Limited War Program: A Project Back Stop Briefing." (Santa Monica, CA: RAND Corporation, D-6354-PR, 1958) p 6.

Two areas of game design development stand out from this effort: the organization of the blue team and the development of scenarios. First, the game design team devoted considerable attention to the organization of the teams (particularly the blue team representing the United States) and regulated communications in order to better reproduce command and control structures of interest. Over subsequent games, more structure was applied to the teams (that is, specific responsibility for political and military decisionmaking with a hierarchy was established to make player roles clearer) and the direct communications between the red and blue team were limited to written channels to allow for more realistic levels of partial and incorrect information

⁵¹² "Simulation of Decisionmaking in Crises : Three Manual Gaming Experiments," (Santa Monica, CA: RAND Corporation, RM-4202-PR, 1964). p 9

to come into play.⁵¹³ While the approaches to structuring within-team interaction were not novel, the care with which they were employed to further the research objective is worth noting.

Second was the extensive thought put into crafting the scenarios that underpinned the games. In contrast to previous games for which the scenario was generally a matter of 10-15 pages, the scenarios for these crisis games run to more than 30 pages of narrative.⁵¹⁴ These scenarios were specifically crafted to meet a set of predetermined conditions of interest that provide guidelines to the player.⁵¹⁵ In particular, the research team felt there was a scarcity of plausible limited war scenarios outside of the often-studied central European case and sought to consider alternative cases that would still be of considerable interest to Air Force missions.⁵¹⁶ As a result, the games considered scenarios for a limited war in Europe in which neither the United States nor the Soviet Union is an instigator,⁵¹⁷ or in which geographically proximate countries had declared positions of neutrality.⁵¹⁸ Greater care was taken to ensure that the scenario had a sensible “scenario past” that connected the present day to the starting point of the game—in other words, in contrast to previous games that were relatively willing to stipulate starting conditions without much reference to current conditions, these games provided traceability so that “nothing should be included in the political-military scenario dealing with the future which differs from the present without giving some explanation as to what happened in the interim, or what caused the change.”⁵¹⁹ These efforts meaningfully advanced the art of scenario construction and served as the foundation of later, scenario-centered research.

Exploring Alternative Conditions’ Impact on Force Structure

A second line of work build on the structured approaches to gaming first developed for system exploration to enable them to more closely examine policy problems under alternative conditions. The Strategic and Force Evaluation (SAFE) game was designed to allow teams to consider potential future investments in the force and subsequent posturing decisions. The game was used to understand what characteristics of the strategic and budgetary environment would drive actual differences in force structure, posture, and warfighting strategy. To this end the game was run six times, each with a different stated strategy and budget level for the United

⁵¹³ Ibid. p 13

⁵¹⁴ For example see: Harvey A. DeWeerd. "A Scenario for a Limited War in the Northern Flank of Nato, 1966." (Santa Monica, CA: RAND Corporation, D-12077-PR, 1964).

⁵¹⁵ "Political-Military Scenarios," (Santa Monica, CA: RAND Corporation, P-3535, 1967). p 6

⁵¹⁶ "A Scenario for a Limited War in the Northern Flank of Nato, 1966." p 2

⁵¹⁷ Marvin M. Lavin. "Blue Military Moves in the Southern Flank Crisis and Limited War Game --1968." (Santa Monica, CA: RAND Corporation, D-12489-PR, 1964).

⁵¹⁸ DeWeerd. "A Scenario for a Limited War in the Northern Flank of Nato, 1966." p 3

⁵¹⁹ "Political-Military Scenarios." p 7. This point is further elaborated in: "A Contextual Approach to Scenario Construction."

States and Soviet Union. This allowed runs to be compared to see how game play differed under different strategic conditions. Game play featured two teams representing the United States and Soviet Union, each of which was asked to look forward in two-year increments to allocate a budget, make investments in new systems to procure, posture their force geographically, and define (in broad strokes) the concept of operation if the forces go to war.⁵²⁰ Figure 9.6 illustrates some of the range of adjudication tools used to support the game. This game is considered in more detail in Chapter 6.

Beyond the initial plays of SAFE, there was interest in using the gaming system for a range of other purposes. First, SAFE was used to support Air Force academic institutions, reframing the original research product for educational ends.⁵²¹ Second, towards the end of the series, it was proposed that the game system could also be utilized for additional research considering the impact of arms control limits on decision-making as well as considering a wider range of red strategic behavior.⁵²² However, there is no evidence that this proposal was taken up. Finally, the game records were re-analyzed in the mid-1970s to consider whether the games had offered any predictive power. The researchers, who had served as part of the original team, noted the serious shortcomings of the games as harbingers of the future due to their (anticipated) failure to consider institutional pressures and domestic constraints.⁵²³ These efforts show how RAND was attempting to expand the utility of games but often bumping up against the limits of a design built for one purpose to support research that needed games to produce different types of information.

Another effort focused on the same themes was Project XRAY—a multi-year effort beginning in 1966 that sought to define force posture options for the execution of a range of flexible deterrence responses rather than merely responding with overwhelming force. The game was run several times with the same basic scenario parameters and process. Three teams—blue, red, and yellow—were each tasked to develop a 10-year strategy, while a green team represented the subordinates staffs, domestic opinion and the perspectives of the allies of the three major powers. Within each team, more detailed roles were assigned to represent different political and military interests.⁵²⁴

⁵²⁰ Brown and Paxson, "A Retrospective Look at Some Strategy and Force Evaluation Games." pp 1-2

⁵²¹ Edwin W. Paxson, "War Gaming," in *Military Operations Research*, ed. Bernard O. Koopman (Operations Research Society of America, 1963). p 17

⁵²² Harvey A. Averch, Allen R. Ferguson, and William M. Jones. "Europe, Sac and Safe: Some Issues for the Next Decade." (Santa Monica, CA: RAND Corporation, D-10895-PR, 1963).

⁵²³ Brown and Paxson, "A Retrospective Look at Some Strategy and Force Evaluation Games."

⁵²⁴ Edwin W. Paxson, "Computers and National Security," (Santa Monica, CA: RAND Corporation, P-4728, 1972). pp 27-29.

Figure 9.6: Adjudication of SAFE Game play



Source: RAND Corporation Archives

Much as in the SAFE games, players first developed a force structure and posture that would implement the strategy, subject to year-specific budget constraints, before being confronted with a tailored crisis to which the three teams responded and reacted to the actions of the other teams.⁵²⁵ In XRAY the primary focus of the first stage of play was the balance between offensive, defensive and global surveillance systems. This placed more emphasis on the role of warning in nuclear conflict than the earlier SAFE games. The second, crisis response, phase of game play also differed in that the players engaged in multi-sided free play with considerable focus on political, as well as military, actions enabled by the selected posture.⁵²⁶ This approach was not an unmitigated success—for example, players complained that the assigned national-level roles were somewhat at odds with the theater-level information provided.⁵²⁷

The mechanics of both phases of play featured innovation in the use of computers to support wargames. In the planning stage, computer support allowed for more realistic modeling of posture costs over time. For example, the second series of XRAY games used a costing model that enabled players to determine the relative price of different force structure packages quickly enough to inform ongoing debate⁵²⁸—in effect, computerizing the bookkeeping that required pre-calculated values in the earlier SAFE game. This innovation allowed for more accurate depictions of the costs of phasing systems in and out of use. The second, crisis response, stage of play also featured the use of computer. For example, play featured a rich scenario but the XRAY game's consideration of crisis response was also informed by a series of computer models that could dynamically update planning factors for participants. For example, the SIMSCRIPT Program for Operational Development (SPOD) was developed to support the later XRAY games with detailed information about the flight times between airfields that could replace more traditional lookup tables.⁵²⁹

Perhaps a more profound change in game play was the role of technology in mid-game communication. The game featured teams that worked from their home agency, communicating via commercial tele-type, and using a time-share computer system for calculations.⁵³⁰ This allowed teams from seven different agencies to participate in game play over the course of a full month while keeping the identities of opponets hidden from fellow players.⁵³¹ Researchers noted

⁵²⁵ H. G. Massey. "The Xray Force Planning Cost Model." (Santa Monica, CA: RAND Corporation, D-19847-ARPA, 1970). p 3

⁵²⁶ Paxson, "Computers and National Security." pp 27-29.

⁵²⁷ Gaylord M. Northrop. "A Resume of Red Actions in War Game X-Ray." (Santa Monica, CA: RAND Corporation, D-15247-ARPA, 1966). p 1

⁵²⁸ Massey. "The Xray Force Planning Cost Model." p v

⁵²⁹ David L. Arnold. "Simscript Program for Operational Deployment (Spod)." (Santa Monica, CA: RAND Corporation, D-20393-ARPA, 1970). p 1

⁵³⁰ Paxson, "Computers and National Security." pp 25, 29

⁵³¹ Ibid. p 25

the growing potential for this style of game play, thanks to contemporary DoD investments to networked computers.⁵³² In retrospect, we can see this effort as an early utilization of the network that would become the internet to enable wargaming.

However, the increasing role of technology to support rapid calculations and communications did not indicate a shift in the type of information that these games were intended to produce. The focus of researcher attention was still on the policy problem of making sense of how players understood the choices they were presented with. These games were not much concerned with evaluating the “goodness” of player decisions.

Towards Evaluation and Away from National Security Policy Analysis Games

In contrast to efforts like the XRAY games, other RAND efforts were using the improving capabilities of computers for a different purpose: evaluating the potential performance of specific systems and operations. While computerized calculation had supported past games, the slow speed, complexity of programming, and need to batch calculations prevented computers from effectively supporting game execution.⁵³³ Newer digital computers allowed for multi-console, time-shared systems providing additional computational power and allowing a broader range of researchers to effectively use the system.⁵³⁴ At the same time, other analytic work at RAND provided more elaborate quantitative models of key operational interactions. These trends enabled an increasing portion of the game rules to be translated into computerized models to support the adjudication of player choices.⁵³⁵ Greater capability to quantitatively model weapons performance also placed more emphasis on building and analyzing technical aspects of warfare, instead of human decisionmaking.⁵³⁶ As a result, over time research became focused on using computerized simulations to evaluate combat outcomes from pre-planned engagements, rather than wargames.

Often, RAND researchers used systems exploration games to build up a model of the problem which could be used as the starting point for computerized simulation development. For example, the Tactical Air Study used a game to provide a baseline set of environmental information and to identify topics for later stages of the study. First, operational gaming was used as a means to build “*representative combat environments* in which the performance [of

⁵³² Ibid. p 31

⁵³³ Gaylord M. Northrop, "Use of Multiple on-Line, Time-Shared Computer Consoles in Simulation and Gaming," (Santa Monica, CA: RAND Corporation, P-3606, 1967). p 1

⁵³⁴ Ibid. p 3

⁵³⁵ The move towards greater automation of wargames is sometimes tied to broader tensions between civilian and military roles in post-WWII military planning, see: Ghamari-Tabrizi, "Simulating the Unthinkable: Gaming Future War in the 1950s and 1960s."

⁵³⁶ Milton G. Weiner, "Trends in Military Gaming," (Santa Monica, CA: RAND, P-4173, 1969). p 4

different modeled systems] can be examined in a consistent and integrated manner,”⁵³⁷ including force disposition, strategy of non-air components, weather, terrain, logistics, and political considerations drawn from a Korean limited war scenario set in 1965. Drawing on this common touchstone ensured that the individual models of different elements of tactical air would all integrate into a common, operationally relevant picture.⁵³⁸ Second, the games helped identify issues worth greater attention in the other stages of analysis. For example, the Korea games raised questions about the scale of capabilities lost should American forces not be able to use Japanese bases and the cost of the deployment of an air assault division by ground forces. Both issues were examined using quantitative tools using the models developed in other aspects of the study.⁵³⁹ The human engagement in the games dwindled over time, moving the approach from gaming to modeling and simulation as the focus of research shifted from establishing the relationships between elements of the systems to evaluating the effectiveness of military options.

