SAND2022-1439J

Experimental Wargaming with SIGNAL

Joshua Letchford[1], Laura Epifanovskaya[2], Kiran Lakkaraju[1], Mika Armenta[1], Andrew Reddie[1], Bethany L. Goldblum[3,4], Jon Whetzel[1], Jason Reinhardt[1], Vamshi Balanaga[3], Andrew Chen[3], Nathan Fabian[1], Sheryl Hingorani[5], Roshni Iyer[3], Roshan Krishnan[3], Sarah Laderman[3], Manseok Lee[3], Janani Mohan[3], Michael Nacht[3], Soravis Prakkamakul[3], Matthew Sumner[1], Jake Tibbetts[3], Allie Valdez[3], Charlie Zhang[3]

1 Sandia National Laboratories, 2 Institute for Defense Analyses, 3 University of California, Berkeley, 4 Lawrence Berkeley National Laboratory,5 Lawrence Livermore National Laboratory

1 INTRODUCTION

Conflict scholars and analysts have placed an increased focus on the role that grey-zone (Gartzke & Lindsey, 2019; Lanoszka, 2016; Lovelace, 2016; Mazarr, 2015), cyber (Schneider 2019; Kreps & Schneider, 2019;Uribe et al., 2020; Fischerkeller, 2017; Libicki, 2018), and low-level conventional (Mearsheimer, 1985; Wirtz, 2018; Mueller, 2018) actions can have in shaping strategic outcomes, along with a renewed interest in the impacts of potential limited nuclear actions (Colby, 2018; Warden, 2018; Larsen & Kartchner, 2014; Halperin, 1961) in a new era of strategic competition. Questions of 'integrated strategy,' escalation control, and cross- or multi-domain conflict interactions have expanded the choices faced by planners and decision makers (Gartzke & Lindsay, 2019), and confounded the assessment of how a nation might best allocate and apply military means to achieve a given end. While a historical view of conflict is still a critical starting point for studying future conflicts, new weapon capabilities (e.g., cyber attacks and cyber-enabled disinformation campaigns, innovations in hypersonic weapons, precision, navigation and guidance) pose challenges to drawing relevant lessons from wars fought decades or centuries ago (Lieber & Press, 2016). The increase in complexity of strategic choices and a lack of relevant historical analogs motivates the design and development of new tools to aid

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

military planners, decision-makers, and policymakers that face a challenging modern strategic reality.

Wargames are a common tool for investigating complex conflict scenarios and have a long history of informing both military and strategic studies (Caffey, 2019; Oberholtzer et al, 2019; Perla & Curry, 2011). Seminar-style wargames (Ministry of Defence, 2017) are often used to investigate scenarios and explore strategic and tactical challenges that planners may face. Historically, many of these games have not rigorously collected quantitative data, and they have often been unique, built primarily for exploration rather than developing firm analytical conclusions. A recent tabletop exercise conducted by the International Institute for Strategic Studies (IISS) involving an artificial intelligence arms race exemplifies this approach (Fitzpatrick, 2019). Longitudinal analyses of seminar-style wargames have identified interesting potential trends and lessons learned (Pauly, 2018; Schneider, 2017). While these analyses are valuable, the games in the datasets are almost never repeated using exactly the same scenario and starting conditions, preventing experimental inference and valid tests of the statistical significance of the outcomes. New wargaming approaches that employ the principles of experimental design are needed to enable an objective basis for the analysis of conflict escalation (Lin-Greenberg, 2018; Reddie et al., 2018; Schneider, 2019b; Valeriano and Jensen, 2019).

Experimental wargames (EWGs) are wargames executed using rigorous experimental techniques to enable hypothesis testing (Reddie et al., 2018). By combining the standard elements of a traditional wargame with experimental design principles, EWGs are focused on data-driven, systematic, and quantitative exploration of player decisions, actions, and interactions taken from highly-controlled scenarios at scale. They produce rich datasets that can be analyzed using a variety of quantitative methods to measure experimental outcomes and identify patterns of interest. An EWG complements and can work in concert with traditional wargames, which have the primary goal of qualitatively exploring the complexities of a scenario to provide insight to sponsors, organizers, and players (Bartels, 2019; Perla & McGrady, 2011; Rubel, 2006).

Strategic Interaction Game between Nuclear Armed Lands (SIGNAL) is an EWG platform built in both table-top (SIGNAL-Board) and digital (SIGNAL-Online) formats. SIGNAL was designed to produce experimentally valid data for quantitative analysis and theory development. To date, the SIGNAL EWG has been played hundreds of times by thousands of players from around the world, creating the largest database of wargame data for academic purposes known to the authors.

The remainder of this section briefly outlines some of the existing data-generating processes used in social sciences. Section 2 outlines the general experimental design concepts that were used as guidelines for development. Section 3 describes the SIGNAL wargame, including elements of the game, gameplay mechanics, and the data collection processes. Section 4 describes the game design choices and connections to experimental principles. Finally, Section 5 provides some lessons learned and a path forward for future experimental wargaming efforts.

1.1 Related Work

In this section, we outline existing data-generating processes used in social science and propose where experimental wargaming might address the challenges associated with existing methodological approaches. Discussions concerning the strengths and weaknesses of extant social science methods are not intended to be exhaustive, but rather to introduce where EWGs might fit in the broader suite of tools to examine strategic questions.

Case-based research has been used to consider a large number of important questions pertaining to foreign policy and security studies (Bennett, 2007). Whether engaging with deviant, most-likely, or most-similar cases, case-based research allows for process tracing, appropriate coding of historical events for subsequent analysis, and refining theoretical arguments. There are, however, challenges for case-based research in terms of replicability as well as the external validity of the analysis (Maoz, 2002).

Capturing large quantities of historical data for statistical analysis is standard in studies of conflict. In this literature, scholars collect historical data related to wars and militarized disputes. The Correlates of War (COW) project, for example, publishes multiple datasets related to conflict, including the Dyadic Military Interstate Dispute dataset used to classify states that threaten, display, or use force against another state from 1816-2010 (Maoz, 2019). These data include characteristics of the nations such as their economic and military strength, the volume of trade between warring states, and alliance relationships. The International Crisis Behavior Dataset takes a similar approach, coding 482 crises and 1,065 crisis actors from 1918 to 2016 (Brechner et al., 2020). These data are subsequently used to understand the impact of different factors on conflict, such as economic interdependence (Barbieri, 1996) or the proliferation of nuclear weapons (Asal & Beardsley, 2007). While both case-based and historical methods are useful, they are limited to conflicts that have occurred, which may differ from future conflicts. Additionally, as is the case with any observational study, researchers lack the precise methodological control to identify causal factors (Hyde, 2015). And perhaps the most challenging aspect is that data is often unavailable for specific types of research questions, particularly those addressing how emerging military capabilities might reshape strategic stability.

In the absence of empirical data, scholars have turned to a variety of modeling approaches to address a series of security challenges, building on insights derived from game theory. For example, formal models provide researchers with a means to "provide a more precise statement of the relations among various concepts in a mathematical form" (Snidal, 2004). These mathematical formulations are then used to make logical inferences about relationships in the real world. Computer-based modeling approaches further extend formal models to scale, accounting for increasing levels of complexity. However, these models are commonly criticized for their simplifying assumptions concerning actor rationality and perfect information, for limiting analyses to small numbers of actors, and for the lack of empirical evidence that support these assumptions (Lebow & Stein, 1989; Schmitter, 2016). Agent-based models, a commonly used class of computational models, may also suffer from calibration and validation problems (Heath, Hill & Ciarallo, 2009; Macal, 2016).

Vignette-based survey experiments address some of these challenges by creating bespoke experimental conditions to test a specific research question and provide the researcher with considerable flexibility in terms of the types of scenarios that they can interrogate (Press, 2013; Sagan and Valentino, 2017; Wess & Dafoe, 2019). These approaches do not impose rationality on research subjects—instead allowing for a behavioral analysis of how subjects might address strategic questions (Mullinix et al., 2015). Concerns remain regarding whether appropriate research subjects are sampled, whether a research subject has internalized the experimental treatment, whether survey experiments allow subjects to consider the intricacies of a particular scenario, and whether the respondent is sufficiently immersed within the experiment environment (Pauly, 2018).

In academia, these methods provide the analytical foundations for much of the existing scholarship. In policy-oriented research, however, wargames—often seminar-based wargames—have also been used as data-generating processes to consider how new types of military capabilities or changes in military doctrine might influence military and strategic engagement (Ministry of Defence, 2017; Perla & McGrady, 2011). Wargames offer researchers broad flexibility to investigate the complexity associated with decision-making in a security context within an immersive environment that can be used to interrogate various types of human behavior, from competition to cooperation. This method, however, has its own challenges. As most existing wargame designs rely on white cell adjudication, researchers cannot precisely replicate scenario conditions and introduce sources of bias. Many wargame designs are also often played only once—usually, with a homogenous set of players. Observing the strengths and weaknesses of wargaming methods, we suggest that combining wargames with experimental design principles—in what we describe as "experimental wargaming"—offers a useful complementary contribution to the methods outlined above.

