SAND2018-6296B

17 Online Games for Studying Human Behavior

Kiran Lakkaraju Laura Epifanovskaya Mallory Stites
Josh Letchford Jason Reinhardt Jon Whetzel

Abstract

Much has been written on the potential for games to enhance our ability to study complex systems. In this chapter we focus on how we can use games to study national security issues. We reflect on the benefits of using games and the inherent difficulties that we must address. As a means of grounding the discussion we will present a case study of a retrospective analysis of gaming data.

INTRODUCTION

Games and simulations have been used as experimental platforms for centuries, particularly in the form of wargames used for planning purposes and for reenactment of historical battles (Sabin, 2014). In these instances, they are used as a type of virtual laboratory to test hypotheses about what will happen (or in the case of historical battles, what might have happened) when a given decision is made under a certain set of conditions. Such games can be highly simplistic or exquisitely detailed; they can be board, card, or internet games; or they can be highly choreographed simulations, sometimes involving multiple (in many cases high-level) personnel from military and government. While the use of games as a didactic and exploratory tool is widespread, they are used less frequently to gather data on player actions for subsequent statistical meta-analysis. Some reasons for this are the small size (in statistical terms) of the participating groups, the differences in the games and simulations themselves from game to game, and the difficulty in acquiring and storing analyzable data in real time.

Online games offer the potential to address some of these difficulties. Online games are designed and developed to be hosted and accessed via the internet, allowing a large and diverse player pool to participate under sets of conditions that are both tunable and recordable by researchers. One particular category of online games, Massively Multiplayer Online Games (MMOGs), is especially intriguing as they can provide data on a large number of players interacting in a shared, persistent world over an extended period of time.

In this chapter we discuss the value of online games[1], including MMOGs, as experimental platforms that might augment or even displace some other data-gathering methodologies in specific areas of research and inquiry. Experimental techniques are a powerful means of identifying causal relationships, which are critical for designing interventions in a system.

We argue that for studying national security issues, where data is sparse, it is difficult to experiment, and behaviors can be complex and varied, online games can serve as a unique and powerful tool to experimentally understand causal relationships.

We begin with a discussion of the potential of online games and the specific applicability to helping address national security issues. We outlines a proof of concept analysis that we performed using data collected from an existing MMOG, which herein will be referred to as Game X to preserve the anonymity of the game, in order to compare conflict phenomena within the game to data from the real world and a scientific analysis of the real-world data selected from the academic international relations literature.

We end with a discussion of considerations to have when using games as experiments. Particular issues arise when developing a game as a vehicle for

[1]See also Guarino et al. (2018) in this volume on online games

studying human behavior. The fundamental problem is the need to engage subjects in the game for extended time periods (on the order of hours to years). In order to naturally encourage this (as opposed to paying subjects), games must be designed in particular ways. These design choices have ramifications on the subject behavior and analysis of data from a game. We outline some research considerations on this topic.

ONLINE GAMES AND MASSIVELY MULTIPLAYER ONLINE GAMES FOR RESEARCH

Online games can span a wide range of characteristics (De Lope and Medina-Medina, 2017; Laamarti et al., 2014). There are many types of online games, ranging from simple single player games such as Fruit Ninja, to more social games such as Words with Friends or Farmville, to tactical heavy action games such as Call of Duty, to highly complex and social role-playing games such as World of Warcraft and Eve Online. The latter games are often referred to as Massively Multiplayer Role Playing Games (MMORPG); however, we will use the term Massively Multiplayer Online Game (MMOG) to refer to all games (whether role playing or not) in which players interact for extended periods of time in a shared, persistent world.

MMOGs are online games that attract players from around the world of all ages, genders, and educational backgrounds to a shared virtual world (Yee, 2006; mmo, 2012). The diversity and size of the player base, which for some games is numbered in millions, is an especially attractive advantage of this type of game as a data-gathering platform. Data on millions of actions performed by a large and diverse sample of people lends itself well to statistical analysis, better than surveys and laboratory experiments with much smaller sample sizes taken from a more homogenous group (e.g., college students, who frequently participate in academic human research studies (Gosling et al., 2010; Henrich et al., 2010)). They also offer the opportunity to see how different types of players respond under different circumstances;

useful for interrogating differences among players, but also, in wargaming, uncovering novel strategies that would not have occurred to personnel typically involved in these games.

There are many types of MMOGs, ranging from simple browser-based games such as Farmville to highly complex and realistic role-playing games such as World of Warcraft and Eve Online. The latter may be of particular interest as an experimental platform, especially in the social sciences, because of their realism, complexity, and degree of player involvement. The use of games as experimental platforms for scientific and other research will, of course, be criticized as providing data that is only meaningful within the game context, and that cannot be used to draw conclusions about the real world (Williams, 2010). The legitimacy of game data for research purposes is a real concern, and one that must be considered both in the experimental design phase and in analyzing and appropriately caveating the results of MMOG experiments. However, research has shown that player behavior in complex and realistic role-playing games may be representative of behavior in the real world because of the investment in time, effort, and reputation made by participants in their player avatar (Castronova, 2008; Lu et al., 2014). In addition, this type of MMOG is often based in a virtual world comprised of highly complex economies and social structures, some of which evolve organically and not as a result of a rule set governing the game platform. Because of the complexity and realism of this type of game, in-game behaviors may mimic real world economic and social behaviors.

