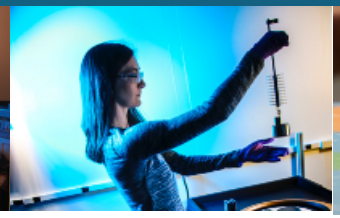
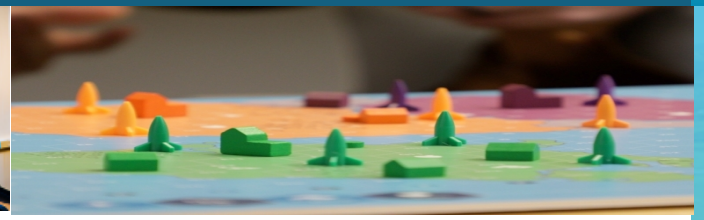
SAND2023-13015C

PEGASIS
PROGRAM for EXPERIMENTAL GAMING & ANALYSIS of STRATEGIC INTERACTION SCENARIOS

Sandia National Laboratories

Experimental Wargaming at Sandia National Laboratories





PRESENTED BY

Jason Reinhardt, Kiran Lakkaraju, Mika Armenta

U.S. DEPARTMENT OF ENERGY NNSA

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.

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.

The Complexity Scarcity Gap

We want to develop better theory and insights in order to inform policy- and decision-making. But, most relevant **conflicts and deterrence scenarios exist in a gap** where models can be overly complex and (thankfully) there is little to no data to study.

Simple Strategic Models

- Two (or few) Actors
- Existential Stakes
- Binary Thresholds
- Assumption of “Lesser and Included” Cases