While EWGs have their own methodological challenges, they usefully address some of those outlined above. Like survey experiments, they do not assign external attributes of rationality to research subjects. They also allow for an exploration of scenarios that do not occur "in the real world", allowing scholars to ask questions that might otherwise go unanswered. And, unlike survey experiments, respondents are immersed for long periods of time in a game environment that, while still abstract, better approximates the complexity associated with real-world decision-making. These attributes drive the pursuit of the experimental wargaming design described in this article.

2 BUILDING EXPERIMENTAL WARGAMES

We start by defining the standard elements of a wargame (Perla & McGrady, 2011):

- **Objectives** – The objectives define the scope and basic purpose of the game.
- **Scenario** – The scenario provides players with the context needed for decision-making. It includes both the goals that players are instructed to achieve and the resources that they have available to attempt to achieve these goals.
- **Database** – This is the supplemental (quantitative) information provided to the players. This information is usually provided to ensure players have the necessary information to make meaningful decisions.
- **Models** – Models are the way that wargames simulate reality via methods such as mathematical expressions. These provide the adjudicators a principled way to evaluate outcomes, but are at best an approximation of reality.
- **Rules and procedures** – These build a common understanding between the participants on how the scenario, databases, and models will be used in the wargame and how the participants are allowed to interact with each other.
- **Players** – The players of a wargame capture not only the sides represented in the game, but also the characteristics and backgrounds of the participants that fill these roles.

Next, consider that our objective is to study realistic human decision-making in strategic contexts in a way that allows identification of relationships between potentially causal factors and outcomes of interest, which could be a behavior, a decision, a motivation, etc. By reducing as many extraneous influences as possible on player actions and by comparing the relative differences in the outcome of interest between experimental conditions, we can answer questions related to correlation and causality. In an experiment, we identify these relationships by instrumenting the hypothesized causal factor as a ‘condition,’ comparing it against other neutral or control conditions, and holding all else equal. The following set of experimental design characteristics may be employed with wargames:

- **Randomization** – Participants are randomly assigned to either the treatment or control groups. This ensures that participants have an equal probability of being assigned each set of starting conditions and serves to lessen the impact of prior knowledge or participant biases on the experimental outcomes.
- **Replicability** – An experiment must be repeatable to allow for data to be accumulated across the different experimental conditions. Such an approach incorporates a wide range of human variability while enabling generation of a sufficiently large data set to make statistically valid conclusions.
- **Controllability** – The experiment must be controllable to allow for systematic manipulation of the independent variable, thereby facilitating the establishment of causal relationships. By changing one variable at a time, resulting changes in player behavior and/or conflict patterns may be attributed to this single change.
- **Instrumented** – The experiment must allow for the capture of player behavior, actions, and other data that provide information about the experimental tasks. To the greatest extent possible, the data collection process should be designed to reduce the need for human-in-the-loop interpretation and coding, as is often the case with transcription of oral conversations.
- **Neutrality** – The experiment should not bias participant behavior with regard to the research question. A variety of behavior and action options should be made available to participants—some that support the hypothesis and some that do not. Interaction between

the experimenter and participants should be limited to reduce potential impacts from experimenter bias and to avoid inadvertent communication of expectations to participants.

In an EWG, participants take part in an experimental task which is conducted as a sequence of activities defined by the experimenter to explore a specific research question or set of hypotheses. Prior to the EWG, participants are randomly assigned different starting conditions, which form the basis for treatment and control groups in the comparative experiment. The treatment conditions are varied relative to the control setting in specific and tightly controlled ways. Before, during, and after the experimental task, data are collected on participant demographics and behavior. The conditions are such that a quantitative evaluation of the relative difference of various outcomes between conditions can allow the use of established statistical methods for hypothesis testing (Maxwell, Delaney, & Kelley, 2017).

While experimental principles offer a framework for wargame design that provides increased rigor and reduces potential biases, a key challenge to EWG design lies in balancing player engagement and experimental rigor – navigating an accuracy-simplicity tradeoff (Hernandez, McDonald, & Ouellet, 2015; Schneider, 2017). While designing SIGNAL, we found that conflicting demands between these design principles required tough decisions. A major tradeoff we faced was between the fidelity of the scenario and the experimental demands, specifically in keeping the experiment simple enough to establish causal relationships. Below, we discuss SIGNAL and the decisions and tradeoffs made in its design, but note that other EWGs might make different decisions when considering these tradeoffs.

3 THE SIGNAL WARGAME

We designed the SIGNAL EWG as a platform to study player decisions and strategy, with a focus on conflict escalation between nuclear-armed states in a multipolar world. The research question was whether, in the context of an experimental wargame, conflict escalation, as measured by in-game likelihood of nuclear use, was affected by the inclusion of tailored output nuclear weapons such as high-precision low-yield and enhanced electromagnetic pulse weapons. A variety of arguments have been made about the potential impact of tailored output weapons on the risk of nuclear war. Some argue that the credibility of tailored output weapons would strengthen deterrence among the major nuclear powers. That is, tailored output weapons will reduce the gap in nuclear capabilities (DoD 2018; Dodge 2018; Lieber and Press, 2009; Williams & Lowther, 2017), send a firm signal to adversaries (Harvey 2017; O'Hanlon 2018), and thereby prevent adversary leadership's miscalculations. However, such characteristics of tailored output weapons could generate other systemic consequences that could serve to escalate a nuclear conflict. Others argue that tailored output weapons may incentivize military commanders to favorably reconsider nuclear options (Mecklin 2017), blur the threshold between conventional and nuclear weapons use (Narang 2018; Weber & Parthemore 2019), or misunderstand an adversary's intentions, leading to a new arms race, and thereby increasing the likelihood of nuclear use (Coyle & McKeon 2017; Doyle 2017; Trenin 2019; ZhaoTong 2019).

While SIGNAL (or similar games) could be used to study a wide range of hypotheses, many of the design elements were explicitly driven by this research question. As a tool for inquiry, SIGNAL was developed as a complement to traditional wargaming to address academic needs unmet by game theory and empirical conflict research while not attempting to compete in areas where those platforms excel. For example, SIGNAL does not provide the immersive experience or battlefield realism of a tactical wargame, or the interactions with professional policymakers that are often a key element of a seminar-style wargame. Instead, SIGNAL is intended to be a laboratory for experimentation, data collection, and hypothesis-testing about the dynamics of armed conflict and strategic interactions. Although current incentives and instrumentation in SIGNAL are focused on studying conflict escalation, we believe that many of the techniques employed here could be modified to answer a wide range of related questions on grey-zone, cyber, nuclear, and low-level conventional conflict.

Beyond the single metric of nuclear use, SIGNAL data can be used to analyze conflict escalation and de-escalation ladders by looking at the choices players make in subsequent turns. For example, if two players engage in conventional conflict through infantry attacks on each other, then we may classify that round as “conventional conflict”. However, if in the next round one of those players launches a nuclear attack, then that round might be classified as “nuclear attack”. Seeing how these classes of behavior evolve over successive rounds can give a sense of escalation or descalation and allow for an analyst to investigate the antecendents to escalation or deescalation.

Gameplay in SIGNAL takes place in a fictional world with three countries. The countries are intentionally abstract, facilitating exploration of how players with specific characteristics respond to discrete strategic challenges, as opposed to how they feel a particular country would act in a specific scenario. Players can take economic, military, and diplomatic action through provided capabilities, each with defined costs and effects. Players follow a set routine of staging and executing these capabilities, but the game otherwise places few limits on their interactions. Players can create their own diplomatic and economic agreements as long as the technical rules of play are not violated. Military alliances can be formed, arms control agreements negotiated, and nuclear umbrellas created at the behest of the players.

SIGNAL was designed to be executed on two platforms: a board game and an online game. The in-person nature of SIGNAL-Board allows closer interpersonal interactions among players and consequently provides a more socially immersive and dynamic experience than the online game. However, data gathering is more challenging with SIGNAL-Board and important verbal exchanges can be missed in the player/rapporteur data-recording process (See Section 3.2). While personal relationships may be weakened in anonymous online interactions, SIGNAL-Online logs all player actions automatically, and interactions among players occur through an in-game chat window that are recorded for later analyses. There are potential advantages to the relative anonymity of SIGNAL-Online as it avoids biases that could be introduced by existing relationships between in-person players or expectations of how a player from a certain organization or role should act. In general, the design of SIGNAL between the two platforms is the same, but cases where differences occured will be denoted by referring to SIGNAL-Board or SIGNAL-Online, as appropriate. Specific design elements of note will be highlighted in this document. Readers interested in more detail can refer to the full set of rules available in the SIGNAL manual (Armenta et al, 2020).