Additionally, it is worth noting that many existing methods of data gathering have their own associated problems. In social science research, for example, surveys and questionnaires are used that may collect incomplete and/or biased information; questions may be misunderstood by respondents or may be formulated in such a way as to bias the respondents' answers. As with all scientific data, the analysis of and conclusions drawn from MMOG data must be understood appropriately and the inherent limits of such games acknowledged. Findings from statistical analysis of game data, for example, are easily validated using results from future game play; however, this only demonstrates that the findings are valid within the game context. Conclusions drawn from player actions in an MMOG will be much more difficult to validate in the real world (Williams, 2010). Game data can still be used for statistical analysis and compared to results in the real world; the work that we will describe below is an example of this. As in the real world, statistical results from game data must also be understood not to provide "the answer" to a given research question, but rather, a range of likely answers, a distribution, a mean, and outlying data points which in and of themselves may be interesting to researchers.

Where Is There Benefit?

Online games can provide exceptional value to research efforts that meet any of the following three criteria: data is sparse and difficult to obtain; it is difficult or impossible to build a real-world experimental laboratory with built-in controls to perform the research; the research is interested in finding not only averages and distributions of player behavior, but also exploring a wide range of possible behaviors and examining results that don't fall neatly within a given distribution (i.e., "distillation games" (Perla et al., 2005)). Of course, these types of unexpected results are always the most interesting to science; online games provide a method of capturing and quantifying conditions that creates these results in a way typically unavailable in some realms of scientific inquiry.

Areas where these conditions apply and online games might be leveraged include the social and behavioral sciences, war studies and planning, and international relations research. The last, international relations, meets the criteria because of the sparsity of data usable for quantitative analysis. As will be discussed in detail below, academics often use the Military and Interstate Disputes (MIDs) data collected by the Correlates of War project to mathematically evaluate the effects of various military, economic, and other parameters on the likelihood of conflict between sovereign states. The MIDS data span instances of conflict occurring over less than a century; they do not capture all variables of potential importance to the analysis of conflict; and, crucially, it is not possible to gather significantly more data over the few years funded by a typical academic research grant. As such, this last area is one in which MMOGs can potentially contribute tremendous value as experimental platforms that can generate large amounts of data in condensed timeframes from games engineered to answer specific research questions. Our work described below shows a first attempt to take data from an MMOG, "operationalize" it so that it is captured in a way approximately equivalent to the MIDs data and the economic variables from an international relations journal article that we chose as a comparison study, and perform the same statistical analysis on the game data as was used in the journal article in order to compare the results.

WAR GAMES AND DATA GATHERING FOR NUCLEAR DETERRENCE POLICY

War studies and planning meet all of the criteria used to evaluate online games as research tools that were enumerated in the Introduction: data from wars that can be used in statistical meta-analysis is sparse and difficult to obtain; it is difficult or impossible to create a real-world experimental laboratory with built-in controls to perform the research; events that create new data naturally are undesirable; and the research is interested in exploring a wide range of possible behaviors and examining results that fall outside of the normal distribution. Games and simulations have been used for more than a century to study important historical battles and to envisage future ones. Because they are referred to as "games," they are often disparaged as serious tools of scholarship, although there have been cases where the game contains more historically accurate detail about a given battle than narrative works on the subject (Sabin, 2014). However, war gaming and simulations are taken quite seriously by military planners and policy makers. During the

Cold War period, Nobel Prize-winning economist Thomas Schelling, working at the RAND Corporation, was instrumental in designing and executing war games for military planning purposes. Schelling believed that wargaming was essential in filling a gap in war planning: analysts don't know what they don't know; and they won't think of every possible future battle or crisis contingency. Wargames allow scenarios to unfold without a human mastermind planning and predicting every step, which means that they can sometimes enter new and unusual territory. Reid Pauly summarized this point nicely in what is perhaps the only meta-analysis of United States politico-military wargames published to date, in recounting events surrounding the Cuban Missile Crisis of 1962. As Pauly tells it: "During the Cuban Missile Crisis, a participant in the office of John McNaughton remarked, 'This crisis sure demonstrates how realistic Schelling's [war]games are.' Another responded, 'No, Schelling's games demonstrate how unrealistic this Cuban crisis is."' In the same paper, Pauly examines historical wargames in order to probe the attitudes of "strategic elites" – that is, military and policy professionals with experience and education relevant to combat and nuclear weapons – toward willingness to use a nuclear weapon in combat. This work followed a publication by Scott Sagan and Benjamin Valentino citing the surprising willingness of the American public at large to use a nuclear weapon against an adversary state, in violation of a hypothesized "nuclear taboo" (Press et al., 2013).

One enormous potential benefit of using an MMOG to answer a similar question about the willingness or unwillingness of particular groups of people to deploy a nuclear weapon (or other weapon of war) in combat is the large, diverse group of people that participate in this type of online game. If some basic data are gathered on the player participants (such as occupation and level of educational attainment), then data from game play can be post-processed to evaluate the decisions of various groups and compare them against each other. In fact, it would be fascinating to compare such an analysis of MMOG data to Sagan and Valentino's finding that the public is not overwhelmingly averse to using nuclear weapons in combat and Pauly's finding that strategic elites are averse to their use. Would we find the same difference in attitudes in player behavior in the game setting?