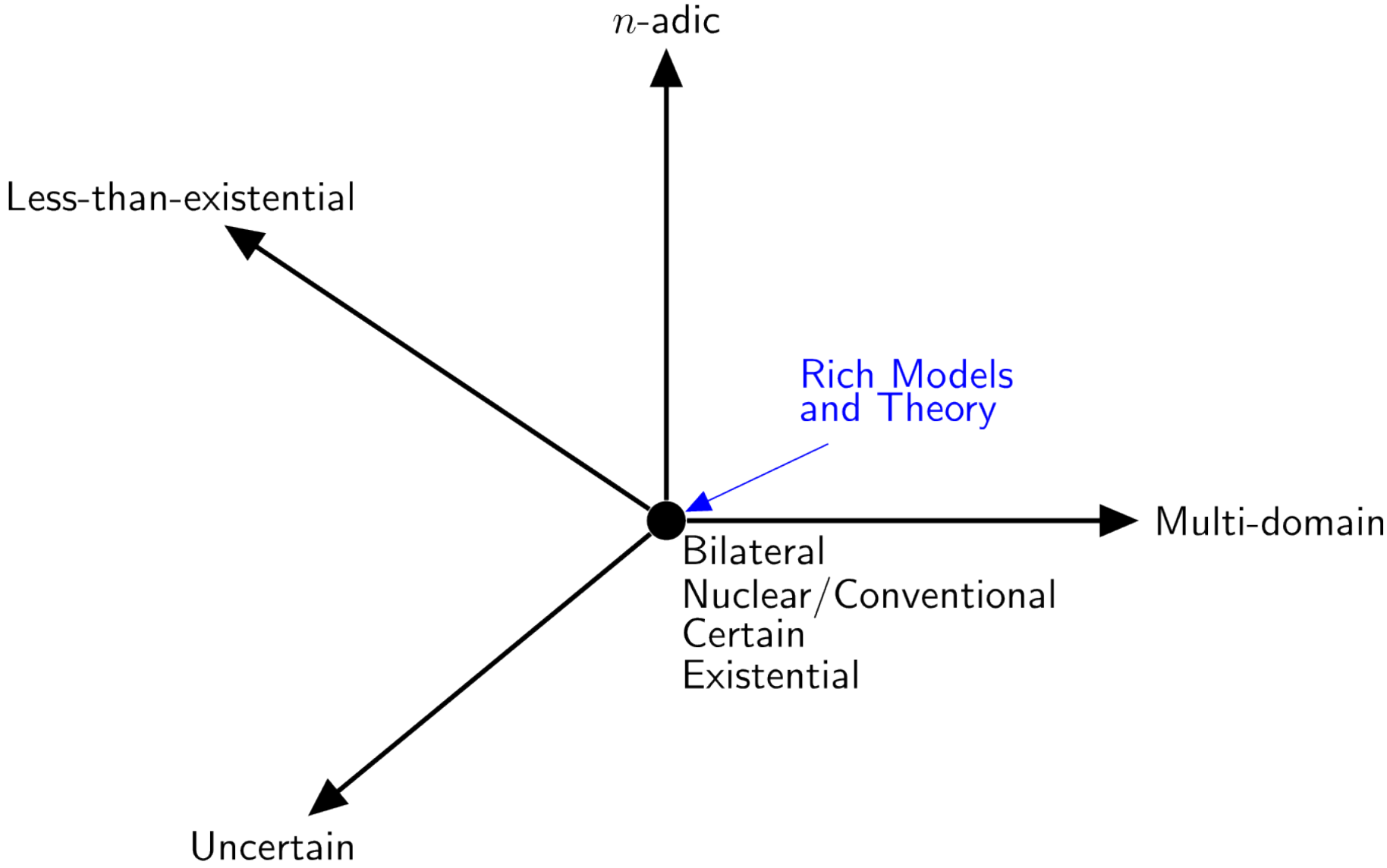
Bigger Models/ Better Algorithms

Rich Empirical Data

- Many Relevant Historical Instances
- Available Data Sources
- Proxy Experimentation
- Crowds vs. Individuals

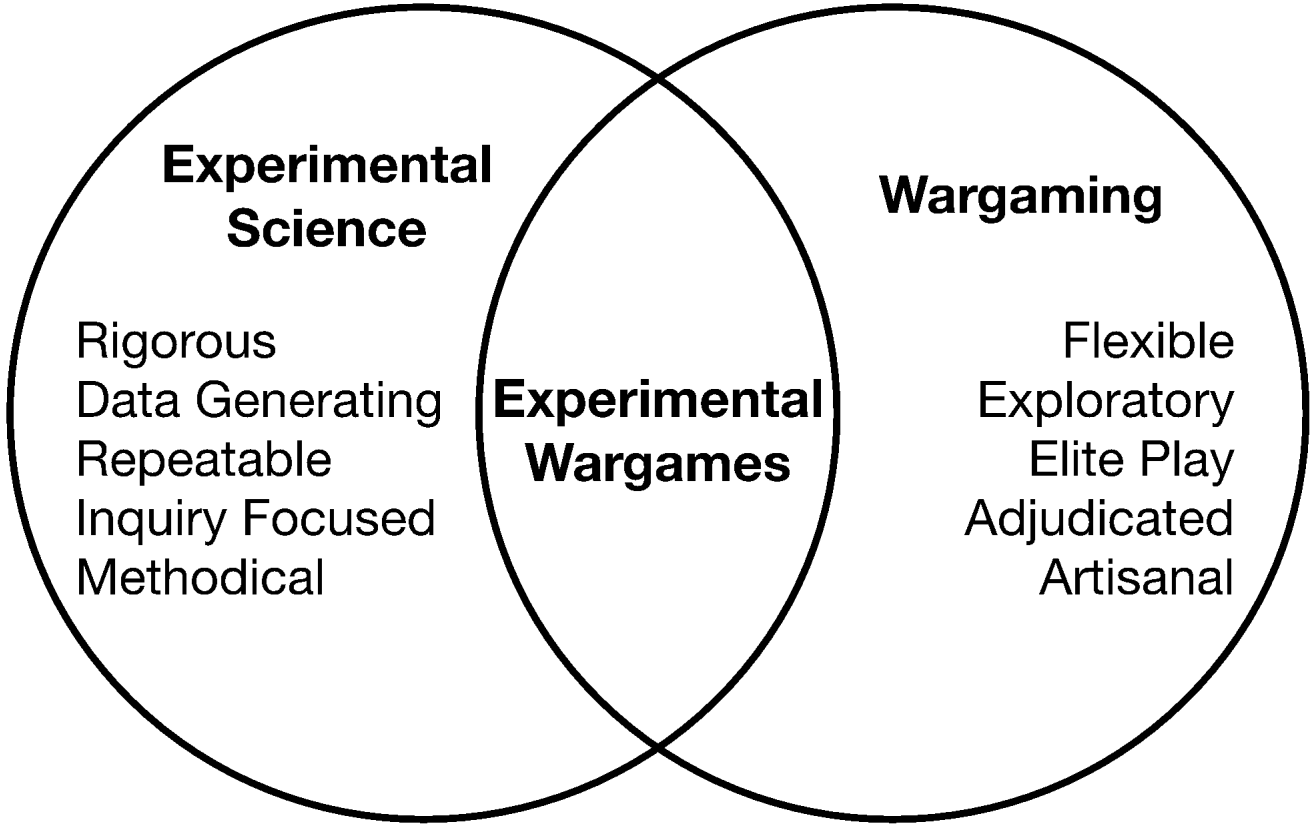
Innovative Data Set Creation and Curation