Games in Eclipse at RAND: 1970-1990

Across the Department of Defense, the early 1970’s represented a low point in the use of gaming. The failures of Vietnam raised questions about the effectiveness of analysis to support military decisionmaking,⁵⁴⁰ and limited budgets shrank available resources to support research. While wargaming began to recover at the institutions of military education, training programs, and operational units, less gaming occurred in strategic and joint operational research in the 1970s.⁵⁴¹ Within many research institutions, the focus was on fully computerized systems. While some, such as the Naval War College’s enhanced Naval Wargame System, were used to support human-in-the-loop play,⁵⁴² many others prioritized computer-only play. RAND writers during this period go out of their way to stress the dominance of fully computerized modeling and simulation efforts in research. For example, a 1970 survey of over 100 DoD research and analysis organizations revealed that only 3% of the modeling, simulation, and gaming output were manual games and 8% human in the loop simulations—vastly outnumbered by fully computerized systems. The authors note that: “Man-machine gaming and simulation was, at its peak of activity [1965-1967] probably oversold. Currently [in 1972] it is undersold, if not totally neglected.”⁵⁴³ A similar observation was made almost 10 years later, when a popular account of

⁵³⁷ "Rand Briefings to the Air Force Advisory Group 21 October 1963." (Santa Monica, CA: RAND Corporation, AR-104, 1963). p 36

⁵³⁸ Ibid. p 36

⁵³⁹ Ibid. p 38

⁵⁴⁰ Caffrey, *On Wargaming: How Wargames Have Shaped History and How They May Shape the Future*. pp 86-87

⁵⁴¹ Ibid. pp 87-92

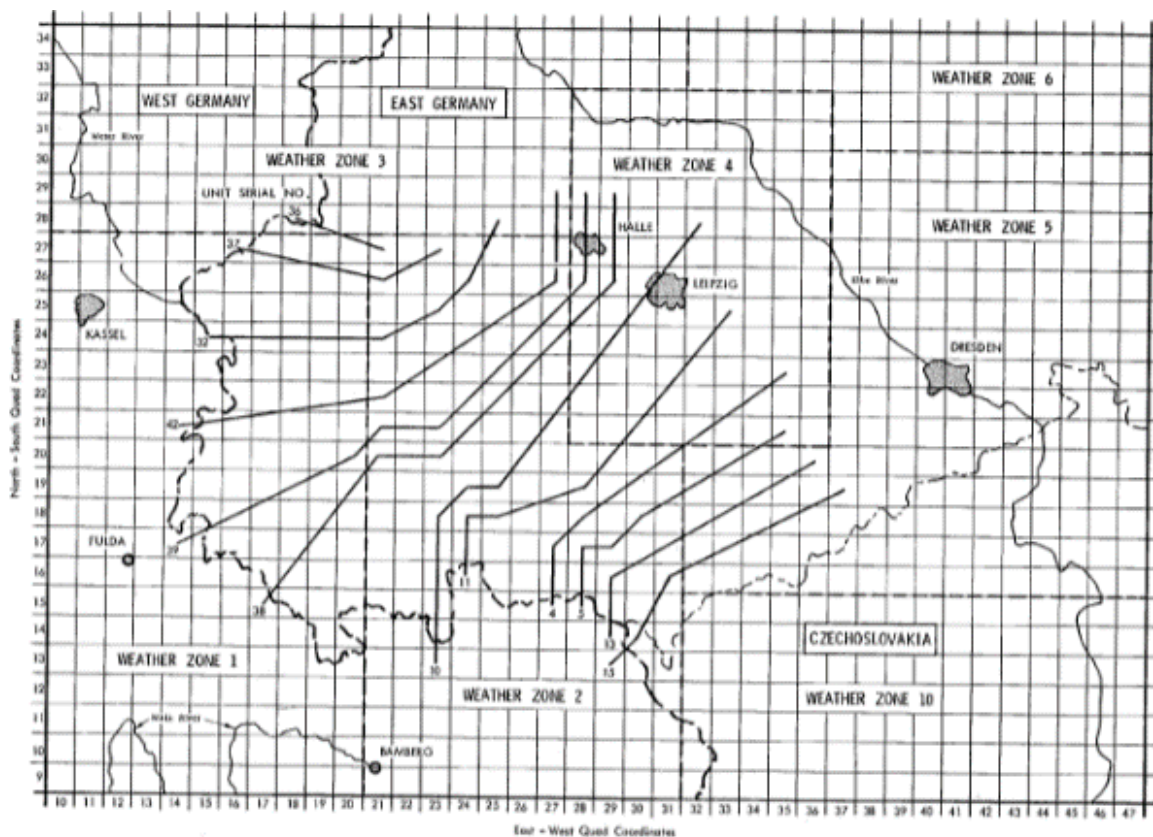
⁵⁴² Ibid. pp 101-103

⁵⁴³ Shubik and Brewer, "Models, Simulations, and Games--a Survey." p 5

wargaming noted that: “the word *game* was out of favor by the early 1980s, when *simulation* and *modeling* became the preferred terms...because they believe that *simulation* more accurately describes [1980s] computerized studies of conflict.”⁵⁴⁴

RANDs research of the 1970s favored the use of computerized modeling and simulation efforts to conduct evaluations. For example, efforts to study air-ground tactics in NATO contingencies in the early 1970s used the TOTEM land warfare and TALLY air combat systems to examine the value of different equipment, postures and tactics. While the terms “war game” and “player” were still used,⁵⁴⁵ human decisionmaking was limited to initial planning, illustrated in Figure 9.7.⁵⁴⁶ While a human was tasked with translating these plans into inputs, this process

Figure 9.7: Sample NATO Planning Map for TOTEM Ground Combat Simulation



Source: Kathleen Harris and Louis H. Wegner, "Tactical Airpower in Nato Contingencies : A Joint Air-Battle/Ground-Battle Model (Tally/Totem)," (Santa Monica, CA: RAND Corporation, R-1194-PR, 1974) p 49.

⁵⁴⁴ Allen, *War Games : The Secret World of the Creators, Players, and Policy Makers Rehearsing World War Iii Today*. p 7

⁵⁴⁵ Kathleen Harris and Louis H. Wegner, "Tactical Airpower in Nato Contingencies : A Joint Air-Battle/Ground-Battle Model (Tally/Totem)," (Santa Monica, CA: RAND Corporation, R-1194-PR, 1974). pp v-vii

⁵⁴⁶ Ibid. pp 47-50

did not stress player decisionmaking or reactions to model outcomes—in effect the system appears to have run on a set path after the initial strategy was selected.⁵⁴⁷ What's more, the focus of analysis was not on these initial player decisions, or the reactions of players to the outcomes of their choices, but rather on the modeling and simulations analysis of technical performance. In other words, regardless of terminology, the focus of research had swung to the computerized simulation, losing the focus on human decisionmaking at the heart of a proper wargame.

While broader defense gaming gained ground in the 1980's,⁵⁴⁸ the dominance of computerized solutions at RAND continued. For example, the RAND Strategy Assessment System (RSAS),⁵⁴⁹ designed to support the Office of Net Assessment⁵⁵⁰ used human-centric gaming very early in its development by running several games and observing those conducted elsewhere,⁵⁵¹ but the project focus was always on developing a fully automated tool without an actual human making decisions.⁵⁵² The concept was that “a war gaming framework would help overcome the otherwise sterile scenarios used in strategic force analysis, while coupling such an approach with analytic models would lend a degree of rigor... [to] gain control over the variables by automating the entire war game.”⁵⁵³ As a result, while the structure of the system closely resembled that of a traditional two sided game, as shown in Figure 9.8, the absence of human players separated the RSAS approach from a traditional game.

While the overall effort was not a game as the term is used in this text, games with human players were occasionally used in specific stages of the project. For example, early work on a ground warfare model, S-Land, used games as a way to prototype rules, before transitioning to a fully computerized approach.⁵⁵⁴ Later on, games were used to explore nuclear play, since players tended to add to and refine the menu of options beyond what was originally included in the model. In particular, games served to falsify options, and generate factors that should be added to

⁵⁴⁷ Ibid. pp 50-56

⁵⁴⁸ Caffrey, *On Wargaming: How Wargames Have Shaped History and How They May Shape the Future*. pp 97-110

⁵⁴⁹ Early stages of the effort referred to the program as the RAND Corporation's Strategic Assessment Center or RSAC. See: Bruce W. Bennett and Paul K. Davis, "The Role of Automated War Gaming in Strategic Analysis," (Santa Monica, CA: RAND Corporation, P-7053, 1984). p 1

⁵⁵⁰ Paul K. Davis and James A. Winnefeld, "The Rand Strategy Assessment Center : An Overview and Interim Conclusions About Utility and Development Options," (Santa Monica, CA: RAND Corporation, R-2945-DNA, 1983). p iii

⁵⁵¹ Interview with Paul Davis, Santa Monica, CA, May 2018.

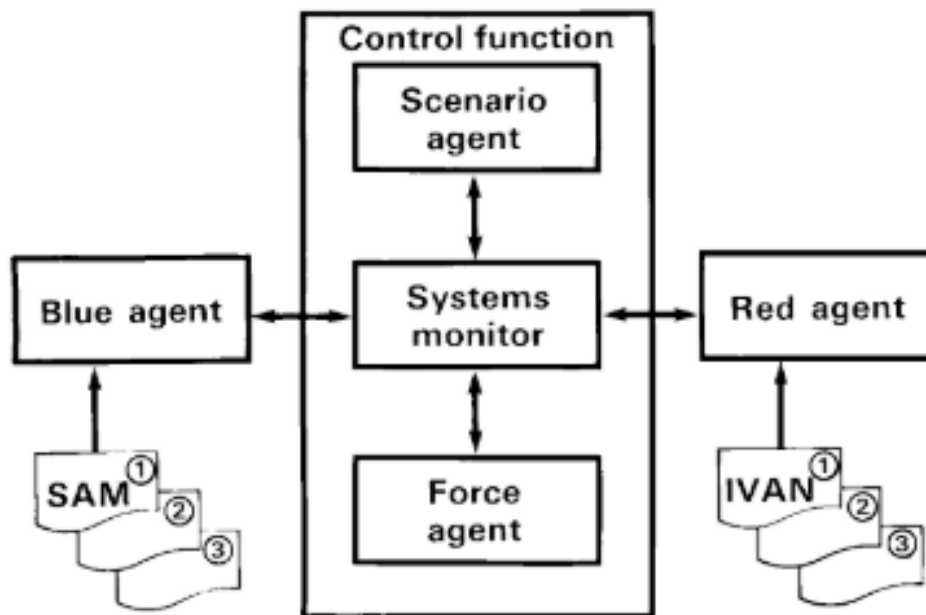
⁵⁵² For a general description of the RSAC program, see: Allen, *War Games : The Secret World of the Creators, Players, and Policy Makers Rehearsing World War Iii Today*.

⁵⁵³ Bennett and Davis, "The Role of Automated War Gaming in Strategic Analysis." p 1.

⁵⁵⁴ Interview with Paul Davis, Santa Monica, CA, May 2018.

the model.⁵⁵⁵ In all cases, the games were designed explicitly to elicit expert understanding to inform the overall modeling and simulation effort.

Figure 9.8. RSAS structure



Source: Bruce W. Bennett and Paul K. Davis, "The Role of Automated War Gaming in Strategic Analysis," (Santa Monica, CA: RAND, 1984) p 3.

Despite the dominance of computer-dominated approaches in this period, some human-centric gaming persisted, particularly on political-military topics. For example, political-military games in the late 1970s considered the potential for US intervention to manage a crisis in the Persian Gulf. The game design featured three teams—a political and military U.S. team, a USSR red team, and a control team that managed the behavior of other key actors (who were able to communicate amongst themselves in writing via fairly unstructured play.)⁵⁵⁶ Later work proposed leveraging computer assistance in largely manual games studying the potential for de-escalation between the U.S. and Soviet Union, though it is not clear that such a game was ever run.⁵⁵⁷ Furthermore, the descriptions of games in general texts on the subject suggest additional games occurred at RAND during these decades. For example, in the mid-1980s noted RAND wargamer William Jones penned several documents on the value of “free-form” political-

⁵⁵⁵ *ibid*

⁵⁵⁶ Carl H. Builder and William M. Jones, "Gaming a Persian Gulf Contingency," (Santa Monica, CA: RAND Corporation, N-1022-AF, 1979). pp 1-2

⁵⁵⁷ James P. Kahan, William M. Jones, and Richard E. Darilek, "A Design for War Prevention Games," (Santa Monica, CA: RAND Corporation, N-2285-RC, 1985). p v

military exercises, the preface and acknowledgements of which list colleagues and games, the output of which is absent from the archive.⁵⁵⁸ Similarly, the preface of a 1991 reissuing of a set of internal papers on the utility of crisis games mentions that the publication was prompted by on-going work using the approach.⁵⁵⁹ These references suggest that even during this long period of decline, manual gaming was still practiced at RAND even if they no longer had the same frequency or prominence they had at their peak. Further, the discussion of gaming within these general works suggest that these games likely continued in the tradition of using games primarily as a tool for systems exploration, suggesting that their role in the research process was relatively unchanged

Repeating the Cycle of Boom and Bust while Expanding Scope: 1990-2014

The fall of the USSR and rise of revolutionary networked technology heralded a resurgence of gaming in the 1990s. Much of the planning for the first Gulf War relied on wargaming, leveraging both past games and efforts specifically designed to support planning for the operation, drawing new attention to the operational value of the tool to develop and test out courses of action.⁵⁶⁰ The result was a massive increase in spending by the U.S. military on games,⁵⁶¹ much of which went to large, service-led "Title 10" wargames that would be used to evaluate the value of new, technologically advanced equipment. RAND's gaming output generally followed two trends. The first continued the tradition of political-military games for systems exploration of emerging challenges. The second leveraged RAND's previous work on computer simulation to support adjudication and analysis of the new Title 10 games.

Gaming to Understand the Post-Soviet Policy System

Perhaps the most iconic of the post-Cold War system exploration games was RAND's Day After games. The first series (early 1990s) were manual games originally designed to look at strategic proliferation issues raised by post-Soviet stockpiles. The dissolution of the Soviet Union and resulting shifts in the international balance of power and technological trends raised critical concerns about the potential for the proliferation of nuclear weapons.⁵⁶² The games were designed to bring together a group which would be:

⁵⁵⁸ See: Jones, "On Free-Form Gaming." p ix and "On the Adapting of Political-Military Games for Various Purposes," (Santa Monica, CA: RAND, N-2413-AF/A, 1986). p iii

⁵⁵⁹ Levine, Schelling, and Jones, "Crisis Games 27 Years Later : Plus C'est Deja Vu." p iii

⁵⁶⁰ Caffrey, *On Wargaming: How Wargames Have Shaped History and How They May Shape the Future*. pp 129-137

⁵⁶¹ Ibid. p 137

⁵⁶² Marc Dean Millot, Roger C. Molander, and Peter A. Wilson, ""The Day After..." Study : Nuclear Proliferation in the Post-Cold War World. Volume I, Summary Report," (Santa Monica, CA: RAND Corporation, MR-266-AF, 1993). p 3

broadly representative of the people who directly support the development of U.S. defense policy and who are likely to participate in the future development, shaping, and “marketing” of policy options to cope with nuclear proliferation and its consequences⁵⁶³

Since these issues were fairly new, and many relevant perspectives were not reflected in the literature, an event offered the opportunity to bring diverse perspectives together.⁵⁶⁴ A core focus of the game design is the scenario, which plays out future events into a future timeline that moves from the present day forward to a crisis in a “future history”, then presents a two-step crisis so participants can consider both the “day of” the crisis and the “day after”⁵⁶⁵. The final step of the game returns to the present day to consider available mitigations and opportunities.⁵⁶⁶ This process is visualized in Figure 9.9.

