

# UC Berkeley

## UC Berkeley Previously Published Works

### Title

Evidence of the unthinkable: Experimental wargaming at the nuclear threshold

### Permalink

<https://escholarship.org/uc/item/2bf9x01p>

### Journal

Journal of Peace Research, 60(5)

### ISSN

0022-3433

### Authors

Reddie, Andrew W

Goldblum, Bethany L

### Publication Date

2023-09-01

### DOI

10.1177/00223433221094734

Peer reviewed

# Evidence of the unthinkable: Experimental wargaming at the nuclear threshold

Andrew W Reddie 

*School of Information, University of California, Berkeley*

Bethany L Goldblum 

*Department of Nuclear Engineering, University of California, Berkeley*

Journal of Peace Research  
2023, Vol. 60(5) 760–776  
© The Author(s) 2022



Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/00223433221094734  
journals.sagepub.com/home/jpr



## Abstract

Ongoing nuclear modernization programs in Russia, China, and the USA have reopened longstanding debates among scholars concerning whether tailored nuclear weapons are likely to have destabilizing consequences for international security. Without data to adjudicate this debate, however, these discussions have remained entirely theoretical. In this article, we introduce an experimental wargaming platform, SIGNAL, to quantify the effect of tailored nuclear capabilities on the nuclear threshold in a simulated environment. We then compare these results with a survey experiment using scenarios related to military basing, cyber operations, and nuclear threats from the wargame environment. While the survey experiments suggest that the presence of tailored nuclear capabilities increases the likelihood of conflict escalation, this trend diminishes in the wargaming context. Across both data-generating processes, we find support for the proposition that lower-yield nuclear weapons are used as a substitute for their higher-yield counterparts. These results have consequences for recent and ongoing policy debates concerning strategic posture and the future of arms control. This work also makes methodological contributions to the design and application of experimental wargaming for social science research, particularly for scenarios where data are limited or non-existent.

## Keywords

experiments, nuclear weapons, wargaming

*'We have no empirical data beyond 1945 about how events may run if nuclear weapons are used.'*—Sir Michael Edward Quinlan

## Introduction

Debates concerning the strategic impact of tailored nuclear weapons – nuclear weapons designed to produce custom effects such as a low explosive yield or electromagnetic pulse (EMP) effects – have existed throughout the nuclear age, with some suggesting that they contribute to stability and others to instability. Nitze, writing during a period in which a doctrine of massive retaliation was ascendant, suggested as early as 1955 that adding tailored nuclear capabilities might reduce the vulnerability of the USA to nuclear blackmail by the Soviet

Union (Buzzard, 1956; Nitze, 1956). Ten years later, McNamara suggested that NATO adopt the doctrine of flexible response with an emphasis on the role of theater nuclear weapons to ensure that the USA had the capability to respond to escalation (Powell, 1988), and in 1974, the Schlesinger doctrine outlined the uses of limited nuclear options as counterforce weapons (Schlesinger, 1975; Burr, 2005). More recently, nuclear modernization in Russia, China and the USA has rekindled academic arguments regarding the opportunities and pitfalls associated with tailored nuclear capabilities (Heginbotham et al., 2017; Podvig, 2018;

---

**Corresponding author:**  
areddie@berkeley.edu

Talmadge, 2019). The release of the 2018 Nuclear Posture Review announcing plans for a new low-yield nuclear warhead (the W76-2 variant), in particular, renewed this theoretical debate with some suggesting that this development would be destabilizing while others argued that the capabilities are necessary for stability (Broad & Sanger, 2016; Narang, 2018; Long, 2018; Roblin, 2019; Facini, 2020).

However, in the absence of empirical data, how do we adjudicate these claims? How, if at all, might tailored nuclear capabilities impact the threshold for nuclear use? To address these questions, we introduce a first application of large- $N$  experimental wargaming as a method of social science inquiry.

Below, we examine the impact of high-precision low-yield and enhanced-EMP nuclear weapons on the nuclear threshold using data from the Strategic Interaction Game between Nuclear Armed Lands (SIGNAL) experimental wargaming platform. Specifically, we compare player behavior and game outcomes with and without these weapons in the arsenal and examine the likelihood of nuclear use. To benchmark the study, we compare these results with a more traditional three-segment survey experiment that uses the same treatment in scenarios designed to approximate those from the wargame setting. Our analysis suggests that the inclusion of tailored nuclear capabilities in an arsenal may increase the likelihood of nuclear use and substitute for high-yield nuclear use. This effect was observed with statistical significance in the survey setting – an important finding given the widespread use of survey methods in the field. Finally, we reflect on the methodological contribution of the article and the potential applications of experimental wargaming to behavioral social science and international relations research.

### *Tailored nuclear options in theory*

A lack of observational data poses a significant challenge to the empirical examination of nuclear issues (Colby & Gerson, 2013; Lieber & Press, 2017). As Gartzke, Kaplow & Mehta (2015) note, the literature often fails to account for the ‘diverse portfolios of [nuclear] weapons with varying range, destructive power, and other characteristics’. Scholars have come to rely on theory and extrapolation from a limited number of cases to examine the potential effects of adding new capabilities to the nuclear arsenal (Brodie et al., 1946; Schelling, 1966; Zagare, 1985; Brewer & Blair, 1979; Powell, 1990; Larsen & Kartchner, 2014; Acton, 2015; Heimer, 2018). This scholarship has contributed a number of assertions

related to the nuclear threshold. While some suggest that the ‘nuclear-ness’ of weapons explains patterns of non-use (Tannenwald, 1999, 2005), others posit that there remain conditions under which states may still engage in limited nuclear war (Larsen & Kartchner, 2014; Freedman & Michaels, 2019). This leaves us with a central question, how do nuclear capabilities with tailored effects shape the likelihood of nuclear use?

In the sections below, we outline two schools of thought pertaining to the impact of tailored nuclear weapons on escalation.

### *Tailored nuclear weapons and stability*

Throughout the 1950s and 1960s, proponents of tailored nuclear capabilities outlined the benefits of using tactical nuclear weapons in a graduated deterrence architecture rather than in the ‘massive retaliation’ strategy that represented the orthodoxy of the period (Blackett, 1958; Kissinger, 1960; Osgood, 1979). Utilitarian arguments for the development of tailored nuclear capabilities go beyond deterrence to consider nuclear capabilities as warfighting tools to address discrete military challenges for which more traditional high-yield nuclear capabilities are ill equipped. Both academics and policymakers have suggested that nuclear weapons are needed that reliably produce ‘special effects’ with much lower collateral damage to destroy or otherwise neutralize targets (Dowler, Howard & Joseph, 1991; Potter et al., 2000; Blair, Carns & Vitto, 2004; Levi, 2004; Tertrais, 2011; Lieber & Press, 2013; Davis et al., 2019). The 2002 US Nuclear Posture Review notes three types of targets for tailored nuclear weapons: ‘hardened or deeply buried facilities; chemical and biological agents; and mobile and relocatable targets’.<sup>1</sup> Others point to the substantially lower levels of collateral damage associated with the use of tailored nuclear weapons (Younger, 2000).

Scholars have also recently argued that tailored nuclear weapons offer a useful tool for improving crisis stability by controlling escalation (Colby, 2014; Kroenig, 2015, 2016, 2018). In work re-examining the withdrawal of nuclear forces in Europe, for example, Kroenig notes that the decision to ‘eliminate tactical nuclear weapons from Europe has left Russia with a wide range of options on the nuclear escalation ladder’ – suggesting that the deployment of a symmetrical nuclear capability might limit these options (Kroenig, 2015, 2016). This

<sup>1</sup> Nuclear Posture Review Report, 8 January 2002. This report and the relevant text is also discussed in Glaser & Fetter (2005).

has led some to argue that low-yield nuclear weapons enable a more credible nuclear deterrent by providing a reasonable response option in certain regional scenarios (Lieber & Press, 2009).

### *Tailored nuclear weapons and instability*

Alternatively, there are two logics that underpin the theory that limited nuclear capabilities are likely to increase the likelihood of nuclear use. First, nuclear weapons tailored to reduce civilian casualties may weaken moral norms associated with their use. Second, tailored nuclear weapons suffer from a discrimination problem – whereby an adversary cannot distinguish between high-yield and low-yield capabilities – contributing to inadvertent escalation and conflict spirals.

Scholars note the potential for tailored nuclear weapons to reduce the nuclear threshold as they provide an attractive means to accomplish military objectives while limiting collateral damage (Halperin, 1961; Von Hippel et al., 1988; Rovere & Robertson, 2013; Doyle, 2016, 2017). The reduced incidental injury to civilians afforded by low-yield or enhanced-EMP nuclear weapons, for example, poses concerns that their deployment weakens the nuclear taboo by being perceived as a less dangerous nuclear option, eroding moral and ethical norms surrounding nuclear non-use (Tannenwald, 1999). Without the indiscriminate blast associated with traditional nuclear weapons, will the use of nuclear weapons be viewed as tolerable, desirable even, leading to a ripple effect that legitimizes nuclear use? Incidentally, some also argue that there is no guarantee that the amount of collateral damage will be substantially altered by these capabilities (Toon et al., 2019).