Attitudes toward nuclear use may be a factor in effective – or ineffective – nuclear deterrence. The United States and its allies rely in part on deterrence, in a communicated willingness to retaliate in kind, to protect them from nuclear aggression by nuclear armed adversaries. As Schelling remarked in his landmark 1966 book Arms and Influence: "The power to hurt is bargaining power. To exploit it is diplomacy – vicious diplomacy, but diplomacy." In order for deterrence to work; that is, in order for it to influence an adversary to avoid the proscribed action, the threat of nuclear retaliation must be deemed to be credible. The cold war-era logic of the threat of mutual annihilation as the foundation of deterrence is summarized neatly by the acronym MAD, or mutual assured destruction, a phrase coined in cold war strategist Herman Kahn's Hudson Institute Deudney (1983). The deterrent power of MAD, in part, drove the U.S. and the Soviet Union to invest hugely in their respective nuclear arsenals in order to maintain nuclear parity with each other and preserve global strategic stability.

Today's political and technological environments are markedly different than those of the cold war. The Soviet Union collapsed and splintered in the 1990s, leaving only a nuclear-armed Russia. China, India and Pakistan have acquired nuclear arsenals. In addition to the change in the nuclear landscape, however, the world has also seen changes in political and economic relationships and the rise of new technologies that may affect strategic stability in new ways. Whereas fear of nuclear destruction may have driven strategic military planning during the cold war, for example, fear of economic devastation may serve as a strong deterrent in today's economically-interconnected environment.

Given the (fortunate) paucity of data on nuclear use in conflict, wargames are a valuable means of generating and analyzing scenarios involving nuclear exchange. Nuclear standoffs have been evaluated using wargames for decades; the additional value that the online game brings to the study of these scenarios is the ability to record massive amounts of data from hundreds of players, for later analysis over multiple instances of game play. A chat feature will allow researchers to gather data on player's thoughts and motives surrounding in-game decisions. In addition, the game allows capture not only of data on nuclear use, but also on the conditions of the game environment at the time of use. These metadata can later be evaluated to determine if there are other heretofore unexamined factors that significantly influence nuclear use (or the likelihood of simple conventional conflict, for that matter).

In the section that follows, we will discuss how MMOG data can inform efforts in the academic areas of international relations and policy studies to understand the influence of multiple different factors on the likelihood of conflict between states. These factors include military, economic, and political elements, and their effect on conflict is analyzed using the MIDs variables first discussed in the Introduction. The data collected using an MMOG may be combined with the real-world MIDs data to better study the effects of these variables. The MMOG allows the collection of significant amounts of additional data in years to come, much more (hopefully) than the data that will be generated by real-world wars. In addition, the game can be designed to gather data on variables that may be relevant to conflict but which are not currently in the MIDs cannon. The utility of the MMOG to further academic debate on the subject of economic interdependence and conflict will be examined in detail in the following section.

MMOG DATA TO TEST INTERNATIONAL RELATIONS THEORY

The relationship between economic interdependencies among states and likelihood of conflict between them has been explored using data in the academic fields of political science and international relations, with divergent results. For decades, academics have performed regression analysis on conflict data from the Correlates of War Project to establish linkages between trade volumes and conflict likelihood. The results of these analyses range from evidence that trade increases instances of conflict between states (the view of the "realist" school of international relations theory) to findings that trade decreases conflict likelihood (the view of the "liberal" school), to work testing the assumption that the impact of trade relations on conflict is more complex than the realist/liberal dichotomy would suggest. To date, this dispute in the academic literature has not been resolved.

The research question of whether economic interdependence positively or negatively affects conflict is one that lends itself to study using an MMOG since additional data is desirable for the type of statistical regression analysis typically performed in the literature but is hard to come by in the real world. Data from game play of an online game such as the one described above may be useful in further elucidating the nature of the relationship between economic interdependencies and their effect on the willingness of nations to go to war with each other. As a proof of concept, however, we took data from an extant online serious game with a steady user base; data which members of our project team had collected and used earlier in unrelated research. Game X contains the elements necessary for comparison to real-world research on economic interdependence and conflict: there are guilds within the game (comparable to nation-states in the real world) that trade and wage war with each other, and there are other quantifiable factors within the game that correlate with those captured by the MIDs variables. These parameters include contiguity (whether states are located next to each other geographically; a strong predictor of conflict in the real world), alliance, and capabilities ratio. The capabilities ratio variable in the MIDs dataset is a relative measure of the Composite Index of National Capabilities (CINC) scores of the states within a dyad. The CINC score is based on the population, urban population, iron and steel production, energy consumption, military personnel, and military expenditure of a state. In Game X, we used combat strength ratio, economic strength ratio, and size ratio, all quantities tracked within the game as a measure of player score, as proxies for capabilities ratio. The MIDs variables and comparable Game X variables are shown below in Table 17.1.