The Gap means our understanding covers only small part of the scenario space we need to cover.



Experimental Wargaming can help fill the gap with synthetic data

Experimental Wargames are games designed to **quantitatively study** national security scenarios of interest where the situation, potential responses, and abstraction are driven by research question(s) of interest.



Experiments are a knowledge-building hierarchy.

Experiments contribute to a **knowledge-building hierarchy**, where the results of one instance inform a progression, which leads to greater confidence that certain strategies are robust and can lead to more desirable scenario outcomes.

Design Study Experiment	Observation Experiment	Control Treatment
Goal: Identify potential variables and relationships that lead to outcomes of interest. Often have a desired objective/outcome for scenario Usually exploratory without well structured theory, identifying points of failure/success Can provide lessons learned and inform further development	**Goal:** Discover/confirm relationships between variables and assess strength. Independent variables are known, but outcomes are usually not Focused on developing possible relationships between factors Enables hypothesis and theory development	**Goal:** Determine if and how the state of the world can be consciously changed. Independent/Dependent variables are suspected/hypothesized Indicator of a causal relationship Provides decision- and policy-makers with stronger evidence of viability of specific courses of
E.g.: TTX/Workshop Wargame	E.g.: ASSENT	E.g.: SIGNAL

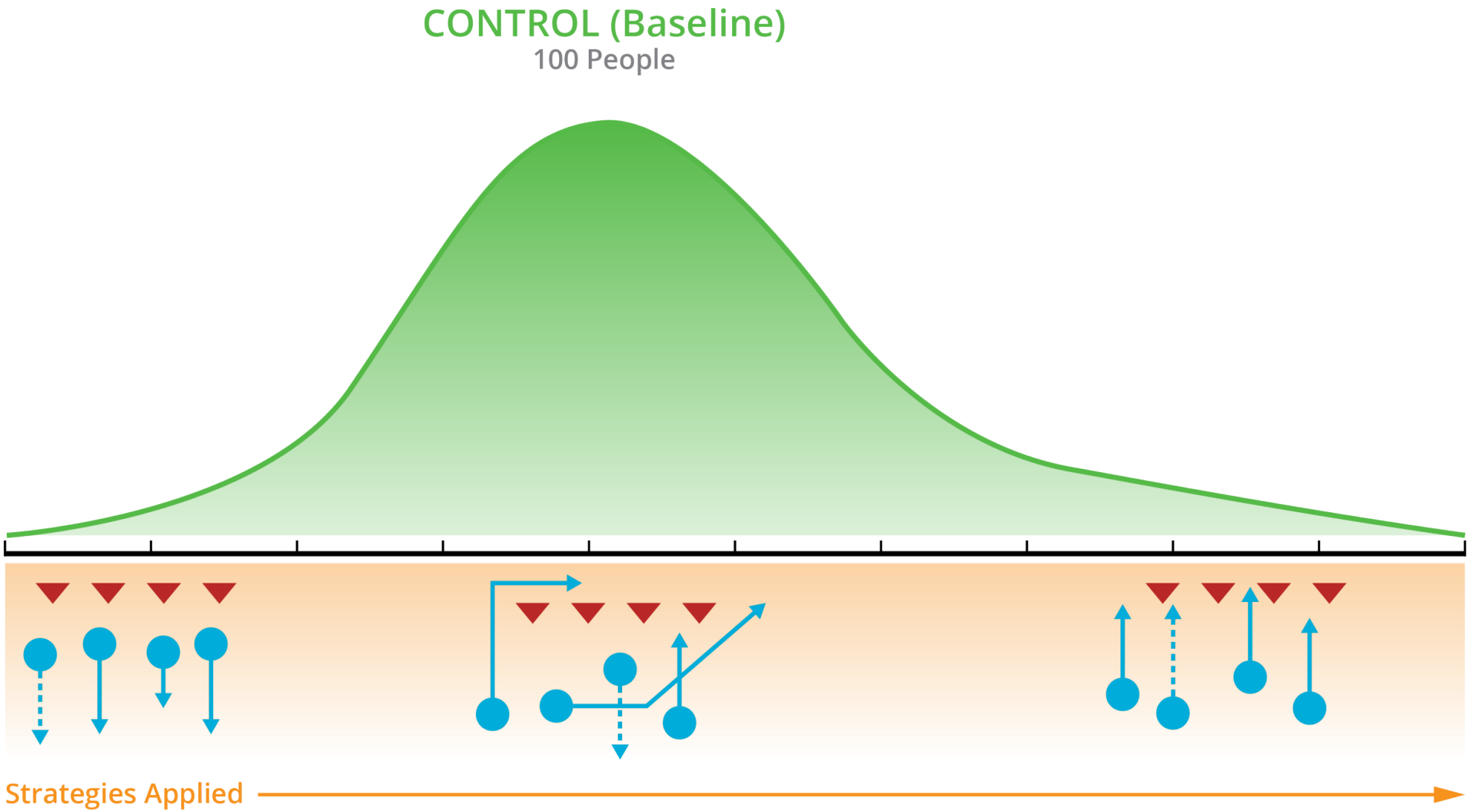
Experimental wargames are a complement to traditional wargames.