The employment of tailored nuclear capabilities may also have escalatory rather than dampening effects on conflict escalation. There is no assurance that a nuclear confrontation that begins with the use of tailored nuclear weapons will remain limited – with the potential for escalation to a strategic nuclear war with existential consequences (Daugherty, Levi & Von Hippel, 1986). Brodie, for example, suggests that when conflict models take into account the reciprocal use of low-yield nuclear weapons, the result is a conflict spiral: ‘we tend in the end to get the same kind of utterly nihilistic result in considering unrestricted tactical war in the future that we get in unrestricted strategic war’ (Brodie, 1955). This theoretical claim was later showcased in subsequent wargames in which ‘practical exercises with simulated tactical nuclear

weapons undermined any claims that such warfare could be kept limited’ (Freedman & Michaels, 2019). More recently, scholars have argued that the deployment of sea-launched, low-yield nuclear weapons reduces the separation between conventional and nuclear escalation as adversaries do not know a priori whether or not an incoming missile is armed with a nuclear payload (Narang, 2018; Weber & Parthemore, 2019). Faced with the uncertainty of what type of nuclear capability an adversary is deploying or launching, there are fears that state leaders will prematurely (or pre-emptively) embark upon an escalatory response.

This theoretical literature yields the following question: if tailored nuclear capabilities are present, is nuclear use more or less likely? While some argue that tailored nuclear capabilities are destabilizing and others argue they are stabilizing, it is also possible that they have no impact. Without data, we cannot adjudicate these theoretical claims. Below, we present an experimental wargaming approach that attempts to provide that data.

### **Experimental wargaming**

Much of our contemporary understanding of conflict escalation dynamics involving nuclear weapons relies on theory rather than empirics. In response, scholars have turned to alternative data-generating processes to examine nuclear issues. Traditional seminar-based wargames, for example, offer a mechanism for senior policymakers to engage with vexing geographical and geopolitical challenges (Pauly, 2018). Formal and computer-based models also serve as longstanding examples of this work (Powell, 1988, 1990). More recently, scholars have used survey and laboratory experiments to investigate nuclear matters – including the conditions under which subjects (often members of the public or undergraduate subjects) would resort to nuclear use (Press, Sagan & Valentino, 2013; Quek, 2016; Sagan & Valentino, 2017). These approaches, like all synthetic data-generating processes, have strengths and weaknesses – from the assumptions that simplify and underpin formal models to the lack of consequences associated with survey responses.

Here, we propose an approach that combines experimental methods with wargaming techniques that have been developed over the past six decades (Perla, 1990; Asal, 2005; Perla & McGrady, 2011; Sabin, 2012; Schofield, 2013). In the process, we provide a new tool for social scientists to interrogate theories on phenomena for which there are limited or no empirical data.

### *Wargames as experiments*

Important characteristics of experimental design are often neglected in traditional wargaming – limiting their utility for quantitative analysis and causal inference. However, when executed using experimental design principles, we suggest that these challenges can be overcome to enable a new methodological tool in the social science toolkit (Reddie et al., 2018). As Pauly notes, one of the major assets of wargaming in comparison with survey experiments is the degree to which participants are ‘immersed’ in the strategic environment of the game (Pauly, 2018). While still an abstraction of reality, wargames provide a rich environment for insight into human decisionmaking, where a wide range of potential scenarios and conflict dynamics can be captured for analysis (Lin-Greenberg, Pauly & Schneider, 2022).

The following characteristics make experimental wargaming particularly well suited to social science inquiry (Tingley & Walter, 2011; Hyde, 2015; Rathbun, Kerzner & Paradis, 2017). First, experimental wargames are repeatable and allow for inference on the basis of player behavior. Second, experimental wargames can be conducted using a control–treatment design, where all conditions within the experiment other than the treatment variable and the characteristics associated with each player are held constant. Third, experimental wargaming provides researchers with control over the variables under examination – in this case, the military capabilities provided to each player. Fourth, the instrumentation of the game can be optimized for data collection – particularly in digital settings. This is important given the data loss in traditional wargaming frameworks that use self-reporting or rapporteurs to collect data on game-level outcomes. Finally, experimental wargames allow for increased fidelity and complexity associated with the scenario in comparison with formal models and survey experiments.

Indeed, the application of experimental design principles to wargaming has already been leveraged by scholars carrying out longitudinal analysis on archived games as well as those creating small-*N*, analog experimental games to address nuclear, cyber, and drone warfare scenarios (Schneider, 2017; Pauly, 2018; Lin-Greenberg, 2018; Jensen & Valeriano, 2019). Below, we provide a brief description of the SIGNAL game architecture and address the mechanisms through which it addresses the research question – do tailored nuclear weapons lead to an increased likelihood of nuclear use?

### *SIGNAL design*

SIGNAL is a three-player (1v1v1), turn-based experimental wargaming platform built upon a hexagonal-

based grid.<sup>2</sup> All players – from a convenience sample, UC Berkeley’s Experimental Social Science Laboratory, and Amazon Mechanical Turk – enter the game through a web browser, watch a short video, and complete a tutorial and demographic survey before competing in a virtual world to achieve the highest relative score across three win conditions over five rounds of play. Two of these win conditions are economic, focused on building infrastructure (i.e. maximizing the number of towns, cities, and/or military bases) and gaining resources (including food, oil, iron, and precious metals) and the other is security-related – centered on minimizing the loss of territory (as opposed to commandeering territory to the greatest degree possible). The zero-sum nature of this competition is specifically designed to provide a competitive environment, but not to force military conflict (Letchford et al., 2022).

As illustrated in Figure 1, SIGNAL uses abstract ‘countries’ (denoted by their color as Green, Purple, and Orange) to reduce the risk of players interpolating real-world cases into the experimental environment. The game world and competitive dynamics are explicitly designed to not map onto any real-world scenario in favor of illuminating how players respond to strategic questions rather than caricaturing a specific conflict or a country’s likely action(s). There are, of course, trade-offs between the internal and external validity of the study in making this choice.

As a between-subjects, control-treatment experiment, the nuclear capabilities provided to the Green and Purple players vary. For the purposes of this experiment, Green and Purple are given nuclear weapons along with a set of conventional capabilities. This dyad either has nuclear forces comprising only high-yield nuclear weapons (the control condition) or those comprising high-yield and tailored nuclear weapons, specifically high-precision low-yield (HPLY) and enhanced-EMP weapons (the treatment condition).<sup>3</sup> Conventional military capabilities (infantry, naval, missile, and defense capabilities) provide alternative means to hold resources in the game and degrade an adversary’s capabilities. The Orange player, while having the same conventional military capabilities as other players and a slightly increased access to resources, does not

<sup>2</sup> For a more complete discussion of the game design and mechanics, see the SIGNAL Game Manual (Armenta et al., 2020).

<sup>3</sup> The high-yield nuclear weapon card permanently destroys six hexes on the map while the HPLY nuclear weapon card permanently destroys one hex. The enhanced-EMP nuclear weapon card disables the infrastructure on a particular hex.



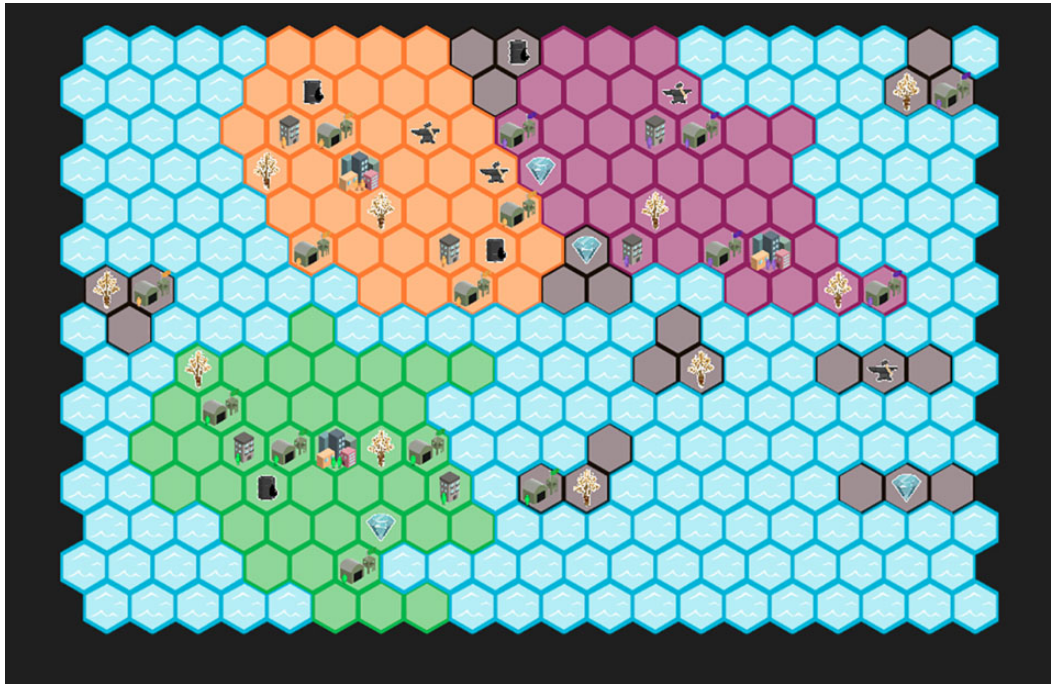


Figure 1. Map of the SIGNAL game environment.