MIDS Variable	**Game X Variable**
Dyadic Trade	Trade between two guilds
Contiguity	Share a border
Alliance	Is Foe $\neq$ True (Is Foe is a designation voluntarily selected by a guild to describe another guild)
Capabilities Ratio	Combat Strength Ratio, Economic Strength Ratio, Size Ratio

Table 17.1: Mapping between MIDS variables and Game X variables

Frequently in the international relations literature, linear regression analysis is performed on one or more economic variables to discern a correlation between them and instances of conflict within a dyadic pair. The variables listed above in Table 1 are used as controls. We performed a similar analysis on 739 days of Game X data "operationalized" to populate the economic, conflict, and control variables for direct comparison of our analysis to the academic literature. We began by taking a sample journal article (Barbieri, 1996) that regresses the MIDs data against three different economic variables defined by the author: salience, symmetry, and interdependence. The variables are defined mathematically as follows:

$$\text{Trade Share}_i = \frac{DyadicTrade_{ij}}{TotalTrade_i} \tag{1}$$

$$\text{Salience} = \text{Sqrt}(TradeShare_i * TradeShare_j) \tag{2}$$

$$\text{Symmetry} = 1 - |\text{TradeShare}_i - \text{TradeShare}_j \tag{3}$$

$$\text{Interdependence} = \text{Salience} * \text{Symmetry} \tag{4}$$

The purpose of our work was not to validate or invalidate the variable definitions or the conclusions of the article, but simply to directly compare statistical analysis of serious game data to similar analysis of real world data reported in the literature; to compare results of the analysis and examine how and why they might be different.

To operationalize the economic and conflict data in Game X, we started by measuring combat and trade between guilds over the entire 739-day period. Guilds change over time; few were present and unchanged over the entire measurement period. As such, we divided the time period in 25 consecutive 30-day month periods and calculated the trade and combat measures on a month-to-month basis. Dyads were the main unit of measurement, as defined by a pair of guilds that engaged in trade, combat, or both any point within a month period. The resulting unit of dyad-month is similar to the dyad-year unit used in the Correlates of War database to store the MIDs data. There were a total of 297 guilds and 13,079 unique dyad pairs over the entire time period. We excluded dyads that included the game itself (i.e., game-controlled entities, or "non-player characters") and guilds engaging in trade or conflict within itself. This resulted in 47,748 observations. On average, only 18% of dyads measured engaged in conflict over the 25 periods of data gathered.

Measurements of trade between members of a dyad were modeled after (Barbieri, 1996). First, the trade share for each member of the dyad was calculated by taking the amount of trade with the dyad partner (imports and exports) divided by the total amount of trade conducted by that dyad member with all trading partners (including trade with other guilds, trade with players not belonging to guilds, and trade with the game). See Equation 1 above. Trade Share is bounded by 0 and 1; its value approaches 1 as trade from $Guild_i$ becomes a larger fraction of overall trade conducted by $Guild_j$. From here, three economic variables were calculated: salience, symmetry, and interdependence, defined by Equations 2, 3, and 4 above. Salience was formulated by the journal article author in order to capture the importance of the trade relationship to the dyadic partners; a high salience score should indicate that the relationship is important to at least one of the partners in the dyad. In fact, a highly asymmetric trade relationship might obfuscate the importance to one trading partner (a very small fraction of one multiplied by a larger fraction of one), but our intention here was not to contest the measurement methods employed by the article but simply to replicate them. The variable salience turned out to be very small on average in the Game X dataset (mean = 0.006, range = [0, 0.09], median = 0.0001). The dyadic trade share was often a very small fraction of overall trade for a given guild.

The variable symmetry is defined above in Equation 3. Using this definition, values close to 1 should indicate a relatively symmetrical trade relationship, while very small values should indicate an asymmetric relationship. In the Game X dataset, most trade share values were very small, as noted above. Because of this, most symmetry values were close to 1 (mean = 0.998, range = [0.3, 1]. Interdependence, defined as in Equation 4 above, reflects the problems inherent in both the salience and symmetry values, and because it multiplies the symmetry values, which approach but are still fractions of 1, by the small salience (approaching 0) numbers, it results in each case in very small interdependence scores (mean = 0.0006, range = [0, 0.074]).

Conflict was defined in broad terms for purposes of this analysis; a dyad was coded as having engaged in conflict if any of its members engaged in any sort of combat with the other dyadic guild within the month period. This method of encoding does not capture whether the conflict was part of a larger effort coordinated by the guild, or simply a "one-off" battle between two individuals. As such, this variable in Game X may not correlate well with the conflict variables from the MIDs data set, which capture state-level disputes and wars.