- Traditional wargames focus on insight generation and exploration of strategies for players.
  - Seminar games.
  - Tactical games.
  - Matrix games.
- Experimental wargames focus on capturing large-n data sets for analysis.
  - Identifying possible causal relationships
  - Proving support for hypotheses
  - Bracketing scenario variables
- There should be a virtuous cycle between traditional wargame approaches and experimental wargames

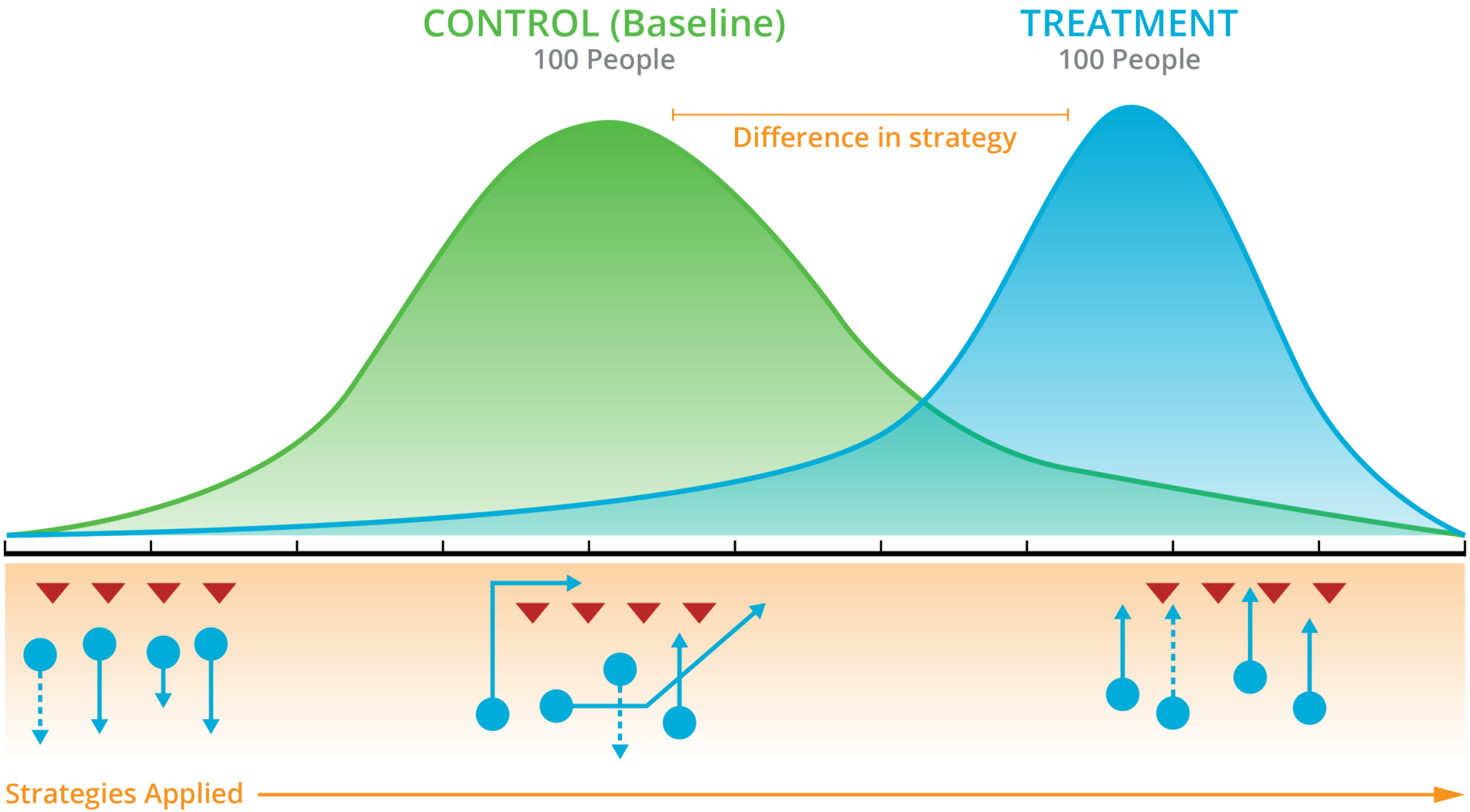
Development, C., and D. Centre. 2017. "Wargaming Handbook". Technical report, Ministry of Defence - The Development, Concepts and Doctrine Centre.