Table I. Treatment and control conditions tested using the SIGNAL framework. Here, *HY* represents high-yield nuclear weapons, *T* represents those players provided HPLY nuclear weapons and electromagnetic pulse (EMP) nuclear weapons, and *CW* represents conventional (non-nuclear) weapons

Condition	Purple	Green	Orange
Control	<i>CW+HY</i>	<i>CW+HY</i>	<i>CW</i>
Treatment	<i>CW+HY+T</i>	<i>CW+HY+T</i>	<i>CW</i>

have nuclear capabilities. The experimental conditions are summarized in Table I.

Game play is governed by a set of rules that do not require external adjudication, with play taking place on a round-by-round basis. In brief, each round comprises three phases: signaling, action, and upkeep.

The signaling phase allows players to simultaneously place signaling tokens on hexes in the game environment and stage (face down) infrastructure or military capability cards to enable potential action in the subsequent phase. The staging of action cards has the dual significance of enabling future actions and ‘signaling’ what types of potential actions may be taken. Players may use signaling tokens to bluff, for example, placing them on hexes where they do not intend to take action. There is also a cost to staging capabilities that resembles ‘costly

signaling’ or a ‘credible commitment’ to act (Powell, 1990; Sagan & Suri, 2003; Colby et al., 2013; Yarhi-Milo, Kertzer & Renshon, 2018).<sup>4</sup>

During the action phase, players make decisions regarding which of the staged action cards they will execute.<sup>5</sup> The turn order is randomized to ensure that no player has a consistent first-mover advantage that may influence their decisionmaking and the game dynamics. During the upkeep phase, players keep score, collect income, and verify that they have sufficient resources to support their population and infrastructure. This gameplay provides a rich and immersive environment for players to grapple with strategies surrounding nuclear weapons. For example, we observed players solicit no first use agreements and nuclear umbrellas, as well as

<sup>4</sup> It costs a player one token to stage an action card during the signaling phase. An interesting extension of the study might vary this cost across games.

<sup>5</sup> In SIGNAL, there are three categories of action cards – infrastructure, conventional military operations, and nuclear weapons – each with a different color backing and with different costs associated with each action based on available resources, distance from a military base, and type of capability. Because cards are played face down in the signaling phase, a player’s opponents know broadly which type of card is being staged but not the specific action on the card or where and whether it will be used. This mechanism is designed to approximate signaling among adversaries.

Table II. Number of games in each experimental condition. In addition to high-yield (*HY*) and tailored (*T*) nuclear weapons, Green and Purple players have access to conventional military capabilities

		<i>Purple capabilities</i>	
		<i>HY</i>	<i>HY+T</i>
<i>Green capabilities</i>	<i>HY</i>	209	N/A
	<i>HY+T</i>	N/A	216

making (and breaking) alliances, while placing an average of 22 signaling tokens on the map per round.

We test two conditions using the SIGNAL experimental wargame. Of the 425 games included in the dataset, there are 209 games in the *control* condition in which the Green and Purple dyad have only high-yield nuclear weapons. There are 216 games in the *treatment* condition in which the Green and Purple dyad have additional tailored nuclear weapons. Table II summarizes the class prevalence of the dataset that, to the best of our knowledge, represents the largest wargaming dataset collected to date.

#### *Levels of analysis and measures of nuclear use*

SIGNAL collects game-based data ( $N = 425$ ) comprising all of the signaling and action moves undertaken by players within the game. Second, SIGNAL collects player data ( $N = 1275$  of which  $N = 850$  have nuclear capabilities)<sup>6</sup> from the game as well as demographic characteristics theorized to influence behavior, including age, political affiliation, occupation, and experience.<sup>7</sup>

At the game level, we extract from the data the incidence and type of nuclear use in each game.<sup>8</sup> When considering player-level data, we are also interested in the characteristics of players that use nuclear weapons.

<sup>6</sup> Seven players chose not to provide their age in the pre-survey. As a result, they have been dropped from the analysis that includes demographic characteristics.

<sup>7</sup> For example, Sagan and Valentino's recent study suggests that liberal, young, college-educated, and female players might be less 'hawkish' (Sagan & Valentino, 2017). We also posit that those with intimate knowledge of nuclear weapons and national security issues may be more reticent to use nuclear weapons. McIntyre et al. have also previously engaged with questions of how demographic characteristics might influence behavior in a wargame setting (McIntyre et al., 2006).

<sup>8</sup> Staging a nuclear capability in the signaling phase is necessary but not sufficient to constitute nuclear use. All nuclear cards, whether 'High-yield', 'HPLY', or 'EMP' are coded as nuclear use – and each card used to field the capability includes 'nuclear' in its title.

To scope the dependent variable, we use two different measures of nuclear use: nuclear first use as well as whether or not a player used nuclear weapons at any point during the game. Indeed, there is good reason to believe that the drivers of nuclear first use might be distinct from nuclear use and we endeavor to analyze both.<sup>9</sup>

As SIGNAL is a fixed round game, there is the potential for players to employ limited backward induction, modifying their strategy in the last round toward the optimal actions required to achieve the win conditions – and knowing that there is no opportunity for retribution. This introduces a potential systematic bias in the analysis whereby player actions are governed by game mechanics as opposed to their own strategic decision-making. To quantify the impact of this, the data are analyzed with and without actions from the last round included in the dataset.

#### *Data analysis*

We use regression-based methods to interrogate the effect of the experimental treatment (nuclear capabilities) and demographic variables on the dependent variable of interest (nuclear use) (Draper & Smith, 1998). As the dependent variable is treated as dichotomous (i.e. nuclear weapons are either used or they are not), we use logistic regression.<sup>10</sup> Specifically, we apply a series of logistic regression models to test the effects of the treatment on the binary wargame outcome of interest,  $Y$ , where 0 indicates no nuclear use and 1 corresponds to nuclear use of any kind. Here,  $x_1, \dots, x_k$  represents a set of predictor variables that might influence nuclear weapon use (e.g. the presence/absence of tailored nuclear capabilities, demographic characteristics, etc.). To determine the conditional probability,  $p$ , of nuclear use

<sup>9</sup> We use the following exclusion criteria to determine those games included in the analysis: games must have lasted at least three complete rounds, all players must have stayed in the game for the duration, and players must have completed the demographic survey associated along with the experiment.

<sup>10</sup> Having a binary outcome variable violates the assumption of linearity in an ordinary least squares regression. Logistic regression addresses this issue via a logarithmic transformation of the outcome variable that allows us to model a non-linear association in a linear way – expressing the regression equation in logarithmic terms (a logit model). For a detailed description of this approach, see Menard (2002). We also include alternative model specifications in the Supplementary Materials to assess the influence of model choice on the result. As logit models assume the absence of multicollinearity, we run tests for multicollinearity between the variables using a variance inflation factor test.

Table III. Logit regression models measuring the probability of nuclear use using the game as the unit of analysis. The values in the table body display the regression coefficients with standard errors in parentheses

	<i>Dependent variable:</i>			
	<i>Nuclear use</i>			
	<i>(1)</i>	<i>(2)</i>	<i>(3)</i>	<i>(4)</i>
Tailored	0.021 (0.218)	0.021 (0.219)	0.006 (0.220)	0.004 (0.220)
Female		-0.056 (0.120)		-0.052 (0.131)
College degree		0.010 (0.127)		-0.004 (0.140)
Age >29		-0.063 (0.123)		-0.081 (0.125)
National security			0.064 (0.144)	0.081 (0.151)
More conservative			0.076 (0.118)	0.083 (0.122)
Reported knowledge			-0.061 (0.127)	-0.075 (0.139)
Constant	0.981 (0.155)	1.112 (0.309)	0.910 (0.257)	1.076 (0.397)
Observations	425	425	425	425
Log likelihood	-248.129	-247.920	-247.675	-247.357

† $p < 0.1$ ; \* $p < 0.05$ ; \*\* $p < 0.01$ .