The initial series of “Day after...” games consisted of four scenarios looking at proliferation in the former USSR, the Middle East, Korea, and South Asia.⁵⁶⁷ The approach was also used to support a 1993 collaboration between a South Korean think tank and the U.S. Center for Army Analysis to develop strategies for negotiating with North Korea on arms control issues.⁵⁶⁸ Later games looked at issues such as strategic information warfare in 1995⁵⁶⁹ and the role of electronic commerce on money laundering. Both had the goal of better defining potential emerging threats to the U.S. by describing the key features of the threat and exploring potential consequences of an incident to the U.S.⁵⁷⁰ The strategic information warfare game designers considered four alternative scenarios, including vignettes focused on Chinese aggression toward Taiwan and instability in Moscow, before finally settling on a Persian Gulf scenario to explore the impact on U.S. military response to an escalating crisis of an attack on critical information infrastructure.⁵⁷¹ The cybercrime study focused attention on potential trajectories of the future of electronic payment systems, and their potential vulnerabilities.⁵⁷² Again, the focus of these exercises were

⁵⁶³ Ibid. p 5

⁵⁶⁴ Ibid. p 6

⁵⁶⁵ Ibid. p 7

⁵⁶⁶ David Mussington, "The "Day after" Methodology and National Security Analysis," in *New Challenges, New Tools for Defense Decisionmaking* (Santa Monica, CA: RAND Corporation, 2003). pp 324-326

⁵⁶⁷ Millot, Molander, and Wilson, ""The Day After..." Study : Nuclear Proliferation in the Post-Cold War World. Volume I, Summary Report." p 8

⁵⁶⁸ Richard E. Darilek and James C. Wendt, "Korean Arms Control : Political-Military Strategies, Studies, and Games," (Santa Monica, CA: RAND Corporation, MR-489-A, 1994). p vii

⁵⁶⁹ Mussington, "The "Day after" Methodology and National Security Analysis." p 327

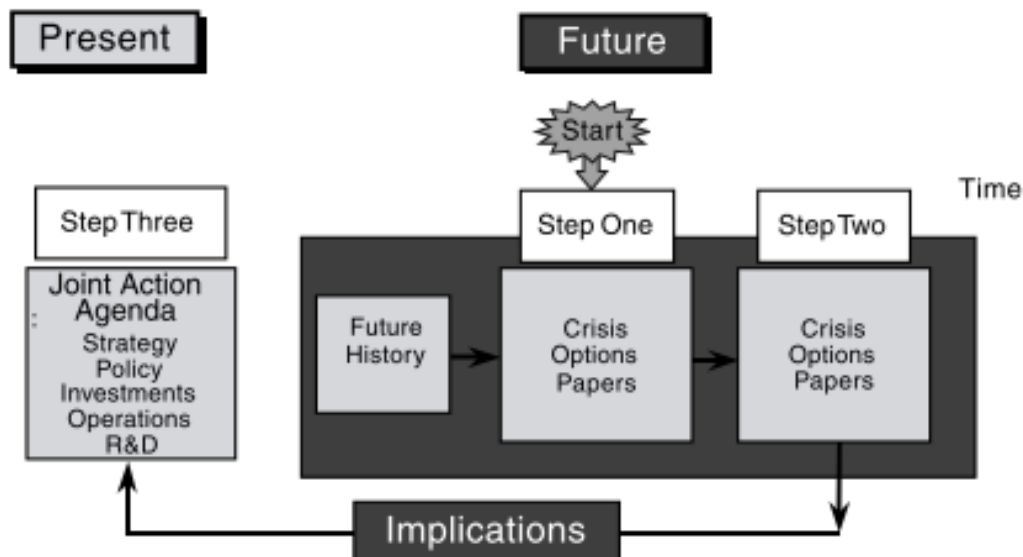
⁵⁷⁰ Ibid. pp 327 and 331-332

⁵⁷¹ Ibid. pp 328-329

⁵⁷² Ibid. p 332

on emerging security threats deriving from the new, post-Cold War asymmetric structure of the security environment.⁵⁷³

Figure 9.9: General Scheme of a Day After Game



Source: David Mussington, "The "Day after" Methodology and National Security Analysis," in *New Challenges, New Tools for Defense Decisionmaking*, (Santa Monica, CA: RAND, 2003) p 325

Supporting Defense Gaming: A New Model

There was one notable change from the previous eras of highly active RAND gaming. Whereas RAND gaming work heretofore almost always focused on RAND- designed games often played by RAND researchers, in the 1990s far more work then in previous periods appears to have been in partnership with DoD institutions. In particular, RAND provided analytical support to games designed and executed by the services—in effect serving as outside advisors and assessors of broader efforts. Efforts such as Army After Next (1996-1999) run by the U.S. Army's Training and Doctrine Command (TRADOC) were large, service-led initiatives to develop and assess new force structures that took full advantage of maturing technology.⁵⁷⁴ RAND supported these efforts in two ways. First, they observed the game development and execution process, and analysis of the games including identification of key trends, assumptions that might impact the credibility of the recommendations, and issues for further analysis.⁵⁷⁵ Second, RAND analysts conducted additional analysis, often using in-game decisions as the

⁵⁷³ Ibid. p 328

⁵⁷⁴ Walter L. Perry and Marc Dean Millot, "Issues from the 1997 Army after Next Winter Wargame," (Santa Monica, CA: RAND Corporation, MR-988, 1998). pp 1-3

⁵⁷⁵ For example, see: *ibid.*

bases for detailed technical analysis using modeling and simulation tools.⁵⁷⁶ However, without direct ownership over the game design, RAND's involvement in the details of improving methods for gaming was fairly limited.

These efforts were altered by the need to support post-September 11 operations in Iraq and Afghanistan. The need to support counterterrorism and counterinsurgencies operations tended to lean on social science-informed approaches, rather than gaming efforts. Across DoD some wargaming efforts were halted, leaving a smaller number of efforts to continue, including the service's large Title Ten games.⁵⁷⁷ At the same time, public controversy over the "fixed" results of wargame run as an early element of Joint Forces Command's Millennium Challenge raised questions about the analytic validity of major force-on-force wargames.⁵⁷⁸ Taken together, the level of wargaming and its profile for defense decisionmaking was in decline.

In response, at RAND some gaming capability remained active but with a reduced profile. This capability included providing support to large service-specific games as outside observer-analysts, continuing lines of work begun in the 1990s.⁵⁷⁹ Occasionally, games involved the RAND team more directly, including one-off policy games using the Day After format⁵⁸⁰ and the 2007 Pacific Vision operational game which drew important attention to the issue of base defense.⁵⁸¹ However, such research was the exception rather than the rule. The purpose of the work tended to follow the pattern of the 1990s with support to outside games generally focusing on assessing the credibility of game design based on other research, while RAND-built games focused on system exploration.

The Current Era of RAND Gaming: 2014-2019

After the most recent period of drought, the emergence of new strategic priorities and corresponding uncertainty about the future of competition has fueled a new generation of RAND gaming, which has seen the expansion of RAND gaming to more fully cover games aligning with all four archetypes. Not surprising given this range of topics, a wide range of formats is also

⁵⁷⁶ John Matsumura et al., "The Army after Next : Exploring New Concepts and Technologies for the Light Battle Force," (Santa Monica, CA: RAND Corporation, DB-258, 1999). p vi

⁵⁷⁷ Caffrey, *On Wargaming: How Wargames Have Shaped History and How They May Shape the Future*. pp 180-182

⁵⁷⁸ Ibid. pp 181-182 and Zenko, "Millennium Challenge: The Real Story of a Corrupted Military Exercise and Its Legacy."

⁵⁷⁹ These included both Air Force (for example, see: Jennifer D. P. Moroney, "Assessing the U.S. Air Force Unified Engagement Building Partnerships Seminars," (Santa Monica, CA: RAND Corporation, DB-605-AF, 2011).) and continued support to Army games.

⁵⁸⁰ For example: Roger C. Molander et al., "The Day After... In Jerusalem : A Strategic Planning Exercise on the Path to Achieving Peace in the Middle East," (Santa Monica, CA: RAND Corporation, CF-271, 2009).

⁵⁸¹ Richard Halloran, "Pacaf's "Vision" Thing," *Air Force Magazine*, 2009. Interview with David A. Ochmanek, Arlington, VA, October 2018.

in use, many calling back to historical RAND formats. The most high profile have used hex and counter board games to explore theater operations against emerging potential adversaries.⁵⁸² However, other structured approaches are being used to consider areas such as explorations of operations short of armed conflict⁵⁸³ and responses to cyber-attacks,⁵⁸⁴ force posture planning under alternative condition,⁵⁸⁵ innovation in tactical equipment⁵⁸⁶ and command and control systems,⁵⁸⁷ and initial evaluation of security force assistance packages⁵⁸⁸ and command and control constructs.⁵⁸⁹ These trends reflect the current questions being asked by defense leadership but also renewed interest in gaming on the part of senior leaders and sponsors.⁵⁹⁰

In part, the new emphasis on gaming was driven by researchers looking for tools that could clarify the challenges posed by a rising China and revanchist Russia. In 2014, RAND researchers initiated a series of games examining a potential invasion of the Baltic states by Russia. Rather than being sponsor-directed, the research was prompted by a small team of researchers who had been previously involved in games and wished to explore what might happen given the relative lack of relevant historical evidence.⁵⁹¹ The high degree of uncertainty about both the nature of the operational problem and the courses of action players would select to counter them caused the game to be designed with open adjudication—that is, players could see why decisions were made, argue for alternative assumptions, and the rules could be modified on the fly based on the collective experience in the room.⁵⁹² The game was originally run internally in order to identify specific challenges blue would have to confront if Russia attempted to invade the Baltic states. Later games evaluated the strengths and weaknesses of potential solutions.⁵⁹³ Once the game had generated analytic findings, additional games were run with players from diverse backgrounds to check the robustness of results and to expose a wider audience to the insights that emerged in

⁵⁸² Karl Mueller, "Paper Wargames and Policy Making: Filling the Baltic Gap or How I Learned to Stop Worrying and Love the D6," *Battles Magazine*, 2016.

⁵⁸³ Wasser et al., "Gaming Grey Zone Tactics."

⁵⁸⁴ Igor Mikolic-Torreira et al., "Exploring Cyber Security Policy Options in Australia," (Santa Monica, CA: RAND Corporation, RR-2008, 2017).

⁵⁸⁵ Bartels et al., "Do Differing Analyses Change the Decision?: Using a Game to Assess Whether Differing Analytic Approaches Improve Decisionmaking."

⁵⁸⁶ Ryan Henry, Steven Berner, and David A. Shlapak, "Serious Analytical Gaming: The 360 Game for Multidimensional Analysis of Complex Problems," (Santa Monica, CA: RAND Corporation, RR-1764, 2017).

⁵⁸⁷ Bartels et al., "Oceans 17 Tabletop Exercise: Findings and Recommendations."

⁵⁸⁸ Bartels et al., "Conceptual Design for a Multiplayer Security Force Assistance Strategy Game."

⁵⁸⁹ Brien Alkire, Sherrill Lingel, and Lawrence Hanser, "A Wargaming Method for Assessing Risk and Resilience of Military Command-and-Control Organizations," (Santa Monica, CA: RAND Corporation, TL-291-AF, 2018).

⁵⁹⁰ Work and Selva, "Revitalizing Wargaming Is Necessary to Be Prepared for Future Wars."

⁵⁹¹ Interview with Karl Mueller, Arlington, VA, May 2018.

⁵⁹² Interview with Barry Wilson, Arlington, VA, May 2018.

⁵⁹³ Interview with Karl Mueller, Arlington, VA, May 2018.

game play. The game was influential in the public debate about the emerging threat posed by Russia,⁵⁹⁴ including influencing Congressional testimony.⁵⁹⁵

Games were also applied to other operational problems that were coming to the forefront of RAND's research agenda. For example, 2012 and 2014, gaming efforts for the Pentagon's policy offices, including elements of the Office of the Under Secretary of Defense for Policy and the Office of Net Assessment, began to contemplate possible contours of a war with China using a similar operational gaming system. These games had serving military plan and execute the U.S. response to Chinese aggression in games that highlighted key operational challenges.⁵⁹⁶ The same approach used for conflict with Russia in the Baltics was also applied to other major theaters that were of interest to DoD.⁵⁹⁷ Like the early Baltic game, these games sought to identify specific operational challenges that would confront U.S. forces in order to prioritize later studies and analysis effort.⁵⁹⁸

At the same time as RAND was engaging in greater operational wargaming, researchers were experimenting with a range of designs appropriate to other emerging problems. In particular, researchers promoted a new emphasis on structured seminar-style games modeled on historical games like the force structure and planning games of the 1950s and 1960s. These games use mechanisms such as game boards, cards, and results tables to consider issues such as crisis management and escalation, portfolio investments, and force posture in ways that are flexible enough to incorporate player input while still being structured enough to generate useful data for analysis.⁵⁹⁹ These approaches are particularly helpful in generating data that supports comparison between games, allowing for more refined analysis.⁶⁰⁰ Less structured approaches to game design are also resurgent, with models such as the "360" seminar-style game⁶⁰¹ being used to explore emerging issues such as whole of society responses to cybercrime.⁶⁰²

⁵⁹⁴ David A. Shlapak and Michael W. Johnson, "Reinforcing Deterrence on Nato's Eastern Flank: Wargaming the Defense of the Baltic," (Santa Monica, CA: RAND Corporation, RR-1253, 2016).