https://warontherocks.com/2019/08/wargaming-has-a-place-but-is-no-panacea-for-professional-military-education/

We want to quantitatively understand behavioral distributions...

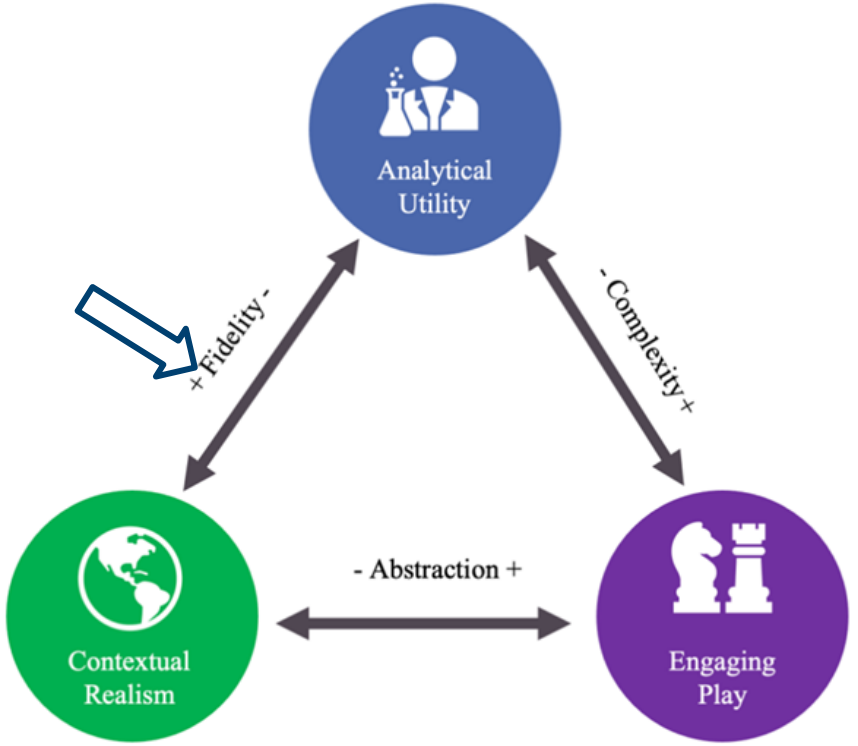


…and how those distributions change in controlled conditions.



Experimental wargames are informed by experimental principles.