( $Y = 1$ ) for a given set of predictors, a logit transformation is applied of the form:

$$\ell = \log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k, \quad (1)$$

where  $\ell$  is the log odds and the coefficient values  $\beta_0, \dots, \beta_k$ , are obtained via maximum likelihood estimation. Using this approach,  $\beta_0$  is an offset parameter corresponding to the log odds of nuclear use when the predictor variables are zero and  $\beta_1, \dots, \beta_k$  represents the expected change in log odds of nuclear use for a one-unit increase in the predictor variable. These log odds can, in turn, be used to calculate the odds and probability of a particular outcome using the log transform. For  $\beta_1, \dots, \beta_k$ , a positive or negative coefficient suggests that the predictor of interest has a positive or negative effect on the likelihood of nuclear use, respectively. The tables below report these coefficients as an estimate of the relationship between the predictors and the dependent variable of interest.

#### *Nuclear use by game*

We begin with an analysis of nuclear use by game wherein the dependent variable represents whether the

players use nuclear weapons (of any type) in a given game. The treatment variable is binary, coded as a 0 for those games where tailored nuclear capabilities (i.e. enhanced EMP and HPLY nuclear weapons) are absent and as a 1 when they are present.<sup>11</sup> As a reminder, this article tests the symmetrical (peer competitor) condition in which both Green and Purple have identical nuclear capabilities.

**Game-level results.** Table III provides the results of a game-level logistic regression. Model 1 examines the effect of the presence of tailored nuclear capabilities on nuclear use. The coefficient and standard error suggest that there is a positive but statistically insignificant relationship between the presence of tailored nuclear capabilities and nuclear use. Put another way, there is only a 2% increase in the odds of nuclear use when tailored nuclear capabilities are present – and this effect does not rise to the level of statistical significance. Indeed, as the ‘tailored’ row shows, there is no statistically significant

<sup>11</sup> It is worth noting that enhanced EMP nuclear weapons were rarely employed in the game in comparison with HPLY and HY nuclear capabilities. Indeed, EMP use represents just under 4% of overall nuclear use in the dataset.



difference between the treatment and control games in the sample; this result is consistent regardless of the covariates included in the analysis.

In the second, third, and fourth models, we include demographic variables theorized to influence the decision of a player to employ nuclear capabilities (Press, Sagan & Valentino, 2013; Sagan & Valentino, 2017). Each characteristic is treated as binary on a per-player basis, e.g. *national security* represents whether or not a player has work experience in the national security field, a player coded as *more conservative* has moderate to conservative political leanings and *reported knowledge* represents whether or not a player reports knowledge of nuclear issues. However, it is important to note that the demographic characteristics in the analysis represent an aggregation of player characteristics in the game. For example, the female characteristic of the game ranges from 0, in the case of a game that includes no women, to 3, for a game that comprises all female players. A game that includes two women and one man is scored as a 2, and so on. This approach does not address the potential for interaction effects between players of different types, as others have observed in team settings (Pauly, 2018).

Despite representing the largest wargaming dataset of its kind, we still have only 425 games in our dataset. As a result, we consider covariates in tranches (Model 2 and Model 3) before including all of the demographic characteristics in Model 4. As is the case with the treatment condition (Model 1), these demographic characteristics do not appear to be shaping a decision to use nuclear weapons inside of the SIGNAL wargame environment in a statistically significant manner. Those games that include more players that are women, have a college degree, report knowledge of nuclear issues, and are over 29 years of age have a lower likelihood of using nuclear weapons in the game.<sup>12</sup> Using log transformation, the coefficients for each can be translated into a percentage that reflects how adding an additional player with a specific characteristic will affect the likelihood of nuclear use. We find in Model 4, for example, that each addition of a player over the age of 29 reduces the probability of nuclear use by 8.4%, all else equal. Those games that include more players that report a background working in national security and

are more conservative report a higher likelihood of nuclear use – although, again, without rising to the level of statistical significance.<sup>13</sup>

The results of these analyses are also depicted in graphical form in Figure 2. It is clear from this visualization that each parameter estimate (i.e.,  $\beta_1, \dots, \beta_k$ ) overlaps with zero within the estimated uncertainties. That is, there is no statistically significant difference in the likelihood of nuclear use between the treatment and control games, irrespective of the particular covariates included in the analysis. While not statistically significant, in Models 1–4, the inclusion of tailored nuclear weapons in the arsenal results in an increased likelihood of nuclear use.

The differences in findings between reported knowledge (decreased likelihood of nuclear use) and experience working in national security roles (increased likelihood of nuclear use) are particularly interesting in light of current debates concerning the appropriateness of sampling elites and non-elites. Specifically, some have questioned the appropriateness of using non-elite samples to address research questions pertaining to national and international security issues – with elites generally understood to be current or former senior policymakers in government (Oberholtzer et al., 2019). Others have found little quantitative evidence for gaps between elite and non-elite behavior (Kertzer, 2022). Using data concerning education, self-reported subject matter expertise, and occupation, there appears to be only negligible differences across games that include higher or lower numbers of players with these markers of ‘elite-ness’.

**Addressing final round effects.** The SIGNAL wargame has a fixed number of rounds known to players at the outset of the game. To explore whether the findings reported above may be driven in part by players deciding to use their nuclear capabilities in the final round of the game without fear of reciprocal action, we extend the analysis of the game-level data to examine a subset in which the final round of play is discarded from the dataset.<sup>14</sup> The results of this analysis are shown

<sup>12</sup> In an attempt to speak to the existing survey experiment literature, our demographic covariates reflect the models in Sagan & Valentino (2017). The age cut-off at 30 years of age represents one of the exceptions and was made given the available sample (very few players of SIGNAL are over the 50 years of age cut-off used in Sagan and Valentino’s work).

<sup>13</sup> The demographic survey allows a respondent to select a moderate political ideology, in addition to liberal and conservative leanings. For the purposes of this analysis, we code moderates as conservative for the results reported in Table III.

<sup>14</sup> For a game that runs to four rounds, for example, only the first three rounds would be included in the analysis. This theoretical concern regarding iterated versus non-iterated games is well-established in the existing literature (Axelrod & Hamilton, 1981).

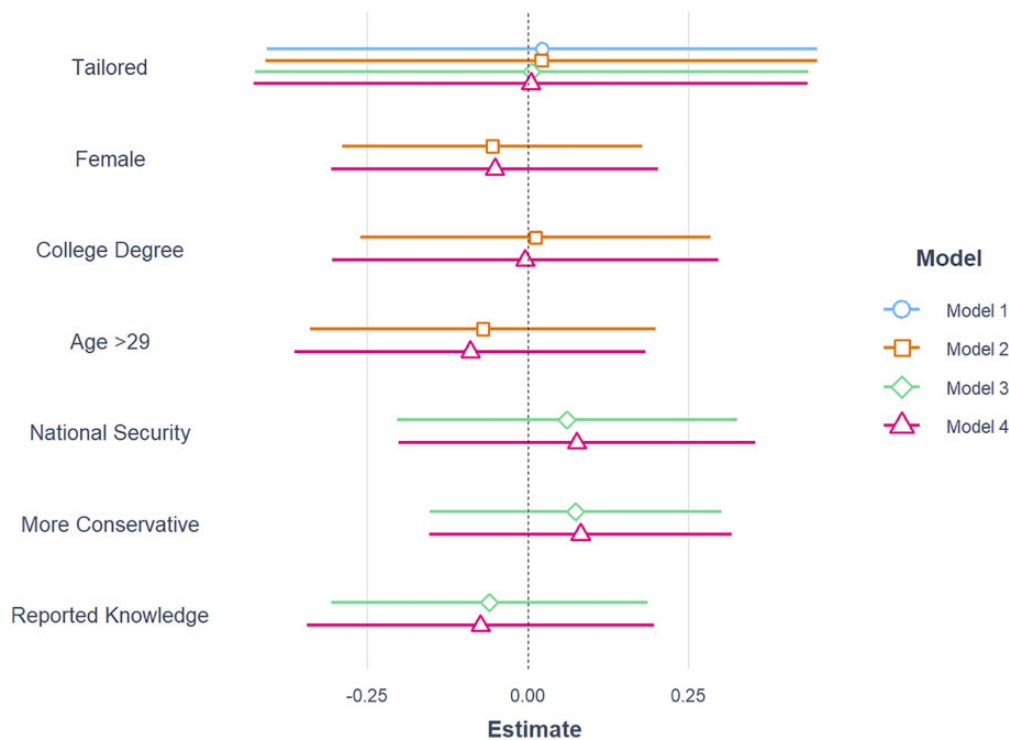


Figure 2. Logit regression models measuring the probability of nuclear use. The estimate represents the coefficient obtained in a game-level analysis via a maximum likelihood approach. The error bar corresponds to the standard deviation of the parameter estimate

in Table IV. Once again, the treatment has a positive (and not statistically significant) effect on the probability of nuclear use, although it is worth noting that the coefficient associated with the experimental treatment is considerably larger in Model 5 (0.214) compared to Model 1 (0.021). Put another way, when the last round is removed, the odds of nuclear use rise 24% when tailored nuclear capabilities are in the game compared with the 2% rise associated with the same treatment noted above. Further, the coefficient obtained in Model 5 is positive within one standard deviation.