⁵⁹⁵ House Armed Services Committee, Subcommittee on Tactical Air and Land Forces, *Deterring Russian Aggression in the Baltic States: What It Takes to Win*, March 1, 2017 2017.

⁵⁹⁶ Interview with David A. Ochmanek, Arlington, VA, October 2018.

⁵⁹⁷ Mueller, "Paper Wargames and Policy Making: Filling the Baltic Gap or How I Learned to Stop Worrying and Love the D6." p 57

⁵⁹⁸ Interview with Karl Mueller, Arlington, VA, May 2018.

⁵⁹⁹ Stacie Pettyjohn and Becca Wasser, "The Promise of Structured Strategic Wargames: Moving Beyond the Seminar" (paper presented at the International Studies Association Annual Conference, San Francisco, 2017 of Conference).

⁶⁰⁰ Bartels, "Short Games as Structured Comparisons: A Discussion of Methods."

⁶⁰¹ Henry, Berner, and Shlapak, "Serious Analytical Gaming: The 360 Game for Multidimensional Analysis of Complex Problems."

⁶⁰² Mikolic-Torreira et al., "Exploring Cyber Security Policy Options in Australia."

Thoughts on the Future

This chapter has laid out two key trends in RAND gaming. The first is a historical tendency to focus on games that explore policy problems in general, and that generate information for systems exploration by eliciting and synthesizing expert opinion in particular, which has only shifted in the last decades. The second trend is a tendency for gaming to wax and wane in popularity. The current moment in RAND gaming is characterized by both a high tempo of gaming, and unusual diversity in gaming output.

Given the convergence of these trends in gaming, it is natural to ask: “whither national security policy analysis gaming at RAND”? Gamers looking at the historical cycle of gaming booms and busts have raised concerns that the current popularity of games will be seen as the high watermark before another period of decline. Many leaders in the field warn that the current expansion is too large for qualified gamers to keep up with the demand for games in DoD. These long-time gamers raise concerns that the influx of underqualified gamers will result in poor quality games that risk alienating sponsors,⁶⁰³ and there is evidence to suggest their fears are not unfounded. Already, sponsoring organizations are warning that games are not meeting the needs of defense decisionmakers.⁶⁰⁴ If these indicators are true, the question then becomes how we can ensure that the wargaming that does persist in the coming period of decline preserves the best of current practices, both to ensure the remaining games are of high quality, and to set the field for success the next time gaming rises to prominence. In particular, how do we preserve the range of purposes and approaches currently in use, and continue to grow diverse, appropriate game design. In part, this monograph is intended to support these efforts by capturing a wide range of practices and making them accessible to a wider audience in a structured, logical way, so that gamers have a better sense of how practice has developed over time.

⁶⁰³ For example, see: Pettyjohn and Shlapak, "Gaming the System: Obstacles to Reinvigorating Defense Wargaming." And Perla, "Now Hear This—Improving Wargaming Is Worthwhile—and Smart."

⁶⁰⁴ Compton, "The Obstacles on the Road to Better Analytical Wargaming."

Chapter 10: Conclusions, Policy Recommendations, and Next Steps

This monograph starts from the premise that national security policy analysis games are an important tool in shaping American defense policy but echoes a concern that they are not always effective in shaping policy analysis. While there are many possible sources of pathological wargames, I diagnose the ineffectiveness stemming from two preentable sources. One potential cause is the lack of documented practice that helps game designers, sponsors and designers make a logical link between the specific research question they want to answer and the design of the game. Second, I argue games are less effective than desired at transferring insights from participants to those who were not directly involved in the game. In my assessment, underpinning both is a common issue: that games find themselves caught between an overly narrow view of what constitutes a scientifically valid analysis on one hand, and a perception that sees games primarily as an art, and thus not subject to the same standards as other types of analysis. This monograph proposes a third way of viewing games through a framework of social scientific practice with the aim of enabling games to better contribute to policy analysis. While recognizing that such foundations are not sufficient to ensure a masterful game design, I argue that they provide critical guiderails for a designer and the basis for sponsors and consumers of games to ensure minimum required competence.

This monograph puts forward a framework that links game purpose to design in several stages. Chapter 3 describes the three potential philosophical bases for a scientific approach to gaming. Chapter 4 describes the types of information games are asked to produce – based on a typology of game archetypes – and discusses which of the philosophies are consistent with each type. Chapters 5-8 expand on each archetype in more depth, explaining common design tradeoffs, illustrated with examples from historic games from the RAND archives, interviews with practicing game designers, and my own work as a game designer for policy analysis. Chapter 9 offers a history of gaming at RAND that contextualizes many of the example games and suggests how the use of gaming has changed over time. This final chapter summarizes the key argument of the monograph, highlights recommendations for game sponsors, designers, and consumers based on this research, and concludes by discussing next steps to test and strengthen the framework proposed here.

Conclusions

The U.S. national security establishment has long used policy analysis games as a tool for research. Senior leaders praise games' ability to help understand emerging problems and develop

potential solutions.⁶⁰⁵ However, despite increased resources and attention paid to gaming over the last five years,⁶⁰⁶ analytic leaders within DoD remain unsatisfied with the quality of gaming to support research. The concerned argue that poorly scoped research questions lead to game designs that do not produce credible information that can feed into analytic or decisionmaking processes.⁶⁰⁷ In response, designers have highlighted the need for sponsors to better connect games into the cycle of research⁶⁰⁸ and for gamers to more aggressively call out events that should not properly be considered games to improve quality.⁶⁰⁹ While these steps are important, this monograph argues that another tool that offers the possibility to improve the quality of games is clearly defining how games, alongside other tools for research, contribute to the advancement of knowledge and understanding.

To date, influential depictions of games treat them as an art form. The designer is tasked with creating a “ludic event”⁶¹⁰ in which players can generate a story about what issues are important to DoD and why as a means to inform game designer, sponsors, and players. This imagining of a game designer’s role is appealing because it highlights the strengths of a truly great game—an engaging event that is able to uncover new understanding and change the minds of players. However, I argue that without a solid, logical connection between the design and the research objectives of the project, it is far too easy for games to go off track when the focus is only on artistry. My experience also suggests that reliance on artistic language can also act as a barrier to accountability—too often, designers may be tempted to dismiss critical feedback as matters of “taste” or a sponsor “not getting it” rather than reckoning with serious concerns about the credibility of the work. Using a scientific approach to games offers a logical base for game design—it furnishes the tools to build a sturdy foundation. A scientific approach alone will not guarantee game architects from designing Frank Lloyd Wright’s Fallingwater, but it will reduce the inclination to unwarily build McMansions out of popsicle sticks.

It is important to be clear what is meant by a scientific approach to research and analysis. Too often our first thought on hearing “science” is a high school chemistry class experiment—rigid instructions dictated by someone else, used to “prove” a hypothesis, rather than the actual practice of researchers. Similarly, in DoD, too often “analysis” is treated as referring only to quantitative tools generally and operations research and systems analysis more specifically. However, it is important to bear in mind the term’s actual definition: “the detailed study or

⁶⁰⁵ Robert Work and Paul Selva, “Revitalizing Wargaming Is Necessary to Be Prepared for Future Wars,” *ibid.* December 8 2015.

⁶⁰⁶ Heath and Svet, “Better Wargaming Is Helping the Us Military Navigate a Turbulent Era.”

⁶⁰⁷ Compton, “The Obstacles on the Road to Better Analytical Wargaming.”

⁶⁰⁸ Phillip E Pournelle, “Can the Cycle of Research Save American Military Strategy?,” *ibid.* and Peter Perla et al., “Rolling the Iron Dice: From Analytical Wargaming to the Cycle of Research,” *ibid.* October 21, 2019.

⁶⁰⁹ ED McGrady, “Getting the Story Right About Wargaming,” *ibid.* November 8, 2019.

⁶¹⁰ *Ibid.*

examination of something in order to understand more about it.”⁶¹¹ In other words, common defense usage constricts the use of both terms to metonymous sets of methods and processes, losing sight of the *purpose* that actually defines the original term. Gamers supporting DoD have too often adopted this usage, referring to games as an “art” that is distinct from “analysis”⁶¹² or even explicitly arguing that games are not analysis even when they are conducted to enhance our understanding of policy.⁶¹³ In doing so, they are ceding ground to researchers who have overly-constrained the meaning of science and analysis.

There is undeniably an art to the best game designs which cleverly assemble mechanisms to tell compelling stories that also provide new understanding. However, this monograph argues that much of the practice of designing games to support research and analysis actually rests on logical approaches to generating new understanding that is fundamentally scientific at heart. For example, key texts on gaming argue that the game’s design should flow logically from the purpose.⁶¹⁴ I simply contend that those logics can be made explicit so they are more transparent and understandable and put in scientific terms so that they more clearly relate to other tools for research and analysis. The goal is not to change expert practice but rather to talk about that practice in a way that is more understandable and guides better decisionmaking by new game designers, sponsors, and consumers.

More nuanced understandings of science argue that the actual process of how we learn things about the world is more complex than a narrow view of science captures, and there is more than one way we conduct scientific inquiry. Borrowing from the literature on social science, which has long studied the same type of human and group decisionmaking that is the focus of games, we find multiple ways of conducting science. Three of these are particularly relevant to policy analysis.⁶¹⁵ The first is *positivist* research based on direct observation and comparison—we try to understand cause and effect by comparing when a potential cause is, and is not, present and see if we can determine whether it causes a particular outcome. Here we are usually trying to understand the importance of a single factor to generate a universal rule—X causes Y. A second approach is *critical realism*, which shares the logic of a detective story or courtroom drama—we may not be able to see cause and effect directly, but by looking at many events around the core issue, we can make a pretty good guess. Usually this “theory of the case” is about a specific set of actors in particular circumstances—we are looking for a “theory of success”⁶¹⁶ to solve a

⁶¹¹ Oxford Learner’s Dictionary, s.v. “analysis,” accessed December, 21, 2019, https://www.oxfordlearnersdictionaries.com/us/definition/american_english/analysis

⁶¹² For example, see: Perla et al., “Rolling the Iron Dice: From Analytical Wargaming to the Cycle of Research.”

⁶¹³ ED McGrady, “Getting the Story Right About Wargaming,” *ibid.* November 8, 2019.

⁶¹⁴ Perla, *The Art of War Gaming: A Guide for Professionals and Hobbyists*.

⁶¹⁵ Jackson, “The Conduct of Inquiry in International Relations: Philosophy of Science and Its Implications for the Study of World Politics.”

⁶¹⁶ Compton. “Analytical Gaming.”

specific problem. The third logic, *analyticism*, is one of model building. We ask smart people how they *think* the world works and combine their perspectives into a simple representation that is a useful shorthand like a schematic or set of rules of thumb. We know these are too simple to capture true reality, so they might not always hold true, but as long as they are a useful guide to decisionmaking, we can still find them helpful as ways to learn. None of these constructs ensure that we are correct—experiments may still go wrong and experts may cling to false predictions—but over time they let us build an understanding of how things work by providing a consistent, transparent logic that can be used to assess scientific design and output.

What is more, all three philosophies align with existing literature on game design, suggesting that at least a portion of leading designers already think about game design in a way that is highly compatible with these approaches. For example, people run sets of games, some of which include manned fighters while others feature drones to understand if people make decisions about how to respond to an unmanned system differently.⁶¹⁷ That is a game designed to generate knowledge about the role of a single factor using comparison—comfortably in line with positivism. Examples of the second type of game include those designed to generate a “theory of success”⁶¹⁸ about what strategies might work in a given operational environment such as a Russian invasion of the Baltics.⁶¹⁹ Other games bring together experts on a given topics, like nuclear escalation on the Korean peninsula,⁶²⁰ to try to capture how they think in a simple but useful model of the key dynamics. So while all these games clearly tie back to one of the core philosophical approaches to learning about the world, they do not all use the same philosophy of science.

At the same time, not all games are trying to generate the same type of information. In surveying games, this monograph defines four ideal types of information games are asked to produce. The first archetype is *systems exploration*. These are games that try to build out an understand of a particular policy problem from a range of perspectives. The second encompasses games that seek to generate *innovation* or new solutions to policy problems. Third are *alternative conditions* games that seek to understand how a key factor shapes decisionmaking processes and choices. Fourth are games designed to *evaluate* policies and strategies. As shown in Figure 11.1, each can be characterized as primarily seeking to better understand a problem or to propose solutions and by the likely audience for the research.





⁶¹⁷ Lin-Greenberg, "Game of Drones: What Experimental Wargames Reveal About Drones and Escalation."

⁶¹⁸ Jon Compton, "The Obstacles on the Road to Better Analytical Wargaming," *ibid.*

⁶¹⁹ Karl Mueller et al., "In Defense of a Wargame: Bolstering Deterrence on Nato's Eastern Flank," *ibid.* 2016.