Analysis And Results

Analysis was performed on the data using linear mixed effects regression models run with the lme4 package for R, which allows the inclusion of multiple observations of the same dyad over the full time period. Random intercepts were included for each dyad, such that the overall estimated intercept was adjusted slightly for the error variance due to each dyad's likelihood to engage in combat. This allowed us to look at the effect of the economic predictor variables, also known as the fixed effects, while accounting for variability in the baseline likelihood of conflict between individual dyads. The dependent variable predicted by the logistic regression model was likelihood of conflict between two dyads (bounded by 0 and 1) during the time period analyzed. Logistic (also known as logit) regression was used because of the categorical nature of conflict (either it occurred or it didn't, there is no combat continuum), resulting in a binary dependent variable that can take either a value of 0 or 1. The fit to the data from the logit model generates an equation that estimates a probability, or likelihood, of the dependent variable occurring (having a value of 1) with a given value of the independent variable. The equation incudes an intercept and a coefficient applied to the independent variable, which are shown in the results tables that follow. The equation takes the form:

$$p(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}} \tag{5}$$

Where x is the dependent variable, and p is the function of x that estimates the probability. The regression coefficients β, shown in the Odds Ratio column in each table indicate the strength of each economic indicator in predicting the dependent variable. These coefficients are a calculation of the odds ratio:

$$\beta = \left(\frac{p(\text{conflict}}{1 - p(\text{conflict})} \right) \tag{6}$$

Three different analyses were performed: the first on all guilds over the full time period, the second only on dyads that include one big guild (defined for purposes of this analysis as guilds with more than 30 members), also over the full time period, and the third on data gathered only during the periods of stability between a period of two large wars, which took place in months 17, 18, and 19. These three different analyses were performed in part to examine whether big guilds behave differently than small guilds, and whether economic predictors behave differently during peacetime and wartime.

Analysis 1: All Guilds, Full Time Period

The dataset for this analysis included all dyads across the entire time period. The three economic indicator variables are not actually independent of each other, since all are a function of dyadic trade share. Fortunately, Barbieri does separate regressions against each variable in addition to her Full Model, which includes all three variables together. We replicated her methodology, using four different models, which we will designate M1, M2, M3, and M4. M1 regresses against only against salience as the economic predictor variable, M2 only against symmetry, M3 against interdependence, and M4 against all the economic predictor variables together (equivalent to the Full Model). The model results can be seen in Figure 17.1 for all guilds.

These results appear to indicate that the symmetry variable is correlated strongly and positively with likelihood of conflict. It is possible that only symmetry registers as statistically significant because it is the only indicator that is not a vanishingly small number, as are salience and interdependence. Simply for purposes of comparison directly to the academic literature, the results of the M4 analysis are shown side by side with the Full Model results from [1] in Table 17.3. As mentioned above, the variables salience, symmetry, and interdependence are not independent of each other, and the M4 results are given here simply for purposes of comparison. Barbieri's work shows a negative correlation between both those variables and likelihood of combat, while our analysis shows a positive correlation. On the other hand, Barbieri shows a positive correlation between interdependence and conflict, (ours is negative), where interdependence is simply a product of the two other economic variables. It is likely that the non-independence of the economic variables is confounding the statistical analysis and is complicating comparison of our results.

Predictor and Control Variables	*Models*											
	M1: Salience			M2: Symmetry			M3: Interdependence			M4: All		
	Odds Ratio	*SE*	*p*	*Odds Ratio*	*SE*	*p*	*Odds Ratio*	*SE*	*p*	*Odds Ratio*	*SE*	*p*
Fixed Parts												
(Intercept)	0.15	0.06	<.001	0.00	0.00	<.001	0.15	0.06	<.001	0.00	0.00	<.001
Salience	0.00	0.00	.005							Inf	Inf	<.001
Symmetry				56887.24	161762.56	<.001				8933765.75	38675407.04	<.001
Interdependence							0.00	0.00	<.001	0.00	0.00	<.001
Economic Strength Ratio	1.51	0.64	.339	1.44	0.62	.391	1.51	0.65	.337	1.42	0.62	.420
Combat Strength Ratio	0.76	0.13	.110	0.76	0.13	.110	0.76	0.13	.110	0.76	0.13	.110
Size Ratio	0.54	0.05	<.001	0.52	0.05	<.001	0.54	0.05	<.001	0.52	0.05	<.001
Contiguity (IS Contiguous)	0.24	0.01	<.001	0.24	0.01	<.001	0.24	0.01	<.001	0.24	0.01	<.001
Alliance (IS Foe)	12.83	0.68	<.001	12.86	0.68	<.001	12.82	0.68	<.001	12.76	0.65	<.001

Figure 17.1: Model results for All Guilds, full time period. Regression coefficients expressed as the Odds Ratio, SE is the standard error, and p is the significance value estimated with Wald's Z using the sjt.lmer package in R

Analysis 2: Large Guilds, Full Time Period

The dataset for the second analysis included only large guilds (guilds with 30 or more members) across the entire time period. The results of all four models run on this dataset are shown in Figure 17.2.

Models M1-M3 show no significant effect of the economic indicators on conflict likelihood. There is, however, a positive correlation between strength and size ratios and combat likelihood (combat is more likely if there is a power asymmetry).