Comparing the demographic coefficients from Model 4 in Table III and Model 6 in Table IV, the coefficients associated with age and education also shift considerably, suggesting that younger players and those with college degrees are most sensitive to last-round effects. All in all, these results suggest that when the last round of game data are taken out of the analysis, the presence of tailored nuclear capabilities may have a greater impact on the likelihood of nuclear use in the game, all else equal.

**Testing for substitution.** Recall that both the instability and stability schools discussed above noted the potential for tailored nuclear capabilities substituting for their

high-yield counterparts. To test this proposition, we compare the use of high-yield nuclear capabilities across the two experimental conditions using the same modeling approaches as outlined above. To construct this dependent variable, we create a dichotomized measure of whether a player uses a high-yield nuclear card in the game or not. The results of these analyses are included in Table V.

Models 7 and 8 suggest that those games that include tailored nuclear capabilities decrease the odds of high-yield nuclear use by approximately 12% compared with those games wherein only high-yield nuclear weapons are present. These negative findings are not statistically significant, however. Interestingly, those players that report a background in national security appear more likely to use high-yield nuclear weapons, all else equal. Taken together, these results suggest that there is good reason to further interrogate the substitution effect associated with tailored nuclear capabilities in future work.

Summarizing the analysis of the game-level data, nuclear use appears more likely when tailored nuclear capabilities are present. Additionally, high-yield nuclear use appears less likely when tailored nuclear capabilities are present. This suggests that the substitution effect

Table IV. Logit models examining the probability of nuclear use using game-level data omitting the final round of play. The values in the table body display the regression coefficients with standard errors in parentheses

	<i>Dependent variable:</i>	
	<i>Nuclear use</i>	
	(5)	(6)
Tailored	0.214 (0.202)	0.205 (0.203)
Female		-0.075 (0.121)
College degree		0.086 (0.130)
Age > 29		-0.167 (0.116)
National security		0.020 (0.138)
More conservative		0.049 (0.113)
Reported knowledge		-0.084 (0.129)
Constant	0.437 (0.142)	0.618 (0.365)
Observations	425	425
Log likelihood	-278.817	-277.565

†*p*<0.1; \**p*<0.05; \*\**p*<0.01.

proposed in the theoretical literature is reflected in player behavior.

*Nuclear use by player*

The second set of statistical models presented here uses the player as the unit of analysis. Rather than coding the game based on whether it is a treatment or control game, these analyses examine player actions given their random assignment of nuclear capabilities. Mirroring the analyses above, a player in the treatment game (coded as 1) has access to high-yield and tailored nuclear weapons. A player in the control game (coded as 0) only has access to high-yield nuclear weapons. Non-nuclear players were removed from the dataset as they could not be reasonably expected to cross the nuclear threshold without the requisite capabilities.

With the additional granularity of the player-level data, we ask two related questions: what are the determinants of an individual player deciding to use nuclear weapons; and what are the determinants of an individual player deciding to use nuclear weapons *first*? The distinction between *nuclear use* and *nuclear first use* is

Table V. Logit regression models measuring the probability of high-yield nuclear use using the game as the unit of analysis. The values in the table body display the regression coefficients with standard errors in parentheses

	<i>Dependent variable:</i>	
	<i>HY nuclear use</i>	
	(7)	(8)
Tailored	-0.110 (0.212)	-0.107 (0.216)
Female		0.096 (0.127)
College degree		-0.111 (0.136)
National security		0.337* (0.149)
Age >29		-0.018 (0.122)
More right		-0.330** (0.121)
Reported knowledge		-0.013 (0.137)
Constant	0.910** (0.153)	1.344** (0.391)
Observations	425	425
Log likelihood	-259.057	-252.390

†*p*<0.1; \**p*<0.05; \*\**p*<0.01.

particularly important given that the cognitive drivers for each may be different.<sup>15</sup> In simple terms, a decision to *escalate* a conventional war to a nuclear conflict is, at least in theory, distinct from a decision to reciprocate in kind.

We once again turn to a series of logit models to examine the effects of the treatment on the likelihood of a player to use nuclear weapons. The results of these analyses are shown in Table VI. Models 9 and 10 examine the effects of the predictors on nuclear first use. The coefficient of 0.009 and standard error of 0.143 reported in Model 9 suggest that the effects of the additional tailored nuclear capabilities have a negligible impact on the likelihood of nuclear first use. Model 10, that takes demographic characteristics into account, also suggests that the treatment has little impact on nuclear first use. The results of Model 10 also suggest that female players and those players that

<sup>15</sup> Incidentally, a number of traditional wargames end at the point at which nuclear weapons are used – missing potentially interesting next-turn dynamics (Pauly, 2018).

Table VI. Logit models predicting the probability of nuclear use and nuclear first use using the player as the unit of analysis. The values in the table body display the regression coefficients with standard errors in parentheses

	<i>Dependent variable:</i>			
	<i>Nuclear first use</i>		<i>Nuclear use</i>	
	(9)	(10)	(11)	(12)
Tailored	0.009 (0.143)	-0.021 (0.144)	0.021 (0.154)	0.017 (0.156)
Female		-0.334* (0.141)		-0.119 (0.146)
College degree		-0.163 (0.168)		-0.018 (0.181)
Age > 29		0.078 (0.160)		-0.095 (0.173)
National security		0.062 (0.214)		-0.005 (0.227)
More conservative		-0.110 (0.150)		0.020 (0.161)
Reported knowledge		-0.303 <sup>†</sup> (0.171)		-0.147 (0.181)
Constant	-0.560 (0.102)	-0.197 (0.178)	0.981 (0.110)	1.133 (0.194)
Observations	850	843	850	843
Log likelihood	-557.664	-547.565	-496.257	-491.282

<sup>†</sup> $p < 0.1$ ; \* $p < 0.05$ ; \*\* $p < 0.01$ .

reported subject matter expertise in the national security field are less likely to use nuclear weapons first in the SIGNAL wargaming environment with significance of  $p < 0.05$  and  $p < 0.10$ , respectively. This is an important finding given that it is at odds with Sagan and Valentino's finding that women are just as likely as men to assume hawkish behavior (Sagan & Valentino, 2017).

Models 11 and 12 return to an analysis of nuclear use rather than nuclear first use. Model 11 reports a similar positive, statistically insignificant result (0.021). Model 12 once again reports the demographic results – none of which are statistically significant and all of which are broadly in line with the game-level analysis – suggesting that the aggregation of player actions to the game level does not meaningfully alter the findings.

These analyses using the player as the unit of analysis provide two important insights. First, there are important differences between nuclear first use and subsequent nuclear use as evidenced by comparison of Tables VI and III. Second, the presence of tailored nuclear capabilities continues to have a positive – but statistically insignificant – effect on nuclear use at the player level (Models 11 and 12).

## Method comparison

When developing a new method of inquiry, we would ideally validate the approach against the empirical record. However, one of the primary justifications for developing a new methodological approach to interrogate nuclear issues is the dearth of empirical data with which to test existing theories regarding nuclear deterrence and conflict escalation, particularly in the context of novel and emerging technologies. To address this challenge, we use a survey experiment to explore the same research question examined using the SIGNAL experimental wargaming platform and compare the findings.

### *SIGNAL survey design*

The SIGNAL survey is a three-segment factorial vignette experiment designed to approximate a series of scenarios faced by players inside of the SIGNAL wargame environment. In the survey, respondents provide recommendations to their state leader in the face of an evolving crisis. We randomly assign military capabilities to both the survey respondent and the fictional adversary in three ways (no nuclear capabilities, high-yield nuclear capabilities, high-yield and tailored nuclear capabilities), resulting in a  $3 \times 3$ , between-subjects survey experiment

Table VII. Logit regression models measuring the likelihood of recommending nuclear use in the SIGNAL survey experiment. The values in the table body display the regression coefficients with standard errors in parentheses

	<i>Dependent variable:</i>			
	<i>Nuclear use</i>		<i>High-yield nuclear use</i>	
	(13)	(14)	(15)	(16)
Tailored	1.316** (0.293)	1.445** (0.322)	-0.652* (0.322)	-0.830* (0.356)
Age > 29		-1.087** (0.356)		-1.227** (0.375)
Female		0.552 (0.338)		-0.167 (0.369)
College degree		0.229 (0.343)		1.128** (0.432)
Reported knowledge		0.481 (0.339)		0.150 (0.383)
National security		1.010 (0.669)		0.622 (0.595)
More conservative		0.641* (0.325)		0.643 <sup>†</sup> (0.358)
Constant	-0.722 (0.209)	-1.025 (0.517)	-0.722 (0.209)	-1.076 (0.558)
Observations	208	207	208	207
Log likelihood	-133.424	-120.165	-118.044	-102.720

<sup>†</sup> $p < 0.1$ ; \* $p < 0.05$ ; \*\* $p < 0.01$ .

design. For the purposes of this article, we are concerned with the two conditions that are corollaries to the treatment and control conditions in the SIGNAL experimental wargame environment described above.<sup>16</sup>

In the first segment of the vignette, respondents faced a scenario in which an adversary plans to build a military base in a near neighbor. Then, respondents faced an unattributed cyber attack. Finally, respondents faced a nuclear threat scenario. The baseline vignettes remained the same across treatments with the experiment introducing two sources of variation.