⁶²⁰ Paul K. Davis, "Illustrating a Model-Game-Model Paradigm for Using Human Wargames in Analysis," (Santa Monica, CA: RAND Corporation, WR-1179, 2017); *ibid.*

Figure 10.1: Archetype of Information Produces by Games and the Philosophies that Underpin Them

	Develop an understanding of the problem	Develop strategies to address the problem
Early research to inform research team and sponsor	System Exploration  Most aligned with analyticism Least aligned with positivism	Innovation  Most aligned with critical realism Least aligned with positivism
Mature research to influence external stakeholders in a decision process	Alternative Conditions  Most aligned with positivism Less aligned with critical realism and analyticism	Evaluation  Only moderately aligned with all three philosophies

The figure also highlights that there is a fair degree of alignment between the different types of information to be generated and the philosophical approach to research that is most likely to underpin it. System exploration games are usually efforts at model building that align well with analyticism. Innovation games often use a critical realist logic to develop a “theory of success” about what strategies might work in a specific context. Alternative conditions games are well suited to the focus on understanding the role of a single factor in causing a decision outcome implied by positivism. All three of the philosophies wrestle with how best to conduct evaluations because of the artificialities of games. While the specific caveats will depend on the philosophical approach, all would agree that games provide only a tentative evaluation more appropriate for highlighting potential flaws in a strategy than providing any type of “validation” of the approach.

So how do these philosophies of science and archetypes of information generated by games actually help with game design? After all, approaches are not a “recipe” for games—there are still many, many choices a designer has to make about what is going to work. But both serve as guiderails for a designer—they signal when a design choice cuts against the argument that will underpin the findings. Each time a researcher makes a design decision, it raises the question: “is this choice consistent with the philosophy of science I am using to learn something from the game?” This process highlights choices that may interfere with the credibility of results, so

potential problems may be mitigated early. Since games are live events, the designer always working in a constrained space—for example ideal players may not be available for the planned period of the game, or the limitation of the physical space may dictate how many teams can play in isolation. Having a clearly articulated scientific construct makes it easier to know if these limitations are manageable, or if they raise serious questions about whether the game is worthwhile.

For example, let's say key players can only stay for part of the game duration—is that a problem that requires rethinking the game design's ability to meet the research objectives? While not ideal, in a “theory of success” game this situation might not be too big a problem—after all, if the concept works, it should not be too sensitive to who is implementing it. In contrast, if the goal is to compare the results of this game to one where all the players stayed the whole time, it represents a major difference that is not the point of the comparison. As a result, a researcher needs to be prepared either to rework the game (for example, by changing game dates, shortening the time required for both games, or altering the composition of the comparison game's player) or describe how the difference in participation might drive key differences in results (in technical terms, you have a confounding factor that must be explored as an alternative cause). In other words, depending on the game's purpose, the constraint is either manageable or a major threat to the credibility of game design.

This same process also works for sponsors looking to evaluate the quality of a game design. A game designer should always be able to explain how the design choices are consistent with the scientific logic of the work, or what will be done to mitigate those constraints that might make it harder to learn something from the game. When design choices do not align with the approach to research, post-game analysis should explain how the design might undercut findings and offer what evidence they can about how the design limits what can be learned from the game.

Finally, I argue that considering the philosophy of science used to generate information from a game can also help when connecting a game to broader programs of research, critical for games to have their proper influence in the DoD. The results of games can be strengthened by repeated play, but the value of repetition will be different for each of the four types of information. A researcher may also want to run a series of games moving between types. For example, an early game could be used to develop a better understanding of the problem, and that knowledge can be used to build a more refined game that can be used to develop strategies to address the challenge. Alternatively, early games can be used to generate initial hypotheses about how the policy system works or what types of strategies might be successful which can then be refined and stress-tested in later games. Finally, games can be linked to research using other analytic tools. This is particularly successful when using non-game tools that have different strengths and weaknesses than a game. For example, games' ability to study decisionmaking in detail can make them productive to couple with approaches that need to treat decisions as fixes by providing a better starting point for the later stage of analysis.

Policy Recommendations

This monograph has argued that national security policy analysis games can and should be designed in a scientific manner and presented a candidate framework for doing so. While feedback from interviews suggests that these practices will be nothing new to an experienced designer, for many sponsors, designers and consumers of games, if this framework is correct there are some clear takeaways about how to use games to support policy analysis. Most boil down to a plea for transparency by all stakeholders—clearly state what information a game is designed to produce, make design tradeoffs that align to that purpose, be clear about the limitation of the results, and respect those limitations when using game results to inform decisions.

Recommendations for the Sponsors of Games

The sponsors of games hold a critical position in improving the quality of games run for DoD. Since games are generally created and analyzed by external, contracting organizations, sponsors play a key role in asking for games to answer appropriate questions and setting out the standards that define a successful game.⁶²¹ They also play a key role in ensuring that gaming is properly connected to other analysis, since the work may be performed by different researchers.⁶²² However, game sponsors can also doom a game to failure by asking the impossible.⁶²³ The following steps can help minimize the incidence of game designs inimical to the games' intended use.

Recommendation 1: Provide clear guidance about the game's purpose and standard of evidence

Since the design of the game will follow from the purpose and objectives of the game, it is critical that these be clear and honest. Too often game objectives are designed by committee and lack the clarity necessary to help designers make the right choices. Even when the "official" wording must be ambiguous, honest conversations with a game designer about what information you need the game to produce will enable them to make better design choices. In particular, think about what decisions or next steps the game is teasing up. Are you looking for good research questions or promising concepts to shape where you spend your next round of study money? Do you want to know likely courses of action that can inform the development of modeling and simulation efforts? Are you looking for evidence to arm your boss about the potential flaws in a proposed process? The clearer you can be about what type of information you need for the game

⁶²¹ Pournelle, "Can the Cycle of Research Save American Military Strategy?."

⁶²² Peter Perla et al., "Rolling the Iron Dice: From Analytical Wargaming to the Cycle of Research," *ibid.* October 21, 2019.

⁶²³ Weuve et al., "Wargame Pathologies." And Downes-Martin, "Your Boss, Players and Sponsor: The Three Witches of War Gaming."

to generate, and the purpose it will be put to, the better chance a game designer can tailor the game to produce information that is actually helpful.

Related to this, it is also helpful to know what standard the sponsor and key consumers will use to judge the results. Of course, this is directly connected to the purpose of the game, since systems exploration and innovation games are designed to provide more preliminary types of evidence than are alternative conditions and evaluation games. However, more information about the audience of the game, its analytical priors, and the foundations of any subsequent analysis can also be helpful. For example, if the key decisionmaker the game is designed to inform has a strong positivist background, he or she may find game results from a positivist game easier to understand and thus more persuasive. If the purpose of the game makes a positivist design inappropriate, the information can still be helpful for designers because they can take specific steps to lay out for the consumer the approach they used, why it was more appropriate, and how such findings can be appropriately used.

Recommendation 2: Limit the purpose of the game to aid coherent design

Games represent a substantial outlay in terms of player time and the costs to support a team of researchers through design, execution, and analysis. To offset these costs, it can be tempting to assign a game multiple objectives. In some cases, it is possible to design a game that seeks to generate multiple types of information. However, the design of such a game will be complicated by the tensions between competing objectives.

In particular, games that combine objectives from more than one of the framework archetypes are likely to be particularly problematic. As discussed in chapters 5-8, the different goals of each of the four types of games generally require different foundational philosophies and suggest different tradeoffs when making choices about how to represent game elements. The differences between a game designed to support system exploration and one designed to conduct evaluation credibly are likely to be distinct enough that both cannot be equally well achieved. Either the game will be designed with the structure required for evaluation and thus will allow relatively little space for players to contribute their understanding of the problem, or players will be modifying the game's structure to accommodate their understanding of the problem making a baseline for evaluation difficult to assess. Differences between the other types are less stark, but the same basic types of tradeoffs are also likely to exist.

While a skilled designer may be able to successfully navigate these tensions to create useful and credible games, more often a game with competing objectives either will have to select which objective to achieve, or worse yet may make compromise design choices that undermine all of the objectives rendering the game less valuable to decisionmakers than might otherwise have been the case. As a result, it will generally be better to limit the purpose of the game to a small number of related research questions. This will allow the wargamer to focus game design choices on producing the highest quality of information possible to answer specific questions. When multiple objectives are necessary, ideally, they should all produce a single type of

information—that is, the game should seek to explore a policy system, consider alternative conditions, generate an innovative idea, or evaluate a potential policy.

Recommendation 3: Use the stated logic of the game design to oversee game development

For a game sponsor who is not an experienced game designer, it can be challenging to oversee game design. Because the range of game designs is so broad, past exposure to games is relatively unlikely to provide good expectations about what a new game should look like. When game designers use craft-based ways of talking about their decisionmaking, it can feel difficult to provide meaningful oversight of a national security policy analysis gaming effort.

The logics provided by the three philosophies of science and the four archetypes can provide standards to assess both in-progress materials and the analysis of the final game. If the connection between a design decision and the information to be generated from the game is not clear, that is an ideal time to follow up with the design team. Given the information you need the game to generate, what philosophy is the design team using and why? When faced with a design choice, what options did the team consider, and why is the selected design most aligned to the scientific logic of how the game is generating information? If a constraint forces a design decision that cuts against the underlying logic, does the team have a strategy for how they will mitigate this weakness in their analysis, or can they at least explain the potential analytical weaknesses that might result? What other types of analysis might be used to balance out inconclusive findings of the game? Such questions allow sponsors to provide meaningful, in stride guidance to design teams, improving the likelihood that the game provides anticipated information that can meaningfully help decisionmakers.

Recommendations for the Designers of Games

In guiding research, game designers have the primary responsibility to ensure that the games they design are intellectually honest and advance the understanding of the audience. They have the primary responsibility for ensuring that sponsors and consumers understand what information a game can and cannot provide and making logical design choices that maximize what can be credibly learned from a game. When a game's design is limited by external constraints (as will almost always happen), the designer has the responsibility to mitigate the risks to the credibility of game findings where possible and be transparent about the potential shortcomings of the game when mitigation is not possible.

Recommendation 1: Advise sponsors on the limits of the information a game can provide

Not all policy problems can be informed by gaming, and any specific game cannot provide all the information that a sponsor or customer might want. Sponsors may not understand these limitations, so it is the responsibility of a designer to advise them on the limits of the tool. Research questions that are not fundamentally about human decisionmaking are unlikely to be appropriate to game. As a result, they are not well suited to answer questions about the

performance of specific technical solutions or to offer predictions rather than indications of patterns of behavior. Games will tend to be most useful early in a process of learning about a policy problem or in cases where historical data are not available to enable other types of analysis.

Similarly, games that try to answer many research questions are unlikely to provide credible answers to them all, since the optimal design to answer one question is unlikely to be the same for a different research question. The four archetypes presented in this monograph can be a helpful tool to guide a discussion with sponsors about a small number of questions that are likely to recommend similar design choices. Designing a game to produce a single type of information is less likely to produce design tensions that could undermine the results. Similarly, ensuring that a game works within a single philosophical logic can help set the game up for success.

Recommendation 2: Make design tradeoffs to maximize the usefulness of information produced by the game, given constraints

All game designers must work in a world of constraints, ranging from the physical spaces available for game play, through to the players and on to the mechanics that can be used under different conditions. What is more, all design requires making tradeoffs about what aspects of a problem to simplify and which to highlight. Good designers make these choices in ways that are consistent with the type of information the game needs to produce. Chapters 5-8 discuss some specific rules of thumb that can help guide these choices, but the overarching recommendation is that design choices should be made to align with the central scientific logic that is being used to generate information from the game. Highlights include:

- **Systems exploration** games generally work to elicit and synthesize the mental models of expert players. Therefore, it is critical to ensure that diverse, truly expert players are brought together, focused on a specific policy problem, and given the tools to explain how they think the policy system works.
- **Alternative conditions** games depend on comparing player decisions under different conditions to understand the cause and effect of a specific factor or factors. Ensuring that key factors of interest are well-represented, and that other aspects of the game conditions very as little as possible, enables this type of analysis.
- **Innovation** games require designers to highlight where new decisions could be made to change how a system works and motivate players to propose new ideas. Therefore, games to spark innovation tend to focus on competition within a less constrained environment than current decisionmakers face.
- **Evaluation** games depend on having a credible representation of the outcomes of player decisions to enable judgement. This type of game focuses on the game rules, particularly the adjudication system.

Recommendation 3: Document the logical links among game purpose, design, and findings

Too often foundational choices about the philosophical underpinnings of a game, the type of information it is designed to produce, and the design choices that are made along the way are not

included in documentation about the game design. The majority of space instead describes how the game looked and how play unfolded. While these choices can feel obvious deep into the project, when both sponsor and designer are well aware of the connection, documenting these fundamental choices is critical to ensure that game reports are credible in the eyes of people not deeply involved in the project. This can range from a stakeholder receiving recommendations to analysts hoping to leverage the findings of past games to conduct broader analysis. As a result, these fundamental choices should always be clearly documented.

The frameworks for thinking about the philosophies underpinning research and archetypical information to be produced from a game are intended to help this process by providing common, accessible language with which to document these key ideas. Such additions to game reports need not be long, but a reader should come away with a clear understanding of the philosophy that guided designer choices and how these choices impact the credibility of design and findings.