3.1 A Brief Summary of SIGNAL Game Play and Mechanics

As shown in Figure 1, SIGNAL is played on a map that contains three major player countries, named for their colors: Orange, Green, and Purple. The map is divided into small hexagonal regions (“hexes”) representing spatial cells on the game map. The contents of most hexes within each country represent control of territory denoted by player color, but some also offer the player resources or infrastructure. Smaller, independent non-player countries (shown in grey) consist of a few hexes each, typically with a single resource. Each major player country has a certain population size , and each player controls an initial supply of currency.

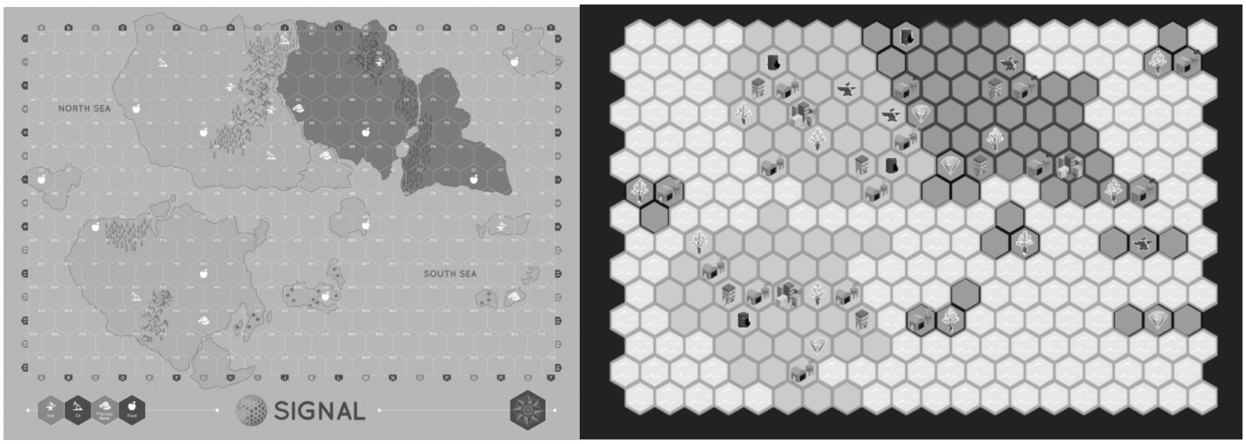


Figure 1 The maps of SIGNAL-Board (left) and Signal-Online (right).

Players are provided with a range of military and economic capabilities to interact with other players in the game environment including trade mechanisms, military actions, chat, and methods of developing infrastructure. In addition, depending on the conditions being tested, players may be given access to various types of nuclear weapon capabilities. These weapon capabilities may be used to threaten or negotiate with other players, and may also be employed for military action. Capabilities, starting resources, and geography vary by country. Most notably, Orange lacks nuclear capabilities, but was bestowed other non-nuclear advantages.

The game consists of five rounds of play, each divided into three phases. In the first phase of each round (the "Signaling Phase"), players simultaneously signal their intentions by placing tokens on hexes to indicate where they *may* take an action in the next phase. Each player also places Action Cards face-down in front of them, representing the actions that could be executed on one of the hexes containing a token. The staging of each Action Card is accompanied by a nominal monetary cost to the player. The backs of the cards show the category of action to be played (nuclear military action, conventional military action, or infrastructure development), but not the specific action type. Up to eight Action Cards and up to twelve Signaling Tokens can be staged during the Signaling Phase. However, at this point, players are not required to tie any of the cards to a particular token. Communication is allowed between players throughout the game, but, in this phase, players can make diplomatic and economic agreements with other players.

Next, SIGNAL transitions to the second phase (the "Resolution Phase") in which players implement actions using the Action Cards and Signaling Tokens that they previously staged in the Signaling Phase. This phase is organized into a series of rotations where each player may combine one of their staged Action Cards with one of their staged Signaling Tokens to create an effect within the game – this is called a turn. Executing an action carries additional costs beyond those of simply staging the Action Card. Some actions have a random element associated with them, such as a chance of failure in executing a military strike, which is resolved at this point. A player may also choose to pass and not play a card. Once all players have had an opportunity to take a turn, a new rotation starts. Once all players have "passed" within a single rotation, the Resolution Phase ends and the game moves to the Upkeep Phase.

In the Upkeep Phase, the consequences of prior actions are calculated. First, player resources, total population, and income are calculated – reflecting changes in access to resources based on trades or hex ownership. Then, any player who has more population than can be supported by their food supplies is forced to reduce their infrastructure (each of which represents a specific amount of "population") until this imbalance is resolved. During each Upkeep Phase, players also calculate their current score. Players are given the same set of three goals for every game: 1) Maximize the value of infrastructure controlled (the infrastructure goal), 2) Maximize the number of resources controlled (the resources goal), and 3) Minimize the number of home country hexes that are lost (the defense goal). A player's score depends on how well they perform in each of these three categories relative to other players in the game. At the end of the fifth round of gameplay, the player with the highest score is the winner. A detailed description of SIGNAL rules and game mechanics is provided in (Armenta et al., 2020).

3.2 SIGNAL Data Collection

Data collection in SIGNAL-Board takes place at the end of the Signaling Phase, after each turn in the Resolution Phase, and at the end of the Upkeep Phase. This is tracked by data sheets that are filled out by participants (See Appendix A). The data collected at the end of the Signaling Phase provides a snapshot of the state of each player's Signaling Tokens and Action Cards, as well as a brief description of any significant diplomacy (including any economic agreements) that occurred between the players. The data collected in the Resolution Phase captures information about each action taken: the action, when and where it occurred, and the outcome (for actions that have an element of randomness). The data collected at the end of the Upkeep Phase captures a summary of the overall game state, including resource counts, income, and each of the player's current scores.

In the SIGNAL-Online setting, a superset of the data recorded in the board game are automatically collected. In addition to the information described above, the sequence and timing of Signaling Token placement and Action Cards staging are recorded, as are data pertaining to Signaling Tokens or Action Cards that are staged and subsequently removed before the end of the Signaling Phase is reached (e.g., because of a threat that was successful in changing the behavior of another player). Detailed chat logs are also recorded.

For the first SIGNAL data collection campaign (which concluded on Dec 20, 2019), the final dataset consisting of 32 Signal-Board games and 447 Signal-Online games was obtained after removing games with project members or bots, and games that did not complete at least three rounds. This dataset was used for all figures in this document. For these games, there were consistently two nuclear-armed players and one non-nuclear player, though the specific nuclear capabilities of the players varied based on the condition.

4 DESIGNING SIGNAL AS AN EXPERIMENAL ENVIRONMENT

In this section, we discuss a subset of the design decisions that were made for SIGNAL and how they were motivated by SIGNAL's purpose as an experimental environment for studying conflict escalation and nuclear use.

4.1 Scenario Concepts Design

Scenarios were intentionally designed to create situations where players were motivated by game objective and win conditions to come into conflict with each other, but were not forced to do so. To achieve this, the map and other elements were arranged to incentivize conflict without restricting the players to a particular decision path. Below are some details on the map and conflict drivers implemented towards these goals.

4.1.1 Map Design

We highlight three main elements of map design: the choice of a hexagonal tessellation, the design of the major countries, and the design of the minor states.

4.1.1.1 Hexagonal Tessellation

We chose a hexagon tessellation of the space for two reasons. First, the tessellation gave us an appropriate level of abstraction for our strategic level of focus. Hexagons, specifically, were employed as they make distance functions straightforward (the distance between the center of all adjacent hexes is the same) as opposed to square grids that require either accepting irrational numbers to handle diagonal movement, operating under "Manhattan distance" (where you count the distance without diagonals), or warping the projection such that diagonal movements are treated as having the same distance as orthogonal movements.

4.1.1.2 Design of Major Countries

To prevent a player's choices from being driven by their perception of how current (or historical) real world states would act, a fictional world was constructed using three primary countries without strong geographic parallels to existing superpowers. Testing of early designs revealed that even with fictional countries, superficial resemblance to real world countries was sufficient to alter player behavior. Thus, the SIGNAL map was constructed to be as distinct as possible from modern geopolitical boundaries, both in terms of landmass shapes and iconic colors (e.g., avoiding a "red" country that players may associate with an adversary nation).