Large Guilds, Inter-War Period

The dataset for analysis 3 included only dyads with at least one large guild, and was restricted to the three months between large wars. The results of

Predictor and Control Variables	*Models*											
	M1: Salience			M2: Symmetry			M3: Interdependence			M4: All		
	Odds Ratio	*SE*	*p*	*Odds Ratio*	*SE*	*p*	*Odds Ratio*	*SE*	*p*	*Odds Ratio*	*SE*	*p*
Fixed Parts												
(Intercept)	1.02	0.84	.983	Inf	Inf	.254	1.02	0.84	.983	Inf	Inf	.071
Salience	0.00	0.00	.040							Inf	Inf	.004
Symmetry				0.00	0.00	.255				0.00	0.00	.071
Interdependence							0.00	0.00	.214	0.00	0.00	.004
Economic Strength Ratio	0.10	0.09	.012	0.10	0.09	.012	0.10	0.09	.012	0.11	0.10	.017
Combat Strength Ratio	0.98	0.33	.961	1.00	0.34	.992	0.98	0.33	.960	1.03	0.35	.941
Size Ratio	0.95	0.19	.785	0.95	0.20	.807	0.95	0.19	.785	0.97	0.20	.896
Contiguity (IS Contiguous)	0.43	0.03	<.001	0.42	0.03	<.001	0.43	0.03	<.001	0.44	0.03	<.001
Alliance (IS Foe)	4.68	0.44	<.001	4.73	0.44	<.001	4.68	0.44	<.001	4.61	0.42	<.001

Figure 17.2: Model results for Large Guilds, full time period. Regression coefficients expressed as the Odds Ratio, SE is the standard error, and p is the significance value estimated with Wald's Z using the sjt.lmer package in R

this analysis can be found in Figure 17.3. Here we see the same correlation between power asymmetries and conflict likelihoods that we saw in Analysis 2, and the same lack of significance in the economic variables.

Caveats These experimental findings must be interpreted in light of several caveats. One important note is that the definition of the economic factors was defined in such a way that many dyads had extreme values (either at the very low or very high end of the scale), and were all derived in some way from trade share (so they were non-independent predictors), both of which could have skewed the results. No interactions were included between any of the control variables and the economic predictors, so we cannot say whether, for example, Symmetry is a more important predictor for contiguous vs. non-contiguous guilds. Future work could address these questions in more detail.

Predictor and Control Variables	*Models*											
	M1: Salience			M2: Symmetry			M3: Interdependence			M4: All		
	Odds Ratio	*SE*	*p*	*Odds Ratio*	*SE*	*p*	*Odds Ratio*	*SE*	*p*	*Odds Ratio*	*SE*	*p*
Fixed Parts												
(Intercept)	0.00	0.00	<.001	0.00	0.00	.643	0.00	0.00	.471	0.00	0.00	<.001
Salience	0.00	0.00	.370							0.00	0.00	<.001
Symmetry				Inf	Inf	.649				Inf	Inf	<.001
Interdependence							0.00	0.00	.795	Inf	Inf	<.001
Economic Strength Ratio	0.08	0.00	<.001	0.12	1.37	.851	0.04	0.45	.784	0.14	0.00	<.001
Combat Strength Ratio	6.57	0.01	<.001	0.39	1.67	.826	12.78	57.87	.574	1.16	0.00	<.001
Size Ratio	3.08	0.00	<.001	1.69	3.49	.799	3.30	7.09	.578	1.45	0.00	<.001
Contiguity (IS Contiguous)	0.19	0.00	<.001	0.23	0.18	.053	0.20	0.16	.041	0.17	0.00	<.001
Alliance (IS Foe)	7.39	0.00	<.001	6.65	6.52	.053	7.68	8.28	.058	7.50	3.65	<.001

Figure 17.3: Model results for Large Guilds, Inter-War period. Regression coefficients expressed as the Odds Ratio, SE is the standard error, and p is the significance value estimated with Wald's Z using the sjt.lmer package in R

Operationalizing MMOG Data

There were several issues as we operationalized the real world variables into Game X .

In Game X the closest analogue to countries were guilds, which were player created and managed. However guilds have no physical boundaries and can vary greatly in size. Countries and states in the real world have strict boundaries. This made the operationalization of the "Contiguity" control variable difficult as it relied on geographic distance between countries.

The lack of publicly documented agreements had a significant effect on operationalizing the "Alliance" control variable. Since we primarily used the "Is Foe" variable, we were aggregating positive relationships that span multiple types, from merely neutral to strongly positive. In the real world treaties and other agreements provide a formal method to assessing the strength of alliances.

Many guilds in Game X have exhibit organizational properties, including roles and a hierarchy (Lakkaraju and Whetzel, 2013). However, there is no public process to codify this organization. In contrast, in the real world, constitutions and laws are drawn up and publicized. This allows observers and the general public to gain information about the organization and can be important for understanding conflict.(Barbieri, 1996) used the control variable of "'Joint Democracy" as a way of capturing the type of government of a country. We could not operationalize that in our current analyses.

Even if a guild establishes laws and codified them into a public document it is not clear if we can define what a "democracy" is, and whether real world definitions can apply.

MMOGs, at their core, are games that are meant for entertainment. An important part of entertainment is the ability to explore and make mistakes in an environment with little consequence. Clearly such behavior can be exhibited by players, especially early on in the game. We must account for this in our analysis. We suspect that as players stay in the game for longer they are more attached to their character and will act in a way to protect their character. We must be aware of this and sample the data to try to avoid the exploration phase of players behavior.

It is impossible to draw any real conclusions about the effects of economic ties and combat likelihood in the game, given the way that the economic variables are defined. However, the purpose and focus of this work was not to do so, but to demonstrate a proof-of-concept: that data from an online serious game can be operationalized in the form of economic, political, and military variables, and that statistical analysis can be performed on that data for direct comparison to academic studies of real world data. Through this exercise, we have identified some of the key issues with using game data for national security research, which we describe in more detail below. With a carefully designed game such as the proposed MMOG described above, we believe that we can populate a database with conflict data to use in future academic research. The work on Game X described here is an example of how MMOGs can be used as experimental platforms and contribute to research efforts in policy areas where data are otherwise sparse.