Recommendations for the Consumers of Games

Last, there are those who are not directly involved with the conception and design of a game, but who may seek the information that the game produces. This might include stakeholders and decisionmakers whom the game is designed to inform as well as researchers attempting to leverage the game's results to extend knowledge. Since these stakeholders have little ability to shape the design, execution, and analysis of the game, recommendations focus more on how to best interpret game results in order to use game findings appropriately.

Recommendation 1: Use care when applying findings from a game to a new purpose

When looking at a game retroactively, it can be easy to read greater confidence into the findings and assume that the game's original architects must have had the same approach to research as a new team hoping to leverage the findings of a game they were not involved in. Such assumptions can be dangerous, particularly if a game report does not include adequate information about the original purpose and design of a game. For example, if a game is run twice to look at two different strategies, it can be easy to assume the results were intended to be comparable, but if the report does not document the efforts that were made to ensure that the games were similar on other dimensions, this may not be a safe assumption. Similarly, it can be easy to take a game intended to build an understanding of a problem and treat it as supporting evidence for a proposed solution. Such analytical jumps apply a different, inappropriate logic to a game, and can cause findings to be misused.

This can also happen at a more granular level. Details of game results may include recordings of specific outcomes such as the number of planes lost or missiles expended. Since such findings are the result of adjudication, they should be treated as the interaction of player decisions with the rules of the game rather than solely attributed to one or the other. The results may perhaps represent extreme behavior of the adjudication model rather than average behavior. At the same time, a different group of players might have opted to make different choices. As a result, any

attempt to overgeneralize a specific outcome, and particularly any claim that such results are predictive, should be avoided. It may be appropriate to use these results more narrowly as the basis for other analysis, but considerable care should be taken to ensure that the logical assumptions of the new analysis are consistent with that of the original game.

Recommendation 2: Evaluate the information generated by a game based on original approach of the game, rather than other standards

Conversely, it can be all too easy to dismiss findings from a game by applying the wrong standard to judge the worth of a game. In particular, the dominance of positivists in research communities can cause games that are not designed using a positivist logic to be dismissed as “unscientific” and thus not a suitable basis for decisionmaking. Alternatively, the goals of positivism, particularly generalizability, can be inappropriately applied to game results produced using other logics. For example, the “theory of success” developed in critical realist research is specific to the context in which it was developed—it would be inappropriate to judge such research for not offering a universal concept (if indeed such a thing could be found). Again, this recommendation comes down to respecting the logic governing research and taking care when applying a different standard to the results.

Next Steps: Options for Testing the Framework

This monograph has argued that traditional artistic approaches to game design are insufficient—in order to advance the state of research a scientific foundation is required. While this work proposes such a framework for consideration, available evidence is insufficient to offer a compelling empirical test of the approach. This final section discusses what data would be needed to conduct such a test and offers alternatives in the face of probable delays and difficulty in conducting detailed empirical research.

Ideal testing of the core hypothesis of this work will require answering two related but distinct questions. The first is whether it is possible to verify that the framework laid out in this monograph is, in fact, a fair description of game design for policy analysis and research purposes. For example, are there other philosophies of science suggesting different logics of inquiry that should also be considered by game designers? Are there additional or alternative archetypes of the information generated by games that should be considered? As currently structured, the approach outlined in this work explicitly does not claim to be mutually exclusive and collectively exhaustive, since it is not built on a comprehensive survey of games. Expanding the base of evidence to take a more comprehensive view would enable refinement of the framework to enhance confidence in its descriptive power.

However, as noted in Chapter 2 there are substantial barriers that prevent this type of survey today. First, without a census of who games within the National Security organizations, it is not possible to know how comprehensive any particular sample of games might be. Second, many

game reports are also not publicly releasable, limiting the forums in which such research can be conducted. Third, and perhaps most importantly, the current standards of documentation of games do not generally capture information about the philosophical underpinnings of a game, its true purpose, or the tradeoffs made by the game designer. As a result, the raw data needed to inform systematic, empirical testing of the framework is not easily available.

Second, future research would need to answer what is perhaps the more difficult question: are games designed using this framework better able to support decisionmakers? Measuring the effect of policy analysis on decisionmaking has long been a struggle for the evaluations community.⁶²⁴ The timeline for both analysis and decisionmaking and implementation is long and it can be difficult to conduct credible process tracing or other approaches to tracing how a specific factor might be causing changes in decisionmaking in complex bureaucratic systems. This is particularly true when decisions are highly consequential and politically salient, as national security often is. In short, observing a complex, sensitive, legally protected decisionmaking process offers many challenges to the toolkit for broad empirical research.

The barriers noted above do not mean that there is not room for progress but suggest that the work needed is beyond the scope of a single project, or even a single researcher. For example, in this work I have focused considerable attention on studying gaming at RAND, because that was the institution to which I had the easiest access. Similar systematic studies could be undertaken by other institutions of their own records and experts that complement the work of this monograph. In particular, direct government sponsorship of future work would enable access to records that could not be included in the context of an academic work that would be made broadly available.

Second, moving forward there is the opportunity to change game reporting standards in order to make empirical work on game design more feasible in the future. Research in this monograph was limited because all too often, even when game documentation was accessible, it lacked key information needed to understand how the game had been designed to meet research and analysis objectives. Going forward, if games are better documented, broader empirical research to identify trends in games design will be possible. To that end, Appendix A offers a sample template for documenting game designs that seeks to capture some of the information I most often found useful in researching this monograph.

One potential tool to change current business practices is the repository of game reports established by former Deputy Secretary of Defense Robert Work. This repository serves as a cross cutting clearing house for game reports, with the goal of enabling senior leader visibility of efforts across the department as a whole.⁶²⁵ One way to potentially change documentation

⁶²⁴ Nancy Cartwright and Jeremy Hardie, *Evidence-Based Policy: A Practical Guide to Doing It Better* (Oxford, UK: Oxford University Press, 2012).

⁶²⁵ See for example: Oleg Svet and Garrett Heath, "How the Joint Staff Calculated a Defense Program's Return on Investment," *Defense One*, 2018.

practices is to use the format of the repository to include types of information not currently included in game design reports. This might entail adopting the elements proposed in this monograph as explicit data collection fields, so that sponsors entering game information would be prompted to consider the philosophy, archetype, and design tradeoffs of their game. Measures of game effectiveness could also be considered for collection. While such prompts will not generate instant or fully predictable change in documentation practices (after all, the structural barriers to good reporting noted in Chapter 2 will not be removed by these actions), the repository offers a “nudge” to improve practices across the board.

Third, in the more immediate future, assessment of both the framework’s descriptive power and utility can be done on a more ad hoc bases by receiving feedback from game designers, sponsors, and consumers who read this work and attempt to apply it in their own practice. While such feedback is unlikely to be systematic, the public release of this monograph will hopefully bring a broader range of perspectives than those I was able to interview as part of this initial research. What is more, over time it will be possible to get a better sense of who this approach resonates with, and where more work is needed to clarify the existing framework, integrate additional ideas, or offer clear alternative models.

In that vein, more than suggesting any particular question or approach for future research, I hope a major contribution of this work is stimulating addition debate and writing in the community. In the course of researching this monograph, I was repeatedly struck by both the enduring nature of fundamental debates in the field, and the relative lack of progress in articulating the theoretical underpinnings of gaming across the decades. All too often, texts from the 1950s and 1960s laid out the same positions as current works, often with more vigorous accessible debate. Today, the incentives of the field have tended to relegate such dialogue to the sidelines of conferences and email exchanges where they are accessible to few. I hope this work is a spur to other national security gamers to bring those discussion out of the margins so they can benefit from more public dialogue. The last few years have seen a number of important new works published, from which this work has benefited considerably. I hope there are many more such works to follow.

Appendix A: Sample Template for Documenting Game Designs

Game Purpose and Objective

- What is the formal purpose and objective of the overall project? Of the game?
- What information was the game intended to produce?
- Who is the audience for the game, and what types of information do they find credible?
- What is the social scientific logic by which the game will generate that information?

Game Design Tradeoffs

- How did you define the limits of the policy problem in the game? What elements might you have included that you opted not to? How would have including that element changed the game?
- How did you decide which parts of the problem to distill and which to include in more detail, given the problem of interest and the research questions?
- How much did the game design opt to formalize or make explicit, versus leaving space for players to incorporate their own understanding of the problem? How did this balance help meet the game's objectives?
- Did you make tradeoffs between the game's environment, actors, and rules (for example, adding rules to compensate for a weakness in how actors are portrayed)? If so, what alternatives did you consider and why was this the best available option?
- How did the processes you selected make game processes transparent or traceable? Where they are not transparent, why should a reader believe that the causal argument you are presenting is credible?
- If making a comparison between multiple games, on what dimensions are they similar? Different? When are those similarities and differences useful for answering the research question, and when might they be confounding?

Environment

- What criteria were used to select the game environment? What other environments were considered, and why is the selected environment most helpful?
- What tradeoffs were made between breadth and depth? Put differently, what level of analysis is the environment represented at, and why?
- To what extent did the design team populate the game environment vs. allow players to make key assumptions? What were the costs and benefits of that approach?

Actors

- Which actors are represented by players in the game? How are actors not represented by players included?
- What actors are not represented? How might that impact game insights?

- What level(s) of analysis are actors represented at? How might this influence key decisionmaking processes and choices?
- How were players in the game selected? Consider factors like demographics, professional expertise and training, past experience, and current position.
- How are the players different than the real-world decisionmakers of interest? How might that impact game insights?
- How are player interactions different than the real-world actors they represent? How might that impact game insights?
- How were players motivated to engage in the game?

Rules

- To what extent were game rules formalized? What was the advantage of this choice?
- When were rules stipulated by the game design team, and when were players to make key assumptions? What was the value of the selected process, given the analytical information the game aimed to generate?
- How were teams able to communicate? In what ways was this artificial, and what impact might that have on game insights?
- What information was accessible to players and what was hidden? In what ways was this artificial, and what impact might that have on game insights?
- What game outcomes were deterministic, and which involved chance? What types of uncertainty is chance designed to introduce into the game, and how does it further the research question?

Insights and Finds

- How was the game documented in order to capture the desired information about decision processes, choices, and outcomes? What observations were collected? What procedures were used? What are the strengths and weaknesses of the selected approach compared to alternatives?
- What information did the game generate? What limits and caveats should we apply?
- If the game did not produce the desired information, why were the findings limited? What could be done to improve the credibility and utility of findings if this approach is used again?

Recommendations

- How should this information shape decisionmaker behavior? What types of decisions could be made based on the findings?
- How can we assess whether this game had the intended effect?

Bibliography

- Alkire, Brien, Sherrill Lingel, and Lawrence Hanser. "A Wargaming Method for Assessing Risk and Resilience of Military Command-and-Control Organizations." (Santa Monica. CA: RAND Corporation, TL-291-AF, 2018).
- Allen, Thomas B. *War Games : The Secret World of the Creators, Players, and Policy Makers Rehearsing World War Iii Today*. New York: McGraw-Hill, 1987.
- Arnold, David L. "Simsript Program for Operational Deployment (Spod)." (Santa Monica, CA: RAND Corporation, D-20393-ARPA, 1970).
- Averch, Harvey A., Allen R. Ferguson, and William M. Jones. "Europe, Sac and Safe: Some Issues for the Next Decade." (Santa Monica, CA: RAND Corporation, D-10895-PR, 1963).
- Averch, Harvey A., and Marvin M. Lavin. "Dilemmas in the Politico-Military Conduct of Escalating Crises." (Santa Monica, CA: RAND Corporation, P-3205, 1965).
- . "Simulation of Decisionmaking in Crises : Three Manual Gaming Experiments." (Santa Monica, CA: RAND Corporation, RM-4202-PR, 1964).
- Averch, Harvey A., and Sorrel Wildhorn. "Risk, Ambiguity and Force Structure: An Analysis of General-War Forces and Strategic Objectives in Cases C and D (of Six Case Studies)." (Santa Monica: RAND Corporation, RM-3511-PR, 1963).
- Bartels, Elizabeth M. "Adding Shots on Target: Wargaming Beyond the Game." *War on the Rocks*, 2017.
- . "Building a Pipeline of Wargaming Talent: A Two-Track Solution." *War on the Rocks*, 2018.
- . "Games as Structured Comparisons: A Discussion of Methods." In *International Studies Association*. San Francisco, CA, 2018.
- . "Gaming – Learning at Play." *ORMS Today*, 2014.
- . "Insights from a Survey of the Wargaming Community." In *Military Operations Research Society Wargaming Community of Practice*. Alexandria, VA, 2017.
- . "Mors 83th Panel Aar: Typologies of Game Standards." In *Paxsims*, edited by Rex Brynen, 2015.
- Bartels, Elizabeth M., Christopher S. Chivvis, Adam R. Grissom, and Stacie L. Pettyjohn. "Conceptual Design for a Multiplayer Security Force Assistance Strategy Game." (Santa Monica, CA: RAND Corporation, RR-2850, 2019).
- Bartels, Elizabeth M., Adam R. Grissom, Russell Hanson, and Christopher A. Mouton. "Oceans 17 Tabletop Exercise: Findings and Recommendations." (Santa Monica, CA: RAND Corporation, RR-2521-OSD, 2019).
- Bartels, Elizabeth M., Margaret McCown, and Timothy Wilkie. "Designing Peace and Conflict Exercises: Level of Analysis, Scenario, and Role Specification." *Simulation & Gaming* 44, no. 1 (2013): 36 - 50.
- Bartels, Elizabeth M., Igor Mikolic-Torreira, Steven W. Popper, and Joel B. Predd. "Do Differing Analyses Change the Decision?: Using a Game to Assess Whether Differing Analytic Approaches Improve Decisionmaking." (Santa Monica, CA: RAND Corporation, 2019).