4.1.1.3 Design of the Minor States

The map also included a number of smaller, independent "minor states" that were not controlled by any player at the start of the game. Each minor state holds a natural resource that is of value to the players. Though the minor states are not controlled by any particular player, each player-controlled state begins the game allied with one minor state, thereby dominating access to its land and resources. Player interaction with minor states may involve "soft" force (where the player builds military infrastructure in the minor state) or military action (where the player invades the minor state using conventional or nuclear military capabilities). Due to their value and locations, minor states provide a possible catalyst for conflict, allowing for low-level conflicts to potentially develop between the players without existential threats. Figure 2 shows the distribution of Signaling Tokens placed on each hex for the 479 SIGNAL-Board and SIGNAL-Online games in the dataset. The number corresponds to the total count of Siganling Tokens in that location and the pie chart shows the fraction of this count that belonged to each player. As shown in Figure 2, the grey minor states (especially the ones centrally located near all three player's borders), received a significant fraction of the total number of Signaling Tokens (39%) while representing approximently 17% of the hexes on the gameboard.

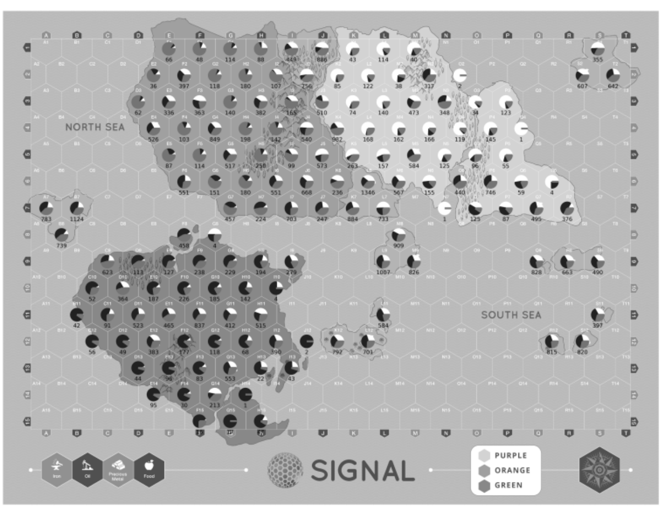


Figure 2 The distribution of where Signaling Tokens were placed in the dataset. Each hex displays the total count of Signaling Tokens and the fraction placed by each of the three players from the entire data set. As noted in the legend, the darkest shade corresponds to Green, the middle shade to Orange and the lightest shade to Purple.

4.1.2 Design of Conflict Drivers

Next, we highlight two specific elements, resource design and the win condition design, both of which are intended to incentivize conflict between players.

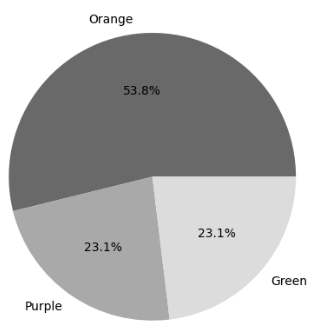
4.1.2.1 Resource Design

We designed the resources in SIGNAL to provide ongoing utility to the players in addition to impacting endgame scoring. Resources were intended to be both a catalyst of conflict and a driver of diplomacy and trade. To enable trade, we needed a variety of resources and the marginal value of these resources to differ between players. SIGNAL's design included four different resources: oil, iron, precious metal, and food. Access to oil and iron lowered the cost of specific actions if a player had access to them. However, they were designed to not be consumed upon use so there would be no direct value in controlling more than one source, and therefore would be more valuable to a player lacking the resource than to the player with a surplus. A third resource, precious metal, directly influenced a player's income. Thus, it did not have decreasing marginal value (beyond the decreasing marginal value generally inherent in money), which meant that these were more often sources of conflict. The fourth resource, food, was a requirement for supporting population and infrastructure. Infrastructure was necessary to generate income (and thus the ability to take actions within the game) and to increase end-game scoring. Since resources were required to develop infrastructure, player ability to develop infrastructure was limited unless they expanded their area of influence – which was a source of conflict. Finally, we varied the distribution of initial resources to balance the starting abilities of the three players, particularly with respect to the availability of nuclear forces. We did this by providing a surplus of iron and oil to the non-nuclear power and a lack for one of these resources to each of the nuclear powers. This provided players a potential economic incentive for making alliances at the start of the game.

4.1.2.2 Win Conditions Design

The three goals assigned to players (infrastructure, resources, and defense) provide a natural path to conflict—both directly (through competition for resources) and indirectly (through needing food to expand infrastructure). The defense goal helps to maintain conflicts once they start, by providing an incentive for players to repel and retaliate against attacks that impact their homelands. Each goal has the same number of points potentially available, and scoring is assessed relative to the performance of other players. This was implemented to encourage players to simultaneously pursue multiple goals to win the game, as it is not possible to win by dominating a single goal at the expense of the other two.

Player goals were symmetric and public to allow the game to be played multiple times by the same player, and to minimize confounding factors. If player goals were private, some degree of randomization would be necessary to allow for multiple playthroughs. Private, randomized player goals would introduce a potential confounding factor in our analysis. Symmetric goals also guarantee that players' aims would be in direct conflict with each other, maximizing the intersection of player interests with the minimum cognitive overhead. Finally, while our research interests do not place significance on the winner of each game, they do place heavy importance on player engagement with the win conditions, which provides the incentives necessary for the goals to influence behavior. Win rates, instead of individual wins, are used to assess balance in the game. To make their position as a non-nuclear nation more competitive, Orange was assigned a slightly larger and more flexible set of starting resources and a bonus point in their score. To evaluate how well we achieved balance, we examined the overall win rates of the three sides in both SIGNAL-Online and SIGNAL-Board. This design resulted in each country winning an approximately equal fraction of the games in SIGNAL-Online, as shown in Figure 3. In contrast, in SIGNAL-Board, Orange won considerably more often than either of the two nuclear-armed players, although it is unclear if this difference is due to significant differences in how players approached the board game platform or the smaller sample size (i.e., 26 Signal-Board games v. 461 Signal-Online games).



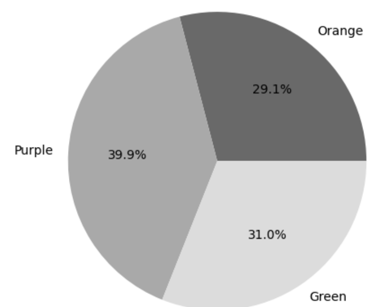


Figure 3 Fraction of games won by each player color for SIGNAL-Board (left) and SIGNAL-Online (right).

4.2 Outcome Arbitration Design

We designed the general rule structure of SIGNAL to address two high-level concerns. First, we avoid the use of the traditional white cell adjudication, avoiding potential experimenter bias and allowing data to be collected at-scale. Second, to ensure neutrality, there were no nuclear-use norms communicated to players.

4.2.1 Automated Rules-Based Arbitration

SIGNAL players are provided with instructions – a fixed ruleset that governs behavior within the game world. This rules-based system was adopted, instead of white cell adjudication, to avoid the possibility of both intentional and unintentional bias by the adjucation team, promoting neutrality and replicability of the game execution. It also enabled us to run SIGNAL-Online at scale. Developing a rule set that was complete enough to avoid the need for referees, but short enough for players to learn quickly, required simplifications and abstractions of many player actions. Additional discussion provided in Section 4.4.

4.2.2 Lack of Imposed Norms Arbitration

The research question is focused on conflict escalation and thus ties to concepts of deterrence theory – the theory of how to deter an action by an antagonist. An important component of deterrence is creating and communicating future normative costs for proscribed actions. When deterrence fails, the costs paid by the attacker are not just the capital costs of the use of the weapon, but may also include the retaliatory costs imposed by the (failed) deterrer. To minimize experimenter bias, we designed SIGNAL so that the norms around nuclear use, retaliatory actions, and costs had to be set by the players. If the game penalized nuclear use with a high retaliatory cost in a mechanical way, it could artificially bias the players against nuclear use and impact the very behavior we were attempting to study. Rather than imposing the structure of deterrence through the game rules, we intentionally chose to let the players develop their own nuclear norms and deterrence systems. SIGNAL does not suggest or presuppose a set of norms around nuclear use, and the in-game cost of nuclear use is similar to the cost of using other military capabilities. However, SIGNAL does provides numerous avenues for the players to develop their own norms, allowing for diplomatic, economic, and military consequences, as well as open-ended agreements between players.

4.3 Round Structure Design

Each of the three phases of the game was designed to fulfill specific needs in the experiment. The Signaling Phase provides an open forum for players to engage in trade, diplomacy, and threats and to plan military actions; it also gives the experimenter the opportunity to collect data on these actions. The Resolution Phase was influenced by our data collection needs and to minimize the amount of complexity in resolving actions. The Upkeep Phase was designed for data collection and accounting purposes and is automated in SIGNAL-Online. The design considerations for the Signaling Phase and Resolution Phase are addressed in turn below.