- **Replicability**
  - To collect data from a large numbers of participants, we need to have a repeatable experiment.
- **Randomization**
  - Participants must be assigned to random conditions.
- **Controllability**
  - The game must be controllable to allow for systematic manipulation of the independent variable.
- **Instrumentation**
  - The game must allow for capturing player behaviors.
- **Neutrality**
  - The game should not bias participant behavior with regards to the research question.



Among the experimental wargaming trilemma, application of experimental principles and design is particularly useful for achieving fidelity

Reddie, A., Booth, R.E., Goldblum, B.L., Lakkaraju, K., Reinhardt, J (Feb. 2023). Wargames as Data: Addressing the Wargamer's Trilemma. *General*

Experimental Wargaming Sandia National Labs

Pre-10/2017 **October** **March 2020** **Present**

- Serious Games Focused on Homeland Security Applications
- Human Subjects and Computation Research
- Exploring Application to Deterrence Problems

Berkeley UNIVERSITY OF CALIFORNIA

Lawrence Livermore National Laboratory

PROJECT ON NUCLEAR GAMING

PEGASIS

PROGRAM for EXPERIMENTAL GAMING & ANALYSIS of STRATEGIC INTERACTION SCENARIOS

SIGNAL

TRACING HOUSE

CASTLE

GRAND GAMING RESEARCH for ALLIANCE NETWORK DYNAMICS

TANTALVS

ASSENT

The Project on Nuclear Gaming had many goals...

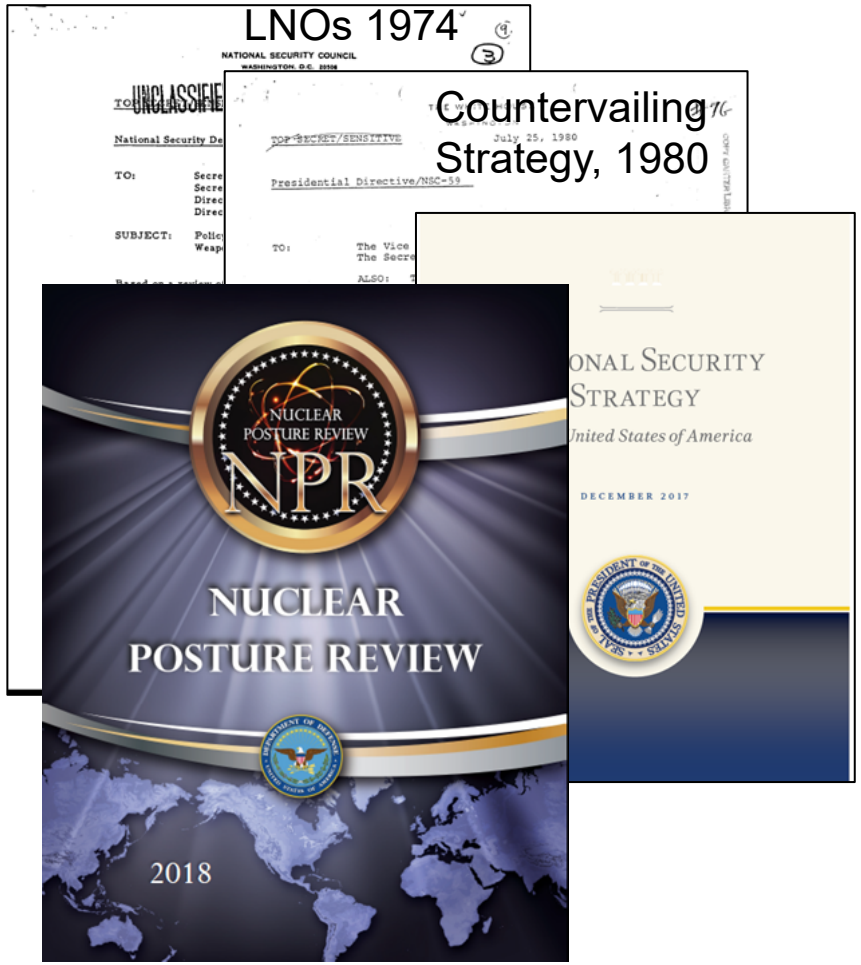
Research Questions:

- How can experimental games be constructed and executed to place players in abstract and hypothetical situations to model escalation challenges, including threats of nuclear use?
- What impact might different weapon capabilities have on deterrence and strategic stability?