- Barzashka, Ivanka. "Wargaming: How to Turn Vogue into Science." *Bulletin of the Atomic Scientists*, 2019.
- Beach, Derek, and Rasmus Brun Pedersen. *Causal Case Study Methods: Foundations and Guidelines for Comparing, Matching, and Tracing*. Ann Arbor, MI: University of Michigan Press, 2016.
- Bennett, Bruce W., and Paul K. Davis. "The Role of Automated War Gaming in Strategic Analysis." (Santa Monica, CA: RAND Corporation, P-7053, 1984).
- Bloomfield, Lincoln P. "Reflections on Gaming." *Orbis* 27, no. 4 (1984): 783-89.
- Brady, Henry E., and David Collier, eds. *Rethinking Social Inquiry: Diverse Tools, Shared Standard*. New York, NY: Rowman & Littlefield Publishers, Inc., 2004.
- Brightman, Hank J., and Melissa K. Dewey. "Trends in Modern War Gaming." *Naval War College Review* 67, no. 1 (2014): 17-30.
- Brown, Thomas A. "Elementary Cost-Effectiveness Computations Based on the Blue Menu of the Safe Game." (Santa Monica, CA: RAND, D-10717, 1962).
- Brown, Thomas A., and Edwin W. Paxson. "A Retrospective Look at Some Strategy and Force Evaluation Games." (Santa Monica, CA: RAND Corporation, R-1619, 1975).
- Bruner, Jerome. *Actual Minds, Possible Worlds*. Cambridge, MA: Harvard University Press, 1986.
- Brynen, Rex. "Setting the (Wargame) Stage." In *Paxsims*, edited by Rex Brynen, 2019.
- Builder, Carl H., and William M. Jones. "Gaming a Persian Gulf Contingency." (Santa Monica, CA: RAND Corporation, N-1022-AF, 1979).
- Caffrey, Matthew B. *On Wargaming: How Wargames Have Shaped History and How They May Shape the Future*. Newport, RI: Naval War College Press, 2019.
- Cartwright, Nancy, and Jeremy Hardie. *Evidence-Based Policy: A Practical Guide to Doing It Better*. Oxford, UK: Oxford University Press, 2012.
- Centers, Marian, Norman Crolee Dalkey, Olaf Helmer-Hirschberg, and F. B. Thompson. "Rules for Straw." (Santa Monica, CA: RAND Corporation, D-1955-PR, 1953).
- Coessens, Kathleen, Darla Crispin, and Anne Douglas. *The Artistic Turn: A Manifesto*. Leuven, Belgium: Leuven University Press, 2009.
- Compton, Jon. "Analytical Gaming." 2014).
- . "The Obstacles on the Road to Better Analytical Wargaming." *War on the Rocks*, 2019.
- Connable, Ben, Michael J. McNerney, William Marcellino, Aaron Frank, Marek N. Posard, S. Rebecca Zimmerman, Natasha Lander, et al. "Will to Fight: Returning to the Human Fundamentals of War." (Santa Monica, CA: RAND Corporation, RB-10040-A, 2019).
- Costikyan, Greg. *Uncertainty in Games*. Cambridge, MA: MIT Press, 2013.
- Culora, Thomas J. "A War-Gaming Renaissance." *Proceedings*, 2016.
- Curry, John, and Tim Price. *Matrix Games for Modern Wargaming: Developments in Professional and Educational Wargames*. History of Wargaming Project, 2014.
- Darilek, Richard E., and James C. Wendt. "Korean Arms Control : Political-Military Strategies, Studies, and Games." (Santa Monica, CA: RAND Corporation, MR-489-A, 1994).
- Davis, J. A., and S. C. Silviski. "Estimated Total Obligational Authority of the Force Structures Generated by the Safe/Acws Games." (Santa Monica, CA: RAND Corporation, D-10948-PR, 1963).
- Davis, Paul K. "The Base of Sand Problem : A White Paper on the State of Military Combat Modeling." (Santa Monica, CA: RAND Corporation, N-3148-OSD/DARPA, 1991).

- . "Illustrating a Model-Game-Model Paradigm for Using Human Wargames in Analysis." (Santa Monica, CA: RAND Corporation, WR-1179, 2017).
- Davis, Paul K., and James A. Winnefeld. "The Rand Strategy Assessment Center : An Overview and Interim Conclusions About Utility and Development Options." (Santa Monica, CA: RAND Corporation, R-2945-DNA, 1983).
- Davison, W. Phillips. "A Summary of Experimental Research on "Political Gaming"." (Santa Monica, CA: RAND Corporation, D-5695-RC, 1958).
- Dekker, Anthony H. "Revisiting "Scudhunt" and the Human Dimension of New: Some Thoughts." (Canberra, Australia: Defence Systems Analysis Division, DSTO, Australian Department of Defence, Undated).
- Derosa, John, and Lauren Kinney. "Narrative Analysis of Wargaming." In *Connections Wargaming Conference US*. Washington, DC, 2018.
- Dewar, J.A., J.J. Gillogly, and M.L. Juncosa. "Non-Monotonicity, Chaos, and Combat Models." (Santa Monica, CA: RAND Corporation, 1991).
- DeWeerd, Harvey A. "A Contextual Approach to Scenario Construction." (Santa Monica, CA: RAND Corporation, P-5084, 1973).
- . "Nato Limited War Crises: Some Research Guidelines." (Santa Monica, CA: RAND Corporation, D-12201, 1964).
- . "Political-Military Scenarios." (Santa Monica, CA: RAND Corporation, P-3535, 1967).
- . "A Scenario for a Limited War in the Northern Flank of Nato, 1966." (Santa Monica, CA: RAND Corporation, D-12077-PR, 1964).
- DeWeerd, Harvey A., T. E. Greene, and F. M. Sallagar. "A Report on the Rand Limited War Program: A Project Back Stop Briefing." (Santa Monica, CA: RAND Corporation, D-6354-PR, 1958).
- Downes-Martin, Stephen. "Group Dynamics in Wargames and How to Exploit Them." In *Connections North 2019 Wargaming Conference*. Montreal, CA, 2019.
- . "Preference Reversal Effects and Wargaming." In *Connections North 2020 Wargaming Conference*. Montreal, CA, 2020.
- . "Your Boss, Players and Sponsor: The Three Witches of War Gaming." *Naval War College Review* 67, no. 1 (Winter 2014 2014): 31-50.
- Evans, John P. "Guide for Ground Force Adjudication in War Games." (Santa Monica, CA: RAND Corporation, D-4765, 1957).
- Farrell, Theo "Military Adaptation in War." In *Military Adaptation in Afghanistan*, edited by Theo Farrell, Frans Osinga and James A. Russell, 1-23. Palo Alto, CA: Stanford University Press, 2013.
- Fielder, James. "Reflections on Teaching Wargame Design." *War on the Rocks*, January 1 2020.
- Fitch, Kathryn, Steven Bernstein, Maria D. Aguilar, Bernard Burnand, Juan Ramon LaCalle, Pablo Lazaro, Mirjam van het Loo, *et al.* *The Rand/Ucla Appropriateness Method User's Manual*. Santa Monica, CA: RAND, 2001.
- Flanagan, Mary. *Critical Play: Radical Game Design*. Cambridge, MA: MIT Press, 2009.
- Frank, Aaron. "The Philosophy of Science and Intelligence: Rethinking Science in Support of Intelligence." In *International Studies Association Annual Conference*. San Diego, CA, 2012.
- Gardner, Howard. *The Arts and Human Development; a Psychological Study of the Artistic Process*. New York, NY: Wiley, 1973.

- George, Alexander, and Andrew Bennet. *Case Studies and Theory Development in the Social Sciences*. Boston, MA: MIT Press, 2005.
- Ghamari-Tabrizi, Sharon. "Simulating the Unthinkable: Gaming Future War in the 1950s and 1960s." *Social Studies of Science* 30, no. 2 (2000): 163-223.
- 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, D-3164-RC, 1955).
- . "Summary of Cold-War Game Activities in the Social Science Division." (Santa Monica, CA: RAND Corporation, D-2850, 1955).
- . "Toward a Cold War Game." (Santa Monica, CA: RAND Corporation, D-2603, 1954).
- Goldhamer, Herbert, and Hans Speier. "Some Observations on Political Gaming." (Santa Monica, CA: RAND Corporation, P-1679, 1959).
- Goldsen, Joseph M. "The Political Exercise: An Assessment of the Fourth Round." (Santa Monica, CA: RAND Corporation, D-3640-RC, 1956).
- Guetzkow, Harold, Chadwick Alger, Richard A. Brody, Robert C. Noel, and Richard C. Snyder. *Simulations in International Relations: Developments for Research and Teaching*. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1963.
- Hall, Joseph M., and M. Eric Johnson. "When Should a Process Be Art, Not Science." *Harvard Business Review*, 2009.
- Halloran, Richard. "Pacaf's "Vision" Thing." *Air Force Magazine*, 2009, 54-56.
- Hanley, John T. "On Wargaming." University of Maryland, 1991.
- . "Some Theory and Practice of Serious "Futures" Games." In *Connections Wargaming Conference*. Carlisle, PA, 2019.
- Harris, Kathleen, and Louis H. Wegner. "Tactical Airpower in Nato Contingencies : A Joint Air-Battle/Ground-Battle Model (Tally/Totem)." (Santa Monica, CA: RAND Corporation, R-1194-PR, 1974).
- Heath, Garrett, and Oleg Svet. "Better Wargaming Is Helping the Us Military Navigate a Turbulent Era." *Defense One*, 2018.
- Helmer-Hirschberg, Olaf. "Strategic Gaming." (Santa Monica, CA: RAND Corporation, P-1902, 1960).
- . "War Game Rules (Fourth Version)." (Santa Monica, CA: RAND Corporation, D-446, 1949).
- Helmer-Hirschberg, Olaf, and Robert E. Bickner. *How to Play Safe--Book of Rules of the Strategy and Force Evaluation Game*. Santa Monica, CA: RAND Corporation, 1961.
- Helmer-Hirschberg, Olaf, and Lloyd S. Shapley. "Brief Description of the Swap Game." (Santa Monica, CA: RAND, RM-2058-PR, 1957).
- Henry, Ryan, Steven Berner, and David A. Shlapak. "Serious Analytical Gaming: The 360 Game for Multidimensional Analysis of Complex Problems." (Santa Monica, CA: RAND Corporation, RR-1764, 2017).
- Heuer, Richard J, and Randolph H. Pherson. *Structured Analytic Techniques for Intelligence Analysis*. Washington, DC: Congressional Quarterly, 2011.
- Holliday, Leo P., and Arnold S. Mengel. "A Tactical Air Superiority War Game." (Santa Monica, CA: RAND Corporation, D-2931, 1955).

- Hughes, Tucker, and Josh Jones. "The Parts and the Whole: Linking Operational and Strategic Wargaming." *Phalanx* 51, no. 2 (2018): 50-57.
- 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.
- Johnson, Dominic D. P., Rose McDermott, Emily S. Barrett, Jonathan Cowden, Richard Wrangham, Matthew H. McIntyre, and Stephen Peter Rosen. "Overconfidence in Wargames: Experimental Evidence on Expectations, Aggression, Gender, and Testosterone." *Proceedings of the Royal Society* 273 (2006): 2513-20.
- Jones, William M. "On Free-Form Gaming." (Santa Monica, CA: RAND, N-2322-RC, 1985).
- . "On the Adapting of Political-Military Games for Various Purposes." (Santa Monica, CA: RAND, N-2413-AF/A, 1986).
- . "One View of Games, Simulations and Analogs." (Santa Monica, CA: RAND Corporation, D-12290-PR, 1964).
- Judge, Sawyer. "The Wargaming Guild: How the Nature of a Discipline Impacts Its Craft and Whether It Matters." Georgetown University, 2019.
- Kahan, James P., William M. Jones, and Richard E. Darilek. "A Design for War Prevention Games." (Santa Monica, CA: RAND Corporation, N-2285-RC, 1985).
- Kaplan, Fred. *The Wizards of Armageddon*. Palo Alto, CA: Stanford University Press, 1991.
- Kecskemeti, Paul. "War Games and Political Games." (Santa Monica, CA: RAND Corporation, D-2849, 1955).
- Kent, Sherman. "Estimates and Influence." In *Sherman Kent and the Board of National Estimates: Collected Essays*, edited by Donald P Steury, 58-67. Washington, DC: Central Intelligence Agency, 1994.
- King, Gary, Robert O. Keohane, and Sidney Verba. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, NJ: Princeton University Press, 1994.
- Klein, Gary. *Sources of Power: How People Make Decisions*. Cambridge, MA: The MIT Press, 1998.
- Kofman, Michael. "Fixing Nato Deterrence in the East Or: How I Learned to Stop Worrying and Love Nato's Crushing Defeat by Russia." *War on the Rocks*, 2016.
- Kuhn, Thomas. *The Structure of Scientific Revolution*. Chicago, IL: University of Chicago Press, 1970.
- Lakatos, Imre. "History of Science and Its Rational Reconstructions." In *The Methodology of Scientific Research Programmes*, edited by John Worrall and Gregory Currie, 102-38. Cambridge, UK: Cambridge University Press, 1978.
- Lange, Matthew. *Comparative-Historical Methods*. Los Angeles, CA: SAGE, 2013.
- Lavin, Marvin M. "Blue Military Moves in the Southern Flank Crisis and Limited War Game -- 1968." (Santa Monica, CA: RAND Corporation, D-12489-PR, 1964).
- Lempert, Robert J. , Steven W. Popper, David G. Groves, Nidhi Kalra, Jordan R. Fischbach, Steven C. Bankes, Benjamin P. Bryant, *et al.* "Making Good Decisions without Predictions: Robust Decision Making for Planning under Deep Uncertainty." (Santa Monica, CA: RAND Corporation, RB-9701, 2013).
- 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, P-7719, 1991).
- Lillard, John M. *Playing War: Wargaming and U.S. Naval Preparations for World War II*. Lincoln, NE: Potomac Books, 2016.