4.3.1 Signaling Phase

The structure of the Signaling Phase provides elements of uncertainty and flexibility to the players. Uncertainty is introduced in the staging of Action Cards, as the displayed side of the card only shows the type of card (infrastructure, conventional weapon use, or nuclear weapon use), but not the specific card within each type. Signaling Tokens may be placed freely, but staging Action Cards imposes a cost, even if they are not used in later phases. In this way, both credible deterrence and flexibility of response carry concrete costs to the players. This allows players to negotiate (make trade deals, form alliances, create non-aggression pacts, etc.) and coerce (make threats, deter, etc.). To enable this, the Signaling Phase is unstructured, allowing all of the players to continuously make real-time updates to their Signaling Tokens and staged Action Card set in reaction to other players.

Additionally, because the players can stage any (or none) of their Action Cards and each card can be later associated with any Signaling Token (or none), players have significant flexibility in both what actions are eventually taken and where those actions are executed—even after they have committed to a set of Action Cards and Signaling Token placements. This flexibility however creates uncertainty in the other players. Because twelve Signaling Tokens but only eight Action Cards can be played, opposing players (and sometimes the player doing the signaling) do not know what and where an action will be taken until a play is made, and which signals, if any, are bluffs.

4.3.2 Resolution Phase

In the Resolution Phase, players execute their plans and effect significant changes in the underlying game state. In this phase, players make explicit the connection between specific Signaling Tokens and Action Cards (as well as revealing to other players the full details of the card that they are executing). To allow us to more clearly disentangle cause and effect, the Resolution Phase was explicitly designed such that only one player can take an action at a time. Finally, to streamline the implementation, each player action is completely self-contained. That is, any response to an action by other players must be executed as a separate action on their next turn. The purpose of this was to reduce pauses in the game due to shifts in player control.

4.4 Design of Capability Mix

Player capabilities addressed two main elements. The general mix of included capabilities was driven largely by the strategic scenario we designed around. In contrast, the specifics of how the nuclear capabilities were designed and how they related to the non-nuclear capabilities were driven by the research questions.

4.4.1 Capability Mix Provided to Players

All players were provided with the following capabilities:

- The ability to build civilian infrastructure
- The ability to build military infrastructure
- The ability to preemptively defend a hex
- The ability to perform a conventional attack on a hex
- The ability to perform a naval attack on a coastal hex
- The ability to destroy infrastructure via conventional missile strike
- The ability to perform a cyber attack on a hex to temporarily disable infrastructure

In addition to these capabilities, nuclear-armed states were always given access to traditional nuclear weapons. Players in the control group only had access to those nuclear weapons, while players in a condition with tailored-output nuclear capabilities were also given access to electromagnetic pulse (EMP) and high-precision low yield (HPLY) nuclear weapons.

Tensions in the design of SIGNAL manifested in balancing the complexity necessary to capture the elements and dynamics of strategic deterrence against the cognitive load imposed by the ruleset. The rules were simplified in dimensions that were tangential to the research question to minimize this complexity. For example, players were given only the high-level abilities described above, instead of detailed combinations of weapons and platforms, reflecting the focus on strategic decision-making, rather than battlefield dynamics.

4.4.2 NW Effect Capability Design

Traditional and tailored-output nuclear weapons differ in weapon effects. In SIGNAL, traditional NWs destroy multiple hexes, removing them for the entirety of a game. Tailored output NWs provide players alternative, but related effects: EMP affects only the infrastructure and lasts only for one round while HPLY takes out a single hex for the entirety of the game. While each of these effects are abstractions of real-world effects, the research question we were studying made details related to delivery systems less vital to model.

4.4.3 Non-Nuclear Capability Design

The non-nuclear options available to the players fulfill two roles. They provide potential rungs on the deterrence ladder, and they offer military alternatives to nuclear conflict. We have provided relatively close non-nuclear analogs to each of the NW capabilities (if each lacking in some fashion) to avoid situations in which players felt they had to use a nuclear weapon to achieve their goals. Both the nuclear and conventional military options offer a long-range attack, the ability to impact an area, the ability to remove territory from a player, the ability to deny resources to another player, and the ability to destroy or temporarily damage infrastructure. The only element present on NW cards that is completely unachievable using a conventional military card is permanent hex destruction.

To evaluate how successful we were in providing non-nuclear alternatives we looked at the distribution of Action Card usage across SIGNAL-Board and SIGNAL-Online (Figure 4). In general, this design choice appeared to be successful, as we saw significant usage of both traditional nuclear and non-nuclear options, suggesting that the choice of action was a meaningful decision for the player. Surprisingly, both EMP cards and their non-nuclear counterparts (cyber strike) were used less often than other cards. This suggests that although cyber strike was a good proxy for EMP, the effect that both provided (temporary damage) was not valued by players in the SIGNAL environment. Finally, while HPLY also had limited use, we saw a more significant employment of its closest non-nuclear proxy, Conventional Missile Strike.

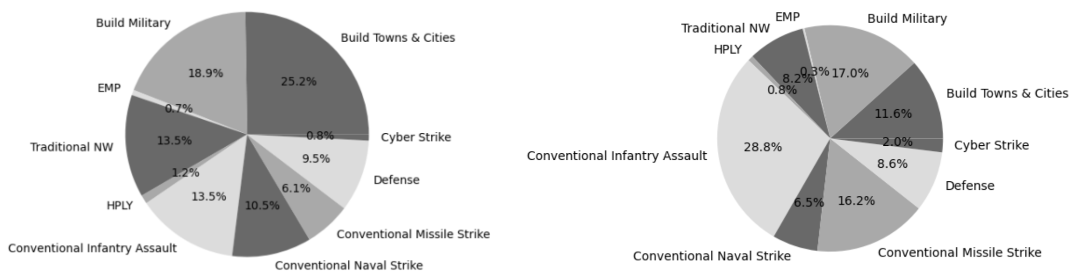


Figure 4 Distribution of Action Card usage for SIGNAL-Board (left) and Signal-Online (right).

4.5 Other Design Considerations

Finally, we summarize several other key design elements. First, we explain the rationale for why SIGNAL is a three-player game, addressing both neutrality and research question concerns. Second, we discuss the elements of randomness and their role in game mechanics. Third, we outline the design considerations for mapping game time to real time. Fourth, we discuss how we chose to handle and record player communications. Finally, we cover the demographics for which SIGNAL was designed.

4.5.1 Number of Players

SIGNAL was designed as a three-player game for several reasons, but most importantly to create the opportunity for cooperation in a competitive environment. The final scoring is zero-sum, meaning that any gains in score that a player achieves come at a matching cost to another player's score. However, the subgames formed by any pairwise interaction between two players are not non-zero sum, as two players can cooperate to take points from the third player. This game characteristic allows for cooperation to be rational even when the greater game is zero-sum. The three-player nature of the game also minimizes the burden to achieve precise explicit balancing of the scenario. If it turns out that elements (such as initial conditions) favor a particular side, the other two players would have incentive to cooperate to blunt this advantage. Finally, three-player games have been given significantly less attention in the academic deterrence literature than the two-player setting (due in part to the significant complexity introduced by moving beyond two players). Therefore SIGNAL provides a unique opportunity to study nuclear dyads in a multipolar environment.

4.5.2 Randomness

While some elements of randomness were included, we aimed to keep them both limited and intuitive to reduce the cognitive load on the players and to emphasize the value of strategy. In SIGNAL-Board, we express the probabilities to the players by requiring them to roll a given value or higher on a pair of 6-sided dice (under the assumption that most people have a reasonable intuition around the relative probabilities of dice games). In SIGNAL-Online, we provide the players with the probabilities of success directly.

The randomness included in the game mechanics falls into three categories: success randomness, effect randomness, and turn-order randomness. Success randomness – i.e., for an action to be successful the player must get a die roll within a certain range – is primarily employed in our design of conventional military actions, which have a chance of failing to achieve their objective. Effect randomness is primarily used for the traditional NW cards. Here, the primary goal of the player is guaranteed to be acomplished, but the secondary effects (additionally impacted hexes near the target hex) are randomly assigned. Appendix B shows the six different secondary effect patterns representing randomness in location of fallout. At the time of use, one of the secondary effect patterns was randomly chosen. The final type of randomness introduced arises in the turn order, where the start player is randomly chosen for each rotation. This was implemented to make it more difficult for players to have an advantage due their position in the turn order.

4.5.3 Time Scale

We designed SIGNAL to focus on strategic-level decision making within an explicit framing scenario. As part of the abstraction, no in-game time duration is tied to the game rounds, and all action cards can be completed in a single turn. This introduces somewhat unrealistic timescales (e.g., allowing the building of infrastructure or the relocation of naval units to occur in the same time as a missile strike). However, by leaving timescales abstract and unstated, the game avoids making this detail salient to the players. This also avoids creating a bias towards weapon types that could be employed quickly. The desire to understand causality for conflict escalation incentivized us to have clear sequences of actions, and our concerns over complexity and cognitive load led us to sacrifice fidelity in this area.