PROJECT ON NUCLEAR GAMING

SIGNAL

PoNG is NOT making an assessment of any specific national policy or conflict scenario. It is the intent of PoNG to generate data that can help inform academic debates.



The Project on Nuclear Gaming's Experimental Approach

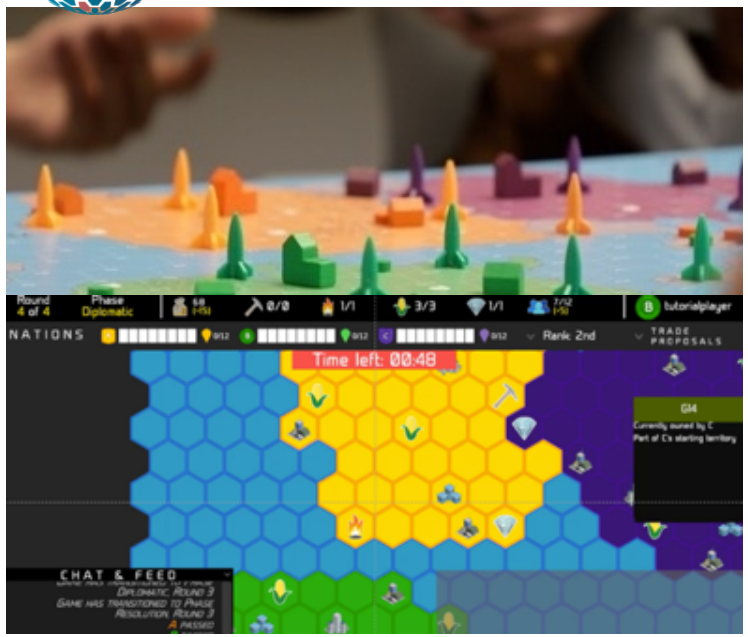
Experimental Games:

- **SIGNAL Board (Knowledgeable and Expert Players)**
  - Highly structured scenarios, rules based adjudication, fluid conversation and over-the-table player dynamics, improved quantitative data collection
- **SIGNAL Online (Any Player)**
  - Highly structured scenarios, rules based adjudication, more structured player dynamics, high quantitative data collection

Benchmarks:

- **SIGNAL TTX (Elite Players)**
  - Fluid exploration of scenario features, player concerns, and boundaries for outcomes, control team adjudication, qualitative and narrative data collection
- **Survey Experiments (Targeted Responders)**
  - Questionnaires focused on evaluating subject responses to specific

SIGNAL



SIGNAL varies initial conditions in weapon capabilities.

Vary military capabilities of players

- Tailored & Traditional NW
- Conventional Forces Only
- Cyber
- Defensive

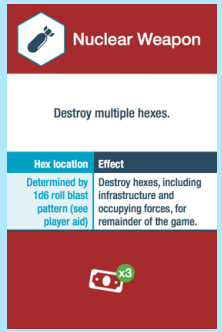
Vary existing economic/political ties

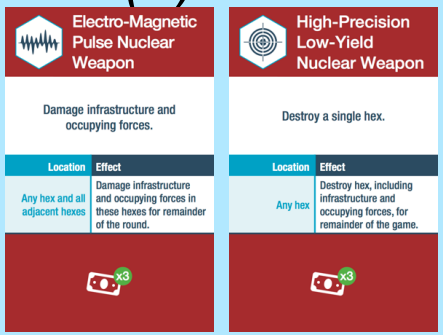
Execute of series of rounds, each with three phases:

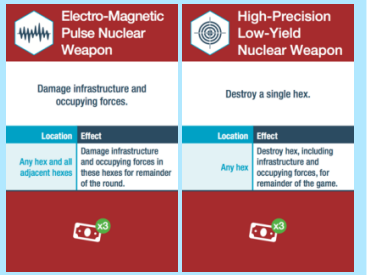
- ***Signaling Phase* for Diplomacy/Threats**
- *Action Phase* for Making Moves
- *Upkeep Phase* for Accounting of Results

SIGNAL Experimental Design

Player 1 Nuclear Capabilities	Player 2 Nuclear Capabilities: *HY (H)*	Player 2 Nuclear Capabilities: *+Tailored (T)*
High-Yield (H)	H, H (**Control**) *210 games*	H, T
+Tailored (T)	T, H	L, L (**Treatment**) *216 games*







Reddie, A. (2020)

Do weapon capabilities matter in the game?