- Lin-Greenberg, Erik. "Game of Drones: What Experimental Wargames Reveal About Drones and Escalation." *War on the Rocks*, 2019.
- . "(War)Game of Drones: Remote Warfighting Technology and Escalation Control Evidence from Wargames." SSRN, 2019).
- Longley Brown, Graham. *Successful Professional Wargames: A Practitioner's Guide*. Edited by John Curry. The History of Wargaming Project, 2019.
- Lourie, Megan, and Elizanth Rata. "Using a Realist Methodology in Policy Analysis." *Education Philosophy and Theory* 49, no. 1 (2017): 17-30.
- Massey, H. G. "The Xray Force Planning Cost Model." (Santa Monica, CA: RAND Corporation, D-19847-ARPA, 1970).
- Matsumura, John, Randall Steeb, Thomas J. Herbert, Scot Eisenhard, John Gordon, Mark R. Lees, and Gail Halverson. "The Army after Next : Exploring New Concepts and Technologies for the Light Battle Force." (Santa Monica, CA: RAND Corporation, DB-258, 1999).
- Mayer, Igor S. "The Gaming of Policy and Politics of Gaming." *Simulation & Gaming* 40, no. 6 (2009): 825-62.
- McEvoy, Phil, and David Richards. "Critical Realism: A Way Forward for Evaluation Research in Nursing?". *Journal of advanced Nursing* 43, no. 4 (2003).
- McGrady, ED. "Getting the Story Right About Wargaming." *War on the Rocks*, November 8, 2019 2019.
- Mikolic-Torreira, Igor, Don Snyder, Michelle Price, David A. Shlapak, Sina Beaghley, Megan Bishop, Sarah J. Harting, *et al.* "Exploring Cyber Securty Policy Options in Australia." (Santa Monica, CA: RAND Corporation, RR-2008, 2017).
- Millot, Marc Dean, Roger C. Molander, and Peter A. Wilson. "'The Day After...' Study : Nuclear Proliferation in the Post-Cold War World. Volume I, Summary Report." (Santa Monica, CA: RAND Corporation, MR-266-AF, 1993).
- Molander, Roger C., David Aaron, Robert Edwards Hunter, Martin C. Libicki, Douglas Shontz, and Peter A. Wilson. "The Day After... In Jerusalem : A Strategic Planning Exercise on the Path to Achieving Peace in the Middle East." (Santa Monica, CA: RAND Corporation, CF-271, 2009).
- Mood, Alexander McFarlane. "War Gaming as a Technique of Analysis." (Santa Monica, CA: RAND Corporation, P-899, 1954).
- Mood, Alexander McFarlane, and Melvin P. Peisakoff. "A Planning Factor War Game." (Santa Monica, CA: RAND Corporation, D-1382-PR, 1952).
- Morgan, Stephen L., and Christopher Winship. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. New York, NY: Cambridge University Press, 2007.
- Moroney, Jennifer D. P. "Assessing the U.S. Air Force Unified Engagement Building Partnerships Seminars." (Santa Monica, CA: RAND Corporation, DB-605-AF, 2011).
- Mueller, Karl. "Paper Wargames and Policy Making: Filling the Baltic Gap or How I Learned to Stop Worrying and Love the D6." *Battles Magazine*, 2016, 53-57.
- Mueller, Karl, David A. Shlapak, Michael W. Johnson, and David Ochmanek. "In Defense of a Wargame: Bolstering Deterrence on Nato's Eastern Flank." *War on the Rocks*, 2016.
- Mussington, David. "The "Day after" Methodology and National Security Analysis." In *New Challenges, New Tools for Defense Decisionmaking*. Santa Monica, CA: RAND Corporation, 2003.

- Nash, John F., and Robert M. Thrall. "Some War Games." (Santa Monica, CA: RAND, D-1379, 1952).
- Northrop, Gaylord M. "A Resume of Red Actions in War Game X-Ray." (Santa Monica, CA: RAND Corporation, D-15247-ARPA, 1966).
- . "Use of Multiple on-Line, Time-Shared Computer Consoles in Simulation and Gaming." (Santa Monica, CA: RAND Corporation, P-3606, 1967).
- Parson, Edward. "What Can You Learn from a Game?". In *Wise Choices: Decisions, Games, and Negotiations*, edited by Ralph 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.
- Paxson, E. W. "The Sierra Project -- a Study of Limited Wars." (Santa Monica, CA: RAND Corporation, B-41 (WITHDRAWN), 1958).
- Paxson, Edwin W. "Computers and National Security." (Santa Monica, CA: RAND Corporation, P-4728, 1972).
- . "War Gaming." In *Military Operations Research*, edited by Bernard O. Koopman: Operations Research Society of America, 1963.
- 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, Michael Markowitz, Albert Nofi, Christopher Weuve, Julie Loughran, and Marcy Stahl. "Gaming and Shared Situtational Awareness." (Alexandria, VA: Center for Naval Anlysis, 2000).
- Perla, Peter, Michael Markowitz, and Christopher Weuve. "Game-Based Experimentation for Research in Command and Control and Shared Situational Awareness." (Alexandria, VA: Center for Naval Analyses, 2002).
- Perla, Peter P. "The Art and Science of Wargaming to Innovate and Educated in an Era of Strategic Competition." In *King's College London Wargaming Network Lecture*. London, UK, 2018.
- . *The Art of War Gaming: A Guide for Professionals and Hobbyists*. Edited by John Curry. 2nd ed.: History of Wargaming Project, 2011.
- . "Now Hear This—Improving Wargaming Is Worthwhile—and Smart." *Proceedings Magazine*, January 2016 2016.
- Perla, Peter P. , and ED McGrady. "Why Wargaming Works." *Naval War College Review* 64, no. 3 (2011).
- Perry, Walter L., and Marc Dean Millot. "Issues from the 1997 Army after Next Winter Wargame." (Santa Monica, CA: RAND Corporation, MR-988, 1998).
- Pettyjohn, Stacie L., and David A. Shlapak. "Gaming the System: Obstacles to Reinvigorating Defense Wargaming." *War on the Rocks*, February 18 2016.
- Pettyjohn, Stacie, and Becca Wasser. "The Promise of Structured Strategic Wargames: Moving Beyond the Seminar." In *International Studies Association Annual Conference*. San Fransciso, 2017.

- . "The Promise of Structured Strategic Wargames: Moving Beyond the Seminar." In *International Studies Association*. San Francisco, 2018.
- Popper, Karl. *The Logic of Scientific Discovery*. New York, NY: Routledge, 1992.
- Pournelle, Phillip E. "Can the Cycle of Research Save American Military Strategy?" *War on the Rocks*, 2019.
- . "Designing Wargames for the Analytic Purpose." *Phalanx* 50, no. 2 (2017): 48-53.
- 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.
- "Research Note: War Games." *RANDom News*, January 21 1949, 2.
- Rittel, Horst W. J., and Melvin M. Webber. "Dilemmas in a General Theory of Planning." *Policy Sciences* 4, no. 2 (1973): 155-89.
- Rubel, Robert C. "Epistemology of War Gaming." *Naval War College Review* 59, no. 2 (2006): 108-28.
- "Rules for an Aggregated War Game." (Santa Monica, CA: RAND Corporation, RM-1046-2, 1954).
- Sampat, Elizabeth. *Empathy Engines: Design Games That Are Personal, Political, and Profound*. CreatSpace Independent Publishing Platform, 2017.
- Schneider, Jacquelyn G. "Cyber Attacks on Critical Infrastructure: Insights from War Gaming." *War on the Rocks*, July 26 2017.
- . "What War Games Tell Us About the Use of Cyber Weapons in a Crisis." edited by Council on Foreign Relations, 2018.
- Schrage, Michael. *Serious Play: How the World's Best Companies Simulate to Innovate*. Boston, MA: Harvard Business School Press, 2000.
- Seawright, Jason. *Multi-Method Social Science: Combining Qualitative and Quantitative Tools*. New York, NY: Cambridge University Press, 2016.
- Senge, Peter M. *The Fifth Discipline: The Art and Practice of the Learning Organization*. Revised and Updated edition ed. New York, NY: Doubleday, 2006.
- House Armed Services Committee, Subcommittee on Tactical Air and Land Forces. *Detering Russian Aggression in the Baltic States: What It Takes to Win*, March 1, 2017 2017.
- Shlapak, David A., and Michael W. Johnson. "Outnumbered, Outranged, and Outgunned: How Russia Defeats Nato." *War on the Rocks*, April 21 2016.
- . "Reinforcing Deterrence on Nato's Eastern Flank: Wargaming the Defense of the Baltic." (Santa Monica, CA: RAND Corporation, RR-1253, 2016).
- Shubik, Martin. "On Gaming and Game Theory." (Santa Monica, CA: RAND Corporation, P-4609, 1971).
- Shubik, Martin, and Garry D. Brewer. "Models, Simulations, and Games--a Survey." (Santa Monica, CA: RAND Corporation, R-1060-ARPA/RC, 1972).
- Simpson Jr., William L. "A Compendium of Wargaming Terms (Updated)." Military Operations Research Society Wargaming Community of Practice, 2017).
- Smith, Kevin B. "Typologies, Taxonomies, and the Benefits of Policy Classification." *Policy Studies Journal* 30, no. 3 (2002): 379-95.
- Specht, R. D. "War Games." (Santa Monica, CA: RAND Corporation, P-1041, 1957).
- Staff, Chairman of the Joint Chiefs of. "Officer Professional Military Education Policy." Washington, DC, 2015.

- Staff, Joint Chiefs of. *Joint Publication 5-0: Joint Planning*. Washington, DC: Joint Chiefs of Staff, 2017.
- Svet, Oleg, and Garrett Heath. "How the Joint Staff Calculated a Defense Program's Return on Investment." *Defense One*, 2018.
- Vickers, Michael, and Robert Martinage. "Future Warfare 20xx Wargame Series: Lessons Learned Report." (Washington, DC: Center for Strategic and Budgetary Assessments, 2001).
- "War Game." *RANDom News*, February 18 1949, 3.
- "War Games." *RANDom News*, October 1 1948, 3-4.
- "War Games." *RANDom News*, October 29 1948, 2-3.
- Wasser, Becca, Jenny Oberholzer, Stacie Pettyjohn, and William Mackenzie. "Gaming Grey Zone Tactics." (Santa Monica, CA: RAND Corporation, RR-2915-A, 2019).
- 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.
- . "Rand Briefings to the Air Force Advisory Group 21 October 1963." (Santa Monica, CA: RAND Corporation, AR-104, 1963).
- . "Trends in Military Gaming." (Santa Monica, CA: RAND, P-4173, 1969).
- . "War Gaming Methodology." (Santa Monica, CA: RAND, RM-2413, 1959).
- . "War Gaming Methodology: Sierra near East Series." (Santa Monica, CA: RAND Corporation, D-4926-PR, 1958).
- . "War Gaming: Two Methods Used in Sierra." (Santa Monica, CA: RAND Corporation, D-4332-PR, 1957).
- Weuve, Christopher A., Peter P. Perla, Michael C. Markowitz, Robert Rubel, Stephen Downes-Martin, Michael Martin, and Paul V. Vebber. "Wargame Pathologies." (Arlington, VA: CNA, 2004).
- Wight, Colin. "Philosophy of Social Science in International Relations." In *Handbook of International Relations*, edited by Walter Carlsnaes, Thomas Risse and Beth A. Simmons, 29-56. London, UK: SAGE Publications Ltd, 2013.
- Williams, Bob, and Richard Hummelbrunner. *Systems Concepts in Action: A Practitioner's Toolkit*. Stanford, CA: Stanford University Press, 2011.
- Wilson, Andrew. *The Bomb and the Computer*. London: Barrie and Rockliff 1968.
- Wittgenstein, Ludwig. *Philosophical Investigations*. Translated by G.E.M. Anscombe, P.M.S. Hacker and Joachim Schulte. Revised 4th edition ed. Chichester, UK: Wiley-Blackwell.
- Wong, Yuna. "Preparing for Contemporary Analytic Challenges." *Phalanx* 47, no. 4 (December 2014 2014): 35-39.
- Wong, Yuna Huh, Sebastian Bae, Elizabeth M. Bartels, and Benjamin Smith. "Next Generation Wargaming for the U.S. Marine Corps: Recommended Courses of Action." (Santa Monica, CA: RAND Corporation, RR-2227-USMC, 2019).
- Work, Robert. Memorandum, February 9 2015.
- Work, Robert, and Paul Selva. "Revitalizing Wargaming Is Necessary to Be Prepared for Future Wars." *War on the Rocks*, December 8 2015.
- Yin, Robert K. *Case Study Research: Design and Methods*. 5th ed. Thousand Oaks, CA: Sage, 2014.
- Zenko, Micah. "Millennium Challenge: The Real Story of a Corrupted Military Exercise and Its Legacy." *War on the Rocks*, 2015.