4.5.4 Player Communication

Much of the SIGNAL gameplay dynamics hinge on the ability of players to threaten and bargain, making communication critical. As such, the capacity to capture player communication was deemed of high importance when it was feasible to do so, such as in SIGNAL-Online. Both global (broadcast) and private (player-to-player) communication channels were provided to players in SIGNAL-Online, to minimize the incentive for players to seek external methods of communication.

4.5.5 Target Demographics

We designed SIGNAL to be approachable by a wide demographic range, from policy experts to the general public. As data collection for SIGNAL-Board required a labor-intensive effort, we targeted three primary populations: national security and military personnel, graduate students, and undergraduate students in relevant fields. These choices were driven by both a desire for populations that were comparable to other experiments (e.g., undergraduates) as well as the need for a large enough population to build up a sizeable library of games. For SIGNAL-Online, in which the data collection overhead was much lower, we targeted a broader population. While we continued to actively recruit from the above three segments, we also recruited via Amazon's Mechanical Turk and released SIGNAL-Online publicly to make it more accessible.

5 DISCUSSION

Wargames, in many forms, provide valuable insights to strategists and decision makers. However, historical wargaming approaches generally lack the characteristics that are necessary for the rigorous testing of concepts and ideas. We propose experimental wargaming as a potential new method for conflict researchers, strategy analysts, and scholars to use in concert with traditional methods of inquiry. This will allow for data-rich, quantitative analyses to help guide theory development and exploration.

5.1 A Wargaming Community Challenge

We have designed and built SIGNAL to satisfy certain requirements that enable rigorous data collection, analysis, and inquiry. Because it contains the elements discussed in Section 2, we argue that SIGNAL represents an EWG. However, the desirable attributes listed here may not be exhaustive for an EWG, and different applications may call for different emphasis. What is needed is a conversation and concerted effort within the wargaming community of practice to formalize the definitions, requirements, and guidelines for building and conducting EWGs as well as analyzing the data they produce. While SIGNAL represents one attempt at developing and executing an EWG with many lessons learned, there are many more to be learned and shared. Finally, for this is to be maximally useful to the broader community, there is value in transparency. This document as well as the ruleset (Armenta et al., 2020) are one aspect of this, but we also plan to release an anonymized version of the data set that we collected with SIGNAL, as a potential resource for others.

5.2 The Value of an Experimental Framework

Data generated in SIGNAL, informed by an experimental framework, is complementary with data generated from other experimental methodologies. Beyond that, however, the structured nature of experimental design has other advantages. One of the less obvious ones is in design reuse. Although some of the design decisions of SIGNAL were driven by the specific research questions, many were driven by the experimental methodology. While another EWG exploring different research questions might be designed differently, there are many elements of SIGNAL that can be used as an informed starting point for new EWGs.

5.3 Demographic Analysis

One of the necessary elements of SIGNAL is the ability to reach a large number of participants to gather data from a large number of games. This was achieved by developing the online version of SIGNAL which leveraged the growth of internet availability across the world, computing capabilities enjoyed by a diverse population, and cloud-based infrastructure that permits scalable data collection.

While primarily designed to allow for replicability, developing an internet-based experiment also allows us to better interrogate the effects of culture and background on behavior in national security scenarios. All participants in SIGNAL take a demographic survey before playing the game. We collect information on age, gender, geographical location, expertise in national security matters, and political affiliation. Each of these properties can be investigated for their independent and combinatorial effect on participant behavior within SIGNAL. This provides a richer data set for generalizable theory development that accounts for potential impact of different environments. One specific opportunity for future work that arises from this demographic data is the ability to contribute to the question of the impact of expertise on behavior. Beyond being a critical question to address if we want to collect data from the general population, it has also arisen in a wide range of fields (Silvia et al., 2015; Wimshurst, Sowden, & Wright, 2016).

5.4 Lessons Learned

Our data collection events reinforced our assumptions about the importance of minimizing the cognitive load on participants. First, increased game complexity potentially increases the time necessary for instruction and comprehension of game rules. This increases the risk of player behavior being driven by difficulties in understanding how the game works rather than strategic considerations in achieving the goals. Game complexity also increases the risk of players disengaging early (causing data loss), either by no longer pursuing the win conditions or exiting the game. High complexity can also discourage players from participating in a game in the first place due to concern about the complexity of the game or because of limited time, increasing the difficulty in recruiting players. A SIGNAL-Board game, from instruction through playthrough, took four hours to complete, while the SIGNAL-Online took one hour.

In our early tests of SIGNAL-Board, we also learned the importance of providing an opportunity for the participants to explore the simulation environment prior to data collection. We solved this issue via a combination of tutorials and an exploratory round that allowed players to take actions with no consequences against temporary opponents; we then reset to the initial conditions and real seating assignments after the sandbox round ended. This both provided a space for players to explore salient strategies prior to data collection as well as supporting alternate learning styles. Through our beta testing, we generally found there was value in increasing this exploration time even at the cost of less time spent collecting data, as it led to fewer players making poor choices in the first round of play that then impacted the remainder of the game. Of course, this was not sufficient to eliminate all problematic behavior. We observed players who chose to ignore the stated goals to pursue goals of their own, which could range from attempting to coordinate a peaceful three-way tie (while mathematically impossible), to trolling other players via irrational behavior intended to be provocative, to random employment of nuclear weapons.

We chose to have a fixed number of rounds rather than having an unknown end round (via a probabilistic ending function) due to the need to the fit the SIGNAL-Board game into fixed-time events. A consequence of this is that a known final round will likely influence player behavior. Although previous experiments have found that players don't tend to be strictly rational under backwards induction (McKelvey & Palfrey, 1992), they do demonstrate that play becomes more rational as the end of the game approaches.

This was not the only instance in which time limitations impacted the design of SIGNAL. The SIGNAL-Online game was most affected by this as retention in online platforms like Amazon Mechanical Turk, is notoriously low. To address this, we strove to keep the SIGNAL-Online duration under an hour by instituting hard limits on phase length. As we had more latitude on the length of SIGNAL-Board, it did not include these time limits and let the players decide when they were ready to move on the to the next phase. In our development process, we identified a potential issue with fixed time limits in SIGNAL-Board: the ability to abuse the time near the threshold to stage an overwhelming suprise attack at the very end of the Signaling Phase without giving the other players time to react before the Phase ended. This has a more limited impact in SIGNAL-Online as it is harder to "pre-stage" actions (as each action requires several precise clicks) as opposed to the physical act of placing multiple cards and tokens. Thus, a player has suffient time for at least a partial reaction in the SIGNAL-Online game in the event of large-scale surprise events in the Signaling Phase.

The difference in game mechanics between the platforms could have a significant impact on behavior. This creates potential future opportunities to explore how behavior could be impacted by the different mechanics in SIGNAL-Board vs. SIGNAL-Online.

5.5 Conclusion

The study of conflict and strategy in modern and future contexts is an important area of focus for a wide range of researchers. Wargames have long provided a tool for exploring scenarios and developing insights across disciplines, ranks, and commands. Experimental wargaming provides another tool in the wargaming and conflict analysis toolbox, with some unique advantages.

To provide the analytic rigor needed to develop insights into complex conflict scenarios, we argue that EWGs have desirable attributes as demonstrated in this paper, including randomization, replicability, controllability, instrumentation, and neutrality. In designing SIGNAL, we have worked to ensure that it upholds these attributes and allows for rigorous analysis of the data collected. This pursuit drove a myriad of design choices and concessions in both the board and online versions of the game and throughout the implementation process. This paper documents the logic behind some of the design challenges and solutions developed by the team, and lessons learned to inform future efforts. Most importantly, we call on the wargaming community to continue a dialogue and debate on EWGs, to more fully develop a working set of definitions and requirements, and to implement future designs as data-generating processes for scientific analysis—openly sharing best practices and lessons learned. With thoughtful and dedicated effort by the community of practice, experimental wargaming can become a tool for research and inquiry that grows sharper the more we use it.

Acknowledgments

We would like to acknowledge the valuable contributions we recieved from Wes Spain and Jarret Lafleur. Additionally, we wish to thank all of the various groups and individuals that supported and participated in our data collection events.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Disclaimer

This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

Bibliography

Armenta, M., Epifanovskaya, L., Letchford, J., Lakkaraju, K.,Whetzel, J., Goldblum, B., and Tibbetts, J. 2020. SIGNAL Game Manual Version 1.0, Sandia National Laboratories Report No. SAND2020-7100.

Asal, V. and Beardsley, K. 2007. Proliferation and international crisis behavior. Journal of Peace Research, Vol 44, No 2, 139-155.

Barbieri, K. 1996. Economic interdependence: A path to peace or a source of interstate conflict?. Journal of Peace Research, Vol 33, No 1, 29-49.