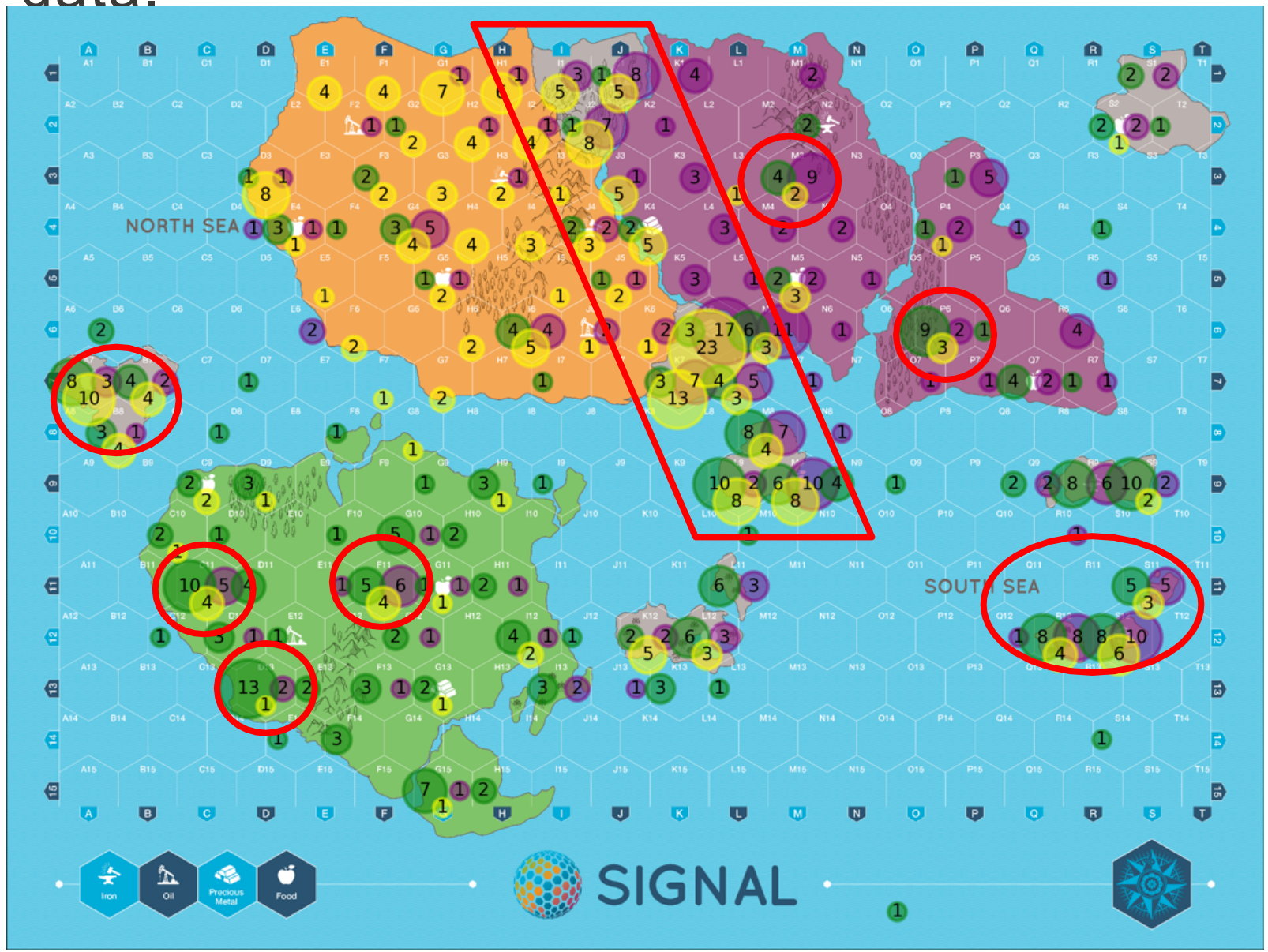
There is **no statistically significant difference** in nuclear use with tailored-output nuclear weapons in the arsenal within the SIGNAL environment.

- Slight positive trend for presence/substitutability of tailored effects, but statistically