GAMES AS EXPERIMENTS: THE FUTURE OF RESEARCH

As discussed and demonstrated in the sections above, games as experiments have both a long history and a bright future. While games have historically been used to probe possibilities in a simulated space and recreate in detail history as it happened, the future of research with games may include gathering large amounts of data for analysis of multiple kinds; the traditional analysis of interesting trajectories that the game play took due to unexpected player actions, but also statistical and other quantitative analysis on the potentially massive amounts of data that can be gathered using a game with a large player base such as a MMOG. This type of game and the accompanying dataset will allow for large-scale analysis of actions taken over time and across multiple games; a type of analysis that has heretofore been extremely difficult to perform. The ability to operationalize game data that we demonstrated in above will allow researchers to compare gameplay results directly to those from the real world, and may enable them to augment real world datasets with MMOG data. Analysis such as the logistic analysis performed above can help inform researchers and policy-makers of the potential impact of particular trade, military, or diplomatic relationships on future outcomes, such as the likelihood for conflict that we tested in our analysis of Game X, which correlated positively with power asymmetries in our analysis.

As we consider the utility of online games for research, we must still be aware of the difference in intention: games are meant, fundamentally, to engage a player through entertainment. To maintain engagement, the design of games may cause issues when used for experimentation. We posit the following considerations when analyzing game data.

Simplification

Entities and processes in the game world are simpler than in the real world. This is for multiple reasons. First, it is to reduce the cognitive burden of learning the game. Second, it is to provide an environment that is focused on the core purpose of the game. The simplification makes the mapping between real world entities (our actual target of interest) to game world entities difficult, as we saw in the mapping of guild to nation-states above.

Option abundance

Game choices (while certainly simplified from the real world) are often provided in plenty. To maximize engagement, players' need to be able to explore and discover new things. However, in experimental contexts one often wants to limit the number of options for a subject in order to study the underlying relationship better.

Event Shaping

The game may push the player to make certain choices or experience certain events. For instance, games often encourage conflict through resource manipulation. When considering the correlation between in-game behavior and real-world behavior one must be careful to account for the potential forces that are driving in-game behavior.

Another factor is that players may be focused on exploring the world initially, and the game may encourage that by providing simple initial environments.

FINAL DISCUSSION

MMOGs will never be perfect representations of the real world; no experimental laboratory is, nor are they intended to be. What both MMOGs and real-world laboratories offer is the possibility to test hypotheses in a controlled setting and to manipulate the controls in order to observe how differences in controls affect experimental results. The potential for MMOGs to serve as experimental platforms for various types of research is enormous given the large sample size that they provide for analysis and the ability to engineer the online environments to address specific research questions.

The exercise we detailed in this chapter, to operationalize variables from an existing MMOG, served a useful purpose in highlighting potential issues that can arise when studying game data, especially for national security issues. Understanding, and addressing, how simplification, option abundance, and event shaping can influence data analysis and interpretation of results is important future work.

AUTHOR BIOGRAPHY

Dr. Kiran Lakkaraju is a Senior Member of the Technical Staff at Sandia National Laboratories, New Mexico in the Cognitive Science and Applications group. He has more than a decade of experience developing models of human behavior and social systems and a has background in artificial intelligence, multi-agents systems and computational social science. He holds a M.S. and Ph.D. in Computer Science from the University of Illinois at Urbana-Champaign. His research has been marked by extensive interdisciplinary efforts that bring together the social and computational sciences. His work has revolved around the question of social influence, whether in linguistic behavior, information diffusion, attitude change, or behavior diffusion. More recently he has focused on developing models that explore the link between social structure (social networks, roles/hierarchy) with cognitive structure (how concepts are interrelated, cognitive consistency, confirmation bias) with respect to problems of information dissemination and attitude change. He is currently building an online social experimentation platform (the Controlled, Large, Online Social Experimentation platform).

Dr. Laura Epifanovskaya has a strong background in nuclear policy and conflict research. She received a doctorate from the University of Southern California in Physical Chemistry in 2010 and went on to work as a nuclear policy intern at the American Association for the Advancement of Science's Center for Science, Technology, and Security Policy in Washington DC. She joined Sandia National Laboratories in 2011 as a Systems Analyst with a focus on nuclear policy before transitioning to a weapons systems engineering role in 2014. In her current role, she develops technologies to inform reasoning about deterrence, including games as experiments.

Dr. Mallory Stites is an Experimental Cognitive Psychologist at Sandia National Laboratories. Her primary research interests lie in understanding the processes involved in skilled reading and visual information processing, using complementary techniques such as eye-tracking and event-related brain potentials (ERPs) to find converging evidence for underlying comprehension mechanisms. The common thread that unites her diverse research areas is to understand how the way visual information is presented to users, and the online processes users engage when viewing it, impact their ability to effectively understand and act on the information.