Bartels, E.M. 2019. The Science of Wargames: A discussion of philosophies of science for research games. Presented at War Gaming and Implications for International Relations Research MIT CIS and US. Naval War College Workshop, Endicott House, July 2019. http://www.elliebartels.com/uploads/1/1/0/6/110629149/bartels-the_science_of_wargames_nwc_mit.pdf , retrieved on August 30th, 2020.

Bennett, A. and Elman, C. Case study methods in the international relations subfield. Comparative Political Studies, Vol 40, No 2, 170-195.

Brecher, M., Wilkenfeld, J., Beardsley, K., James, P., and Quinn, D. 2020. International Crisis Behavior Data Codebook, Version 13. http://sites.duke.edu/icbdata/data-collections/

Brehmer, B. and Dörner, D. 1993. Experiments with computer-simulated microworlds: Escaping both the narrow straits of the laboratory and the deep blue sea of the field study. Computers in Human Behavior, Vol 9, No 2-3, 171-184.

Caffrey Jr, M.B., 2019. On Wargaming.

Elbridge, C. 2018. If You Want Peace, Prepare for Nuclear War. Foreign Affairs. Vol 97, No 25.

Coyle, P. and McKeon, J. 2017. The huge risk of small nukes. Politico.

Dodge, M. and Nikolova, D. 2018. 5 Myths About the Nuclear Posture Review. Washington, D.C.: The Heritage Foundation. https://www.heritage.org/missile-defense/commentary/5-myths-about-the-nuclear-posture-review.

Doyle, J. 2017 Mini-Nukes: Still a bad choice for the United States. Bulletin of the Atomics Scientists.

Fischerkeller, M. 2017. Incorporating Offensive Cyber Operations into Conventional Deterrence Strategies. Survival, Vol 59, No 1, 103-134.

Fitzpatrick, M. 2019. Artificial Intelligence and Nuclear Command and Control. Survival,Vol 61, No 3, 81-92.

Gartzke, E. and Lindsay, J.R. 2019. Cross-Domain Deterrence: Strategy in an Era of Complexity. Oxford University Press.

Halperin, M.H. 1961. Nuclear weapons and limited war. Journal of Conflict Resolution, Vol 5, No 2, 146-166.

Harvey, J.R. 2017. Low-Yield Nuclear Weapons Are Worth a New Look. War on the Rocks. November 10. https://warontherocks.com/2017/11/low-yield-nuclear-weapons-worth-new-look/.

Heath, B., Hill R., and Ciarallo, F. 2009. A survey of agent-based modeling practices (January 1998 to July 2008). Journal of Artificial Societies and Social Simulation, October.

Hernandez, A., McDonald, M., and Ouellet, J. 2015. Post wargame experimentation and analysis: Re-examining executed computer-assisted wargames for new insights. Military Operations Research, Vol 20, No4, 19–37.

Huth, P.K. 1999. Deterrence and international conflict: Empirical findings and theoretical debates. Annual Review of Political Science, Vol *2*, No 1, 25-48.

Hyde, S.D. 2015. Experiments in International Relations: Lab, Survey, and Field. Annual Review of Political Science, Vol 18, No 1, 403–424.

Jensen, B. and Valeriano, B. 2019. Cyber Escalation Dynamics: Results from War Game Experiments International Studies Association, Annual Meeting Panel: War Gaming and Simulations in International Conflict March 27, 2019.

Kreps, S. and Schneider, J. 2019. Escalation firebreaks in the cyber, conventional, and nuclear domains: Moving beyond effects-based logics. Journal of Cybersecurity, Vol 5, No 1.

Kroenig, M. 2018. *The logic of American nuclear strategy: why strategic superiority matters*. Oxford University Press.

Lanoszka, A. 2016. Russian hybrid warfare and extended deterrence in eastern Europe. International affairs, Vol 92, No 1, 175-195.

Larsen, J.A. and Kartchner, K.M. 2014. *On limited nuclear war in the 21st century*. Stanford University Press.

Lastowka, G. 1999. University of V. S. of L. Wargames.

Lebow, R.N. and Stein, J.G. 1989. Rational deterrence theory: I think, therefore I deter. World politics, Vol 41, No 2, 208-224.

Libicki, M.C. 2018. Expectations of cyber deterrence. Strategic Studies Quarterly, Vol 12, No 4, 44-57.

Lieber, K.A. and Press, D.G. 2009. The Nukes We Need-Preserving the American Deterrent. Foreign Affairs, Vol 88, No 6, 39-51.

Lieber, K.A. and Press, D.G. 2016. The New Era of Nuclear Weapons, Deterrence, and Conflict. Strategic Studies Quarterly, Vol 10, No 5, 31-42.

Lin-Greenberg, E. 2018. Game of Drones: The Effect of Remote Warfighting Technology on Conflict Escalation (Evidence from Wargames).

Lovelace, D. Jr. 2016. *Hybrid warfare and the gray zone threat*. In (eds.) Lovelace, D. Jr. Terrorism. Vol. 141. Oxford University Press.

Macal, C.M. 2016. Everything you need to know about agent-based modelling and simulation. Journal of Simulation, Vol 10, No 2, 144–156.

Maoz, Z., 2002. Case study methodology in international studies: From storytelling to hypothesis testing. *Evaluating methodology in international studies*, p.163.

Maoz, Z., Johnson, P.L., Kaplan, J., Ogunkoya, F. and Shreve, A.P., 2019. The dyadic militarized interstate disputes (MIDs) dataset version 3.0: Logic, characteristics, and comparisons to alternative datasets. *Journal of Conflict Resolution*, *63*(3), pp.811-835.

Maxwell, S. E., Delaney, H. D., and Kelley, K. Designing experiments and analyzing data: A model comparison perspective. Routledge, 2017.

Mazarr, Michael J. *Mastering the gray zone: understanding a changing era of conflict*. US Army War College Carlisle, 2015.

McKelvey, R.D. and Palfrey, T.R., 1992. An experimental study of the centipede game. Econometrica: Journal of the Econometric Society*,* Vol 60, No 4, 803-836.

Mearsheimer, J.J. 1985. *Conventional deterrence*. Cornell University Press.

Mecklin, J. 2017. Mini-nukes: The attempted resurrection of a terrible idea. Bulletin of the Atomic Scientists. September 14. https://thebulletin.org/2017/09/mini-nukes-the-attempted-resurrection-of-a-terrible-idea/#sf_form_salesforce_w2l_lead_1.

Ministry of Defence. 2017. Defence Wargaming Handbook. Ministry of Defence, U.K.

Mueller, K.P. "Conventional deterrence redux: avoiding great power conflict in the 21st century." Strategic Studies Quarterly, Vol 12, No 4, 76-93.

Mullinix, K.J., Leeper, T.J., Druckman, J.N. and Freese, J. 2015. The generalizability of survey experiments. Journal of Experimental Political Science, Vol 2, No 2, 109-138.

Narang, V. 2018. The Discrimination Problem: Why Putting Low-Yield Nuclear Weapons on Submarines is so Dangerous. War on the Rocks. Febuary 8. https://warontherocks.com/2018/02/discrimination-problem-putting-low-yield-nuclear-weapons-submarines-dangerous/.

Oberholtzer, J., Doll, A., Frelinger, D., Mueller, K., and Pettyjohn, S. 2019. Applying wargames to real-world policies. Science, Vol 363, No 6434, 1406–1407.

O'Hanlon, M.E. 2018. Trump's nuclear plan mostly makes sense. Washington, D.C.: Brookings Institute. https://www.brookings.edu/blog/order-from-chaos/2018/02/06/trumps-nuclear-plan-mostly-makes-sense/.

Palmer, G., D'Orazio, V., Kenwick, M. R., and McManus, R. W. 2020. Updating the Militarized Interstate Dispute Data: A Response to Gibler, Miller, and Little. International Studies Quarterly, Vol 64, No 2, 469–475.

Pauly, R.B. 2018. Would US Leaders Push the Button? Wargames and the Sources of Nuclear Restraint. International Security, Vol 43, No 2, 151-192.

Perla, P.P. 2011. *Peter Perla's the Art of Wargaming: A Guide for Professionals and Hobbyist*. The History of Wargaming Project.

Perla, P.P. and McGrady, E.D. 2011. Why wargaming works. Naval War College Review, Vol 64, No 3, 111-130.

Perla, P.P., Markowitz, M. and Weuve, C. 2005. Game-Based Experimentation for Research in Command and Control and Shared Situational Awareness. CNA.

Press, D.G., Sagan, S.D. and Valentino, B.A. 2013. Atomic aversion: Experimental evidence on taboos, traditions, and the non-use of nuclear weapons. American Political Science Review, Vol 107, No 1, 188-206.

Reddie, A. W., Goldblum, B. L., Lakkaraju, K., Reinhardt, J., Nacht, M., & Epifanovskaya, L. 2018. Next-generation wargames. Science, Vol 362, No 6421, 1362–1364.