First, we vary the capabilities ascribed to the adversary based upon the experimental condition randomly assigned to the research subject, where the notional adversary was randomly assigned no nuclear capabilities, only high-yield nuclear weapons, or HPLY nuclear weapons, enhanced-EMP nuclear weapons, and high-yield nuclear weapons. Second, we vary the military, economic, and diplomatic policy responses that players could choose to advise on the basis of their randomly assigned treatment.

#### *Survey experiment results*

Here, we interrogate the impact of the presence of tailored nuclear capabilities on the respondent's policy advice. To best approximate the decisionmaking process faced by players in the SIGNAL wargame environment, we used all three segments in the multisegment survey as the unit of analysis. The treatment condition refers to the case in which respondents (and the fictional adversary) are provided with high yield and tailored nuclear capabilities. The control condition refers to the case in which respondents (and the adversary) are provided with only high-yield nuclear capabilities. Nuclear use, for the purposes of the survey experiment, refers to respondents choosing any of the three nuclear use policy options.

Table VII provides the results of these analyses. As shown in Model 16, the coefficient estimating the effect of tailored nuclear capabilities (1.136) suggests that respondents provided with HPLY and enhanced-EMP nuclear weapons in addition to high-yield nuclear capabilities are approximately three times as likely to use nuclear weapons in comparison with respondents who are provided with only high-yield nuclear capabilities. This finding is statistically significant to the  $p < 0.01$  level. In Model 17, we assess the same demographic

<sup>16</sup> There are 208 players in the sample with all survey respondents answering two comprehension questions during the survey.



Table VIII. Logit regression models measuring the likelihood of recommending nuclear use by rank in the SIGNAL survey experiment. The values in the table body display the regression coefficients with standard errors in parentheses

	<i>Dependent variable:</i>		
	<i>Top 5 rank (17)</i>	<i>Top 3 rank (18)</i>	<i>Top rank (19)</i>
Tailored	0.933** (0.291)	0.835** (0.294)	0.652 (0.442)
Constant	-0.856 (0.214)	-0.950 (0.219)	-2.357 (0.349)
Observations	208	208	208
Log likelihood	-135.370	-133.468	-75.273

† $p < 0.1$ ; \* $p < 0.05$ ; \*\* $p < 0.01$ .

covariates used above and find that those respondents 30 years of age and older (-1.087) have a lower likelihood of recommending nuclear use, while those that identify as more politically conservative (0.641) have a higher likelihood of recommending nuclear use – consistent with Sagan and Valentino's recent work (Sagan & Valentino, 2017).

In Models 15 and 16, we test the likelihood of respondents recommending the use of high-yield nuclear capabilities with and without the presence of tailored nuclear capabilities in the arsenal. If there is a substitution effect, as found in the wargaming above, we would expect to see a negative coefficient associated with the tailored condition – particularly as overall nuclear use, as established in Models 13 and 14, is higher in this condition. As in the SIGNAL experimental wargame, the survey data suggests that the presence of tailored nuclear capabilities has a negative effect (-0.652) on the likelihood to recommend high-yield nuclear use. That is, respondents with tailored nuclear capabilities are half as likely to employ their high-yield nuclear weapons. Unlike the results of the wargame analysis above, this finding is statistically significant.

As respondents may select as many policy recommendations as they deem appropriate, an analysis of how respondents rank nuclear options in their guidance may help to better understand whether respondents viewed their recommendation as an important strategic decision. The results of these analyses are included in Table VIII. Here, we examine the effect of the additional tailored nuclear capabilities on the likelihood of respondents ranking nuclear use in the top five, top three, or as the top policy option shown in Models 17–Model 19, respectively. In Model 17, the presence of tailored

nuclear capabilities has the same positive, statistically significant effect (0.933) reported above – although the coefficient is lower than the unranked analysis. In Model 18, which examines when nuclear use is ranked within the top three recommendations, the coefficient (0.835) further decreases. Model 19, wherein nuclear use is ranked as the top policy guidance, reports a positive estimate (0.652), but the finding is no longer statistically significant. Taken together, these models suggest that the effect of tailored nuclear capabilities on the decision to recommend nuclear use lessens as the respondent's commitment to that option strengthens. This suggests that respondents may ultimately prefer alternative capabilities but that the presence of tailored capabilities places nuclear weapon use squarely on the table.

In summary, our analysis of the SIGNAL survey data suggests that there is a positive, statistically significant relationship between the presence of tailored nuclear capabilities and nuclear use – under symmetric conditions. The survey data also provide further evidence for a substitution effect, whereby tailored nuclear capabilities are likely to be used in lieu of their high-yield counterparts. While there are similarities in the findings between the wargame and survey analysis, there are also differences that point to the different laboratory effects across the two experimental environments worthy of further study.

## Conclusion

Across wargaming and survey methods, the evidence presented in this article finds limited support for the proposition that tailored nuclear capabilities increase the likelihood of crossing the nuclear threshold. Amid policy debates concerning the appropriate mix of nuclear capabilities as nuclear weapons states modernize their arsenals, in general, and concerns surrounding the proliferation of low-yield nuclear weapons, in particular, this finding suggestive of the destabilizing consequences of tailored nuclear capabilities raises important questions for both academics and policymakers to consider. Our analysis also finds support for the proposition that low-yield nuclear weapons substitute for their high-yield counterparts – suggesting that even in nuclear conflict, players internalize distinctions in the use of different types of force.

This article also showcases the use of experimental wargaming methods to create an immersive environment for carrying out quantitative social science research. The results discussed above also point to important differences between experimental wargames and surveys as

data-generating processes. Further work is undoubtedly needed to understand the laboratory effects associated with wargame design. For example, do team-based decisionmaking processes yield different results than the individual-level decisionmaking implemented in SIGNAL? Does the number of players inside a game setting matter – would three-player games evolve differently than 10-player games? Would different win conditions yield different results? How might digital vs. analog settings influence player behavior?<sup>17</sup> To answer these methodological questions, we look forward to scholars of behavioral social science implementing, manipulating, and testing experimental wargame designs.

Perhaps most significantly, this work represents a model framework that combines experimental and gaming methods to interrogate research questions pertaining to international security. Our approach addresses some of the methodological concerns associated with alternative synthetic data-generating processes – from formal models that bake in simplifying assumptions to traditional wargames that rely on adjudication and offer idiosyncratic results. Moreover, the development of this method and its comparative benefits vis-à-vis existing approaches offers particular advantages with regard to data-starved policy and academic debates concerning the risks posed by emerging military capabilities – from hypersonic missiles to the integration of ‘AI technologies’ – that the existing literature struggles to adjudicate.

While Quinlan – quoted at the top of this article – is right that scholars do not have empirical data regarding nuclear weapons use, it is our hope that this work serves as an initial demonstration of the potential utility of experimental wargaming for large-*N* analysis by revisiting a long-held and policy-relevant research question related to deterrence, strategy and international security (Quinlan, 2009). For research questions in which observational data are limited, experimental wargaming methods represent a compelling new tool for social science inquiry.

### Replication Data

The dataset, Online appendix, codebook, and replication files for the empirical analysis in this article are available at <https://www.prio.org/jpr/datasets/>. All analyses were conducted using R.

<sup>17</sup> A board game version of SIGNAL was developed as an aid in the design of the online SIGNAL wargame. The laboratory effects across the SIGNAL board game, online game, table-top exercise, and survey experiment are the subject of a forthcoming article.


### Acknowledgements

We would like to acknowledge Jon Whetzel, Nathan Fabian and their team of graduate and undergraduate students at UC Berkeley for their work building the SIGNAL Online game. We would also like to thank our colleagues who have sharpened our methodological process and theoretical discussion. In particular, we would like to thank Jason Reinhardt, Kiran Lakkaraju, Joshua Letchford, Jacquelyn Schneider, Reid Pauly, Joshua Kertzer, Rose McDermott, Andrew Goodhart, Erik Gartzke, Sheryl Hingorani, Brad Roberts, Wes Spain, and Michael Nacht along with two anonymous reviewers for their helpful comments.

### Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This publication was made possible in part by a grant from Carnegie Corporation of New York. This material is based upon work supported in part by the Department of Energy National Nuclear Security Administration through the Nuclear Science and Security Consortium under Award Number DE-NA0003180. The statements made and views expressed are solely the responsibility of the authors.

### ORCID iDs

Andrew W Reddie  <https://orcid.org/0000-0003-3231-8307>

Bethany L Goldblum  <https://orcid.org/0000-0002-9274-9240>

### References

- Acton, James M (2015) Hypersonic boost–glide weapons. *Science & Global Security* 23(3): 191–219.
- Armenta, Mikaela Lea; Laura Wyman Edwards Epifanovskaya, Joshua Letchford, Kiran Lakkaraju, Jonathan Whetzel, Bethany Goldblum, Andrew Reddie & Jake Tibbetts (2020) SIGNAL game manual. Technical Report SAND-2020-7100. Sandia National Laboratories (<https://doi.org/10.2172/1643226>).
- Asal, Victor (2005) Playing games with international relations. *International Studies Perspectives* 6(3): 359–373.
- Axelrod, Robert & William Donald Hamilton (1981) The evolution of cooperation. *Science* 211(4489): 1390–1396.
- Blackett, PMS (1958) Nuclear weapons and defence: Comments on Kissinger, Kennan, and King-Hall. *International Affairs (Royal Institute of International Affairs 1944–)* 34(4): 421–434.
- Blair, Dennis; Michael Carns & Vincent Vitto (2004) Report of the Defense Science Board Task Force on future strategic strike forces. Technical report. Office of the

- Under Secretary of Defense For Acquisition, Technology, and Logistics (<https://apps.dtic.mil/sti/pdfs/ADA421606.pdf>).
- Brewer, Garry D & Bruce G Blair (1979) War games and national security with a grain of salt. *Bulletin of the Atomic Scientists* 35(6): 18–26.
- Broad, William J & David E Sanger (2016) As US modernizes nuclear weapons, ‘smaller’ leaves some uneasy. *New York Times* 11 January (<https://www.nytimes.com/2016/01/12/science/as-us-modernizes-nuclear-weapons-smaller-leaves-some-uneasy.html>).
- Brodie, Bernard (1955) Strategy hits a dead end. *Harpers Magazine* (<https://harpers.org/archive/1955/10/strategy-hits-a-dead-end/>).
- Brodie, Bernard; Frederick Sherwood Dunn, Arnold Wolfers, Percy Ellwood Corbett & William Thornton Rickert Fox (1946) *The Absolute Weapon: Atomic Power and World Order*. New York: Harcourt.
- Burr, William (2005) The Nixon administration, the ‘horror strategy,’ and the search for limited nuclear options, 1969–1972. *Journal of Cold War Studies* 7(3): 34–78.
- Buzzard, Anthony W (1956) Massive retaliation and graduated deterrence. *World Politics* 8(2): 228–237.
- Colby, Elbridge A (2014) The need for limited nuclear options. In: David Ochmanek & Michael Sulmeyer (eds) *Challenges in US National Security Policy: A Festschrift Honoring Edward L (Ted) Warner*. Santa Monica, CA: The Rand Corporation, 141–168.
- Colby, Elbridge A & Michael S Gerson (2013) *Strategic Stability: Contending Interpretations*. Carlisle, PA: Strategic Studies Institute and US Army War College Press.
- Colby, Elbridge; Avner Cohen, William McCants, Bradley Morris & William Rosenau (2013) The Israeli ‘nuclear alert’ of 1973: Deterrence and signaling in crisis. Technical Report DRM-2013-U-004480-Final. Center for Naval Analyses, Strategic Studies Research Department ([https://www.cna.org/CNA\\_files/PDF/DRM-2013-U-004480-Final.pdf](https://www.cna.org/CNA_files/PDF/DRM-2013-U-004480-Final.pdf)).
- Daugherty, William; Barbara Levi & Frank Von Hippel (1986) The consequences of ‘limited’ nuclear attacks on the United States. *International Security* 10(4): 3–43.
- Davis, Paul K; J Michael Gilmore, David R Frelinger, Edward Geist, Christopher K Gilmore, Jenny Oberholtzer & Danielle C Tarraf (2019) *Exploring the Role Nuclear Weapons Could Play in Deterring Russian Threats to the Baltic States*. Santa Monica, CA: RAND Corporation.
- Dowler, Thomas W; II Howard & H Joseph (1991) Countering the threat of the well-armed tyrant: A modest proposal for small nuclear weapons. *Strategic Review* 19(4): 34–40.
- Doyle, James E (2016) Better ways to modernise the US nuclear arsenal. *Survival* 58(4): 27–50.
- Doyle, James E (2017) On integrating conventional and nuclear planning. *Arms Control Today* 47(2): 43.
- Draper, Norman R & Harry Smith (1998) *Applied Regression Analysis, Vol. 326*. New York: John Wiley & Sons.
- Facini, Andrew (2020) The low-yield nuclear warhead: A dangerous weapon based on bad strategic thinking. *Bulletin of Atomic Scientists* 28 January (<https://thebulletin.org/2020/01/the-low-yield-nuclear-warhead-a-dangerous-weapon-based-on-bad-strategic-thinking/>).
- Freedman, Lawrence & Jeffrey H Michaels (2019) *The Evolution of Nuclear Strategy* (4th ed.). London: Palgrave Macmillan.
- Gartzke, Erik; Jeffrey M Kaplow & Rupal Mehta (2015) The determinants of nuclear force structure. In: Neil Narang, Erik Gartzke & Matthew Kroenig (eds) *Non-proliferation Policy and Nuclear Posture*. Routledge, 191–220.
- Glaser, Charles L & Steve Fetter (2005) Counterforce revisited: Assessing the nuclear posture review’s new missions. *International Security* 30(2): 84–126.
- Halperin, Morton H (1961) Nuclear weapons and limited war. *Journal of Conflict Resolution* 5(2): 146–166.
- Heginbotham, Eric; Michael S Chase, Jacob L Heim, Bonny Lin, Mark R Cozad, Lyle J Morris, Christopher P Twomey, Forrest E Morgan, Michael Nixon & Cristina L Garafola, et al. (2017) *China’s Evolving Nuclear Deterrent: Major Drivers and Issues for the United States*. Santa Monica, CA: RAND Corporation.
- Heimer, Brandon Walter (2018) Standoff over the LRSO: Assessing the long-range stand-off missile’s impact on strategic stability. Technical Report SAND2018-6969 R. Sandia National Laboratories.
- Hyde, Susan D (2015) Experiments in international relations: Lab, survey, and field. *Annual Review of Political Science* 18: 403–424.
- Jensen, Benjamin & Brandon Valeriano (2019) Cyber escalation dynamics: Results from war game experiments. Working paper. International Studies Association, Annual Meeting Panel: War Gaming and Simulations in International Conflict (<http://web.isanet.org/Web/Conferences/Toronto%202019-s/Archive/71e7820c-e61c-4187-ab8c-28de83dfd660.pdf>).
- Kertzer, Joshua D (2022) Re-assessing elite–public gaps in political behavior. *American Journal of Political Science* 66(3): 539–553.
- Kissinger, Henry A (1960) Limited war: Conventional or nuclear? A reappraisal. *Daedalus* 89(4): 800–817.
- Kroenig, Matthew (2015) Facing reality: Getting nato ready for a new cold war. *Survival* 57(1): 49–70.
- Kroenig, Matthew (2016) Toward a more flexible NATO nuclear posture: Developing a response to a Russian nuclear de-escalation strike. Technical report. Atlantic Council ([https://www.atlanticcouncil.org/wp-content/uploads/2016/11/Toward\\_a\\_More\\_Flexible\\_NATO\\_Nuclear\\_Posture\\_web\\_1115.pdf](https://www.atlanticcouncil.org/wp-content/uploads/2016/11/Toward_a_More_Flexible_NATO_Nuclear_Posture_web_1115.pdf)).