Dr. Joshua Letchford received his Ph.D. in Computer science from Duke University. His research areas of interest are Game Theory, Optimization, and Machine Learning with a focus on understanding how adversarial behavior can be accurately incorporated in models. He has spent his past five years as a Research Scientist at Sandia National Laboratories applying these techniques to problems ranging from cybersecurity and infrastructure resilience to modeling residential solar adoption and nation-state conflict.

Jon Whetzel is a Senior Member of the Technical Staff at Sandia National Laboratories within the Interactive Simulation, Systems, and Analysis organization. Jon received a BS and MS in Computer Science from Texas A&M University and has several years of experience as a software engineer at Sandia, specializing in artificial intelligence and serious game development. Jon is also an Adjunct Professor at the University of New Mexico (UNM), where he teaches courses on video game development and acts as a game development liaison between Sandia and UNM.

Dr. Jason Reinhardt is a national security systems analyst with Sandia National Laboratories specializing in nuclear and cross-domain deterrence and quantitative risk studies. His research focuses on the development of quantitative risk models to examine the trade-offs faced by nuclear armed nations in escalatory deterrence challenges. Other research interests include game theoretic applications to risk analysis and management, as well as adversary models. He joined Sandia National Laboratories in August of 2002, and has worked on a diverse set of projects both as an engineer and as an analyst.

ACKNOWLEDGEMENTS

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.

REFERENCES

(2012). MMOG Data. Technical report.

Barbieri, K. (1996). Economic Interdependence: A Path to Peace or a Source of Interstate Conflict? *Journal of Peace Research*, 33(1):29–49.

Castronova, E. (2008). A Test of the Law of Demand in a Virtual World: Exploring the Petri Dish Approach to Social Science. SSRN Scholarly Paper ID 1173642, Social Science Research Network, Rochester, NY.

De Lope, R. P. and Medina-Medina, N. (2017). A Comprehensive Taxonomy for Serious Games. *Journal of Educational Computing Research*, 55(5):629–672.

Deudney, D. (1983). *Whole Earth Security: A Geopolitics of Peace*. Worldwatch Paper Series. Worldwatch Institute.

Gosling, S. D., Sandy, C. J., John, O. P., and Potter, J. (2010). Wired but not WEIRD: the promise of the Internet in reaching more diverse samples. *The Behavioral and Brain Sciences*, 33(2-3):94–95.

Guarino, S., Eusebi, L., Bracken, B., and Jenkins, M. (2018). Using sociocultural data from online gaming and game communities. In Davis, P. K., O'Mahony, A., and Pfautz, J., editors, *Social-Behavioral Modeling for Complex Systems*. Wiley & Sons.

Henrich, J., Heine, S. J., and Norenzayan, A. (2010). The weirdest people in the world? *The Behavioral and Brain Sciences*, 33(2-3):61–83; discussion 83–135.

Laamarti, F., Eid, M., and El Saddik, A. (2014). An Overview of Serious Games.

Lakkaraju, K. and Whetzel, J. (2013). Group Roles in Massively Multiplayer Online Games. In *Proceedings of the Workshop on Collaborative Online Organizations at the 14th International Conference on Autonomous Agents and Multiagents Systems*.

Lu, L., Shen, C., and Williams, D. (2014). Friending your way up the ladder: Connecting massive multiplayer online game behaviors with offline leadership. *Computers in Human Behavior*, 35:54–60.

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*, 107(01):188–206.

Sabin, P. (2014). *Simulating War: Studying Conflict through Simulation Games*. Bloomsbury Academic, London Oxford New York New Delhi Sydney, reprint edition edition.

Williams, D. (2010). The Mapping Principle, and a Research Framework for Virtual Worlds. *Communication Theory*, 20(4):451–470.

Yee, N. (2006). The Demographics, Motivations, and Derived Experiences of Users of Massively Multi-User Online Graphical Environments. *Presence: Teleoperators and Virtual Environments*, 15(3):309–329.

Count of dyads engaging in conflict and/or trade in each month period.			
	Conflict Present		Conflict Absent
Month ID	Trade Present	Trade Absent	Trade Present
1	3	2	11
2	73	66	398
3	120	97	786
4	145	142	1016
5	123	124	1271
6	104	160	1421
7	149	118	1566
8	130	162	1589
9	124	133	1731
10	153	230	1817
11	133	172	1772
12	153	186	1951
13	159	195	1981
14	147	193	1977
15	209	705	2000
16	155	398	1944
17	127	203	1971
18	177	155	2013
19	147	138	2061
20	218	433	2137
21	183	494	1971
22	182	255	2213
23	199	189	2249
24	195	129	2083
25	108	77	1558

Table 17.2: Game X Conflict Data Summary

	Game X			**Barbieri**		
Variable	β	*se*	p	β	*se*	p
Salience	463.27	131.95	$\leq$ 0.001	-22.64	6.69	$\leq$ 0.01
Symmetry	16.01	4.33	$\leq$0.001	-4.46	0.80	$\leq$ 0.01
Interdependence	-490.88	140.78	$\leq$ 0.001	26.60	7.28	$\leq$ 0.01

Table 17.3: Game X M4 analysis compared with Barbieri's Full Model results. β is regression coefficient from the logistic regression, expressed as an odds ratio, *se* is the standard error, and p is the significance value estimated with Wald's Z using the sjt.lmer package in R