Rubel, R.C. 2006. The epistemology of war gaming. Naval War College Review, Vol 59, No 2, 108-128.

Sagan, S. D., and Valentino, B. A. 2017. Revisiting Hiroshima in Iran: What Americans Really Think about Using Nuclear Weapons and Killing Noncombatants. International Security, Vol 42, No 1, 41–79.

Schneider, J. 2017. Cyber and crisis escalation: insights from wargaming. In *USASOC Futures Forum*.

Schneider, J. 2019a. Cyber and Cross Domain Deterrence: Deterring Within and From Cyberspace.

Schneider, J. G. 2019b. War Gaming for Nuclear Research. Albuquerque.

Schmitter, P.C. 2016. Comparative politics: its past, present and future. Chinese political science review, Vol 1, No 3, 397-411.

Snidal, D., 2004. Formal Models of International Politics. In (eds.) Sprinz, D.F. and Wolinsky, Y. *Models, numbers, and cases: methods for studying international relations*, University of Michigan.

Silva, A., Emmanuel, G., McClain, J.T., Matzen, L. and Forsythe, C. 2015, Measuring expert and novice performance within computer security incident response teams. In *International Conference on Augmented Cognition,* 144-152. Springer, Cham.

Trenin, Dmitri. 2019. Russian views of US nuclear modernization. Bulletin of the Atomic Scientists, Vol 75, No 1, 14-18.

Uribe, E., Bonin, B., Minner,M.F., Reinhardt, J.C., Hammer,A., Teclemariam,N.P., Miller,T.H., Forrest, R., Apolis,J.J., and Yang,L.I. 2020. *Why does cyber deterrence fail and when might it succeed? A framework for cyber scenario analysis*. No. SAND2020-5016. Sandia National Lab.(SNL-CA), Livermore, CA (United States); Sandia National Laboratories, Minneapolis, MN.

US Department of Defense. 2018. Nuclear Posture Review. February. https://media.defense.gov/2018/feb/02/2001872886/-1/-1/1/2018-nuclear-posture-review-finalreport.pdf

Warden, J.K. 2018. Limited Nuclear War: The 21st Century Challenge for the United States. Livermore Papers on Global Security, Vol 4.

Washburn, A., and Kress, M. 2009. Combat Modeling. In F. S. Hiller (Ed.), *International Series in Operations Research and Management Science*. NY: Springer.

Weber, A. and Parthemore, C. 2019. Smarter US modernization, without new nuclear weapons. Bulletin of the Atomic Scientists, Vol 75, No 1, 25-29.

Weiss, J.C. and Dafoe, A. 2019. Authoritarian Audiences, Rhetoric, and Propaganda in International Crises: Evidence from China. International Studies Quarterly, Vol 63, No 4, 963-973.

Williams, D and Lowther, A.B. 2017. Lower-Yield Weapons Will Raise, Not Lower, The Threshold for Nuclear Use. *Defense One.*

Wimshurst, Z. L., Sowden, P. T., and Wright, M. 2016. Expert-novice differences in brain function of field hockey players. Neuroscience, Vol 13, 31–44.

Wirtz, James J. 2018. How Does Nuclear Deterrence Differ from Conventional Deterrence?. Strategic Studies Quarterly, Vol 12, No 4, 58-75.

Z. Maoz, P. L. Johnson, J. Kaplan, F. Ogunkoya, and A. P. Shreve. 2019. The Dyadic Militarized Interstate Disputes (MIDs) Dataset Version 3.0: Logic, Characteristics, and Comparisons to Alternative Datasets. Journal of Conflict Resolution, Vol 63, No 3,811–835.

Zhao, T. 2019. What the United States can do to stabilize its nuclear relationship with China. Bulletin of the Atomic Scientists, Vol 75, No 1, 19-24.

Appendix A: Data Collection Sheets

Game ___

- [ ] Purple
- [ ] Green
- [ ] Orange

Signaling Token Locations:

_____ _____ _____ _____ _____ _____

_____ _____ _____ _____ _____ _____

Action Cards staged:

Quantity		Quantity		Quantity	
___	Build Military Base	___	Electro-Magnetic Pulse Nuclear Weapon	___	Conventional Defense
___	Build Towns & Cities	___	High-Precision Low-Yield Nuclear Weapon	___	Conventional Infantry Assault
		___	Nuclear Weapon	___	Conventional Missile Strike
				___	Conventional Naval Assault
				___	Cyber Strike

Trades

Other Player	Trade Terms	Notes

1 Purple ☐ Green ☐ Orange ☐ Location: ______ Cost: ______

- ☐ Build Towns and Cities
- ☐ Build Military Base
- ☐ **PASS**

Success?
☐ Yes ☐ No

Notes:

- ☐ Electro-Magnetic Pulse Nuclear Weapon
- ☐ High-Precision Low-Yield Nuclear Weapon
- ☐ Nuclear Weapon

- ☐ Conventional Defense
- ☐ Conventional Infantry Assault
- ☐ Conventional Missile Strike
- ☐ Conventional Naval Assault
- ☐ Cyber Strike

2 Purple ☐ Green ☐ Orange ☐ Location: ______ Cost: ______

- ☐ Build Towns and Cities
- ☐ Build Military Base
- ☐ **PASS**

Success?
☐ Yes ☐ No

Notes:

- ☐ Electro-Magnetic Pulse Nuclear Weapon
- ☐ High-Precision Low-Yield Nuclear Weapon
- ☐ Nuclear Weapon

- ☐ Conventional Defense
- ☐ Conventional Infantry Assault
- ☐ Conventional Missile Strike
- ☐ Conventional Naval Assault
- ☐ Cyber Strike

3 Purple ☐ Green ☐ Orange ☐ Location: ______ Cost: ______

- ☐ Build Towns and Cities
- ☐ Build Military Base
- ☐ **PASS**

Success?
☐ Yes ☐ No

Notes:

- ☐ Electro-Magnetic Pulse Nuclear Weapon
- ☐ High-Precision Low-Yield Nuclear Weapon
- ☐ Nuclear Weapon

- ☐ Conventional Defense
- ☐ Conventional Infantry Assault
- ☐ Conventional Missile Strike
- ☐ Conventional Naval Assault
- ☐ Cyber Strike

Trades/Notes:

Upkeep Sheet

Game ___

- [ ] Purple
- [ ] Green
- [ ] Orange

Complete?

☐ Step 1: Return all cards and signaling tokens to your hand.

☐ Step 2: Count your food: ☐

☐ Step 3: Count your population and support:

________ = 3 + (3 x ☐)

Support

________ = ☐ + ☐ + ☐ + 2 x ☐

Population

☐ Step 4: Adjust your population until there is enough food to support them:

IF: ________ > ________

Population Support

THEN:

Remove: **OR** Remove: **OR** Remove: **OR** Convert: TO

☐ Step 5: Count your resources. Remember to count resources in allied states and resources from trade deals..

☐ ☐ ☐ ☐

☐ Step 6: Count your non-damaged infrastructure.

☐ ☐

☐ Step 7: Collect income - use infrastructure counts from Step 6.

________ = 2 x + 4 x + 3 x

Income

☐ Step 8: Repair damaged infrastructure and remove defense tokens.

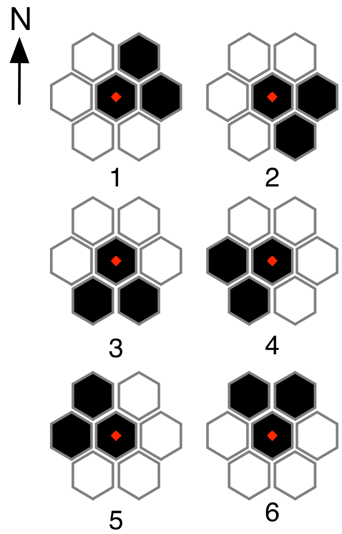
Scoring

	Purple			Green			Orange		
			Total Points			Total Points			Total Points
Survival: Maintain Territory *(# of owned hexes that have been destroyed or occupied. Winner is the one with the lowest score.)*	Score	Rank	Points	Score	Rank	Points	Score	Rank	Points
			+			+			+
Gaining resources *(# of food, oil, iron, precious metals and credits)* Score =  + $\frac{\text{Credits}}{5}$	Score	Rank	Points	Score	Rank	Points	Score	Rank	Points
			+			+			+
Gaining infrastructure *(# of towns, cities and military bases)* Score = 1X  + 1 X  + 3 X 	Score	Rank	Points	Score	Rank	Points	Score	Rank	Points
									+1

Which player is leading?

Purple	Green	Orange

Appendix B: Traditional Nuclear Weapon Secondary Effect Patterns



NW Blast Patterns.
Roll die to choose one.