- Kroenig, Matthew (2018) The case for tactical US nukes. *Wall Street Journal* 24 January (<https://www.wsj.com/articles/the-case-for-tactical-u-s-nukes-1516836395>).
- Larsen, Jeffrey A & Kerry M Kartchner (2014) *On Limited Nuclear War in the 21st Century*. Stanford, CA: Stanford University Press.
- Letchford, Joshua; Laura Epifanovskaya, Kiran Lakkaraju, Mika Armenta, Andrew Reddie, Bethany L Goldblum, Jon Whetzel et al. (2022) Experimental wargaming with SIGNAL. *Military Operations Research* 27(2): 59–82.
- Levi, Michael A (2004) Dreaming of clean nukes. *Nature* 428(6986): 892.
- Lieber, Keir A & Daryl G Press (2009) The nukes we need: Preserving the American deterrent. *Foreign Affairs* 88: 39–51.
- Lieber, Keir A & Daryl G Press (2013) The new era of nuclear weapons, deterrence, and conflict. *Strategic Studies Quarterly* 7(1): 3–14.
- Lieber, Keir A & Daryl G Press (2017) The new era of counterforce: Technological change and the future of nuclear deterrence. *International Security* 41(4): 9–49.
- Lin-Greenberg, Erik (2018) *Wargame of Drones: Remotely Piloted Aircraft and Crisis Escalation. Working Paper*. Massachusetts Institute of Technology ([https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3288988](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3288988)).
- Lin-Greenberg, Erik; Reid B.C. Pauly & Jacquelyn G. Schneider (2022) Wargaming for political science research. *European Journal of International Relations* 28(1): 83–109.
- Long, Austin (2018) Discrimination details matter: Clarifying an argument about low-yield nuclear warheads. *War on the Rocks* 16 February (<https://warontherocks.com/2018/02/discrimination-details-matter-clarifying-argument-low-yield-nuclear-warheads/>).
- McIntyre, Matthew H; Stephen Peter Rosen, Dominic DP Johnson, Rose McDermott, Emily S Barrett, Jonathan Cowden & Richard Wrangham (2006) Overconfidence in wargames: Experimental evidence on expectations, aggression, gender and testosterone. *Proceedings of the Royal Society: Biological Sciences* 273(1600): 2513–2520.
- Menard, Scott (2002) *Applied Logistic Regression Analysis*. No. 106. Thousand Oaks, CA: SAGE.
- Narang, Vipin (2018) The discrimination problem: Why putting low-yield nuclear weapons on submarines is so dangerous. *War on the Rocks* 8 February (<https://warontherocks.com/2018/02/discrimination-problem-putting-low-yield-nuclear-weapons-submarines-dangerous/>).
- Nitze, Paul H (1956) Atoms, strategy and policy. *Foreign Affairs* 34(2): 187–198.
- Oberholtzer, Jenny; Abby Doll, David Frelinger, Karl Mueller & Stacie Pettyjohn (2019) Applying wargames to real-world policies. *Science* 363(6434): 1406–1406.
- Osgood, Robert E (1979) *Limited War Revisited*. Boulder, CO: Westview.
- Pauly, Reid B C (2018) Would US leaders push the button? Wargames and the sources of nuclear restraint. *International Security* 43(2): 151–192.
- Perla, Peter P (1990) *The Art of Wargaming: A Guide for Professionals and Hobbyists*. Annapolis, MD: Naval Institute Press.
- Perla, Peter P & ED McGrady (2011) Why wargaming works. *Naval War College Review* 64(3): 1–20.
- Podvig, Pavel (2018) Russia's current nuclear modernization and arms control. *Journal for Peace and Nuclear Disarmament* 1(2): 256–267.
- Potter, William C; Nikolai Sokov, Harald Müller & Annette Schaper (2000) *Tactical Nuclear Weapons: Options for Control*. Geneva: United Nations Institute for Disarmament Research.
- Powell, Robert (1988) Nuclear brinkmanship with two-sided incomplete information. *American Political Science Review* 82(1): 155–178.
- Powell, Robert (1990) *Nuclear Deterrence Theory: The Search for Credibility*. Cambridge: Cambridge University Press.
- Press, Daryl G; Scott D Sagan & Benjamin A Valentino (2013) Atomic aversion: Experimental evidence on taboos, traditions, and the non-use of nuclear weapons. *American Political Science Review* 107(1): 188–206.
- Quek, Kai (2016) Nuclear proliferation and the use of nuclear options: Experimental tests. *Political Research Quarterly* 69(2): 195–206.
- Quinlan, Michael (2009) *Thinking About Nuclear Weapons: Principles, Problems, Prospects*. Oxford: Oxford University Press.
- Rathbun, Brian C; Joshua D Kertzer & Mark Paradis (2017) Homo diplomaticus: Mixed-method evidence of variation in strategic rationality. *International Organization* 71(S1): S33–S60.
- Reddie, Andrew W; Bethany L Goldblum, Kiran Lakkaraju, Jason Reinhardt, Michael Nacht & Laura Epifanovskaya (2018) Next-generation wargames. *Science* 362(6421): 1362–1364.
- Roblin, Sebastien (2019) US submarines are getting new W76-2 tactical nuclear warheads (and it might be a giant mistake). *National Interest* 21 December (<https://nationalinterest.org/blog/buzz/us-submarines-are-getting-new-w76-2-tactical-nuclear-warheads-and-it-might-be-giant>).
- Rovere, Crispin & Kalman Robertson (2013) Non-strategic nuclear weapons: The next step in multilateral arms control. *Australian Strategic Policy Institute: Special Report* 62: 1–10.
- Sabin, Philip (2012) *Simulating War: Studying Conflict Through Simulation Games*. New York: Bloomsbury Academic.
- Sagan, Scott D & Jeremi Suri (2003) The madman nuclear alert: Secrecy, signaling, and safety in October 1969. *International Security* 27(4): 150–183.
- Sagan, Scott D & Benjamin A Valentino (2017) Revisiting Hiroshima in Iran: What Americans really think about using nuclear weapons and killing noncombatants. *International Security* 42(1): 41–79.

- Schelling, Thomas C (1966) *Arms and Influence*. New Haven, CT: Yale University Press.
- Schlesinger, James R (1975) The theater nuclear force posture in Europe (<https://www.archives.gov/files/declassification/isca/pdf/2013-066-doc01.pdf>).
- Schneider, Jacquelyn (2017) Cyber and crisis escalation: Insights from wargaming USASOC Futures Forum (<https://paxsims.files.wordpress.com/2017/01/paper-cyber-and-crisis-escalation-insights-from-wargaming-schneider.pdf>).
- Schofield, Julian (2013) Modeling choices in nuclear warfighting: Two classroom simulations on escalation and retaliation. *Simulation & Gaming* 44(1): 73–93.
- Talmadge, Caitlin (2019) The US–China nuclear relationship: Why competition is likely to intensify (<https://www.brookings.edu/research/china-and-nuclear-weapons/>).
- Tannenwald, Nina (1999) The nuclear taboo: The United States and the normative basis of nuclear non-use. *International organization* 53(3): 433–468.
- Tannenwald, Nina (2005) Stigmatizing the bomb: Origins of the nuclear taboo. *International Security* 29(4): 5–49.
- Terraïs, Bruno (2011) In defense of deterrence: The relevance, morality and cost-effectiveness of nuclear weapons. *Proliferation Papers* (39).
- Tingley, Dustin H & Barbara F Walter (2011) The effect of repeated play on reputation building: An experimental approach. *International Organization* 65(2): 343–365.
- Toon, Owen B; Charles G Bardeen, Alan Robock, Lili Xia, Hans Kristensen, Matthew McKinzie, RJ Peterson, Cheryl S Harrison, Nicole S Lovenduski & Richard P Turco (2019) Rapidly expanding nuclear arsenals in Pakistan and India portend regional and global catastrophe. *Science Advances* 5(10): 1–13.
- Von Hippel, Frank N; Barbara G Levi, Theodore A Postol & William H Daugherty (1988) Civilian casualties from counterforce attacks. *Scientific American* 259(3): 36–43.
- Weber, Andrew & Christine Parthemore (2019) Smarter US modernization, without new nuclear weapons. *Bulletin of the Atomic Scientists* 75(1): 25–29.
- Yarhi-Milo, Keren; Joshua D Kertzer & Jonathan Renshon (2018) Tying hands, sinking costs, and leader attributes. *Journal of Conflict Resolution* 62(10): 2150–2179.
- Younger, Stephen M (2000) Nuclear weapons in the twenty-first century. Technical Report LAUR-00-2850. Los Alamos National Laboratory (<https://nuke.fas.org/guide/usa/doctrine/doe/younger.htm>).
- Zagare, Frank C (1985) Toward a reformulation of the theory of mutual deterrence. *International Studies Quarterly* 29(2): 155–169.
- ANDREW W REDDIE, b. 1989, PhD in Political Science (University of California, Berkeley, 2019); Assistant Professor of Practice, School of Information, University of California, Berkeley (2021– ).
- BETHANY L GOLDBLUM, b. 1979, PhD in Nuclear Engineering (University of California, Berkeley, 2007); Associate Research Engineer, Department of Nuclear Engineering, University of California, Berkeley (2012– ); Physicist, Nuclear Science Division, Lawrence Berkeley National Laboratory (2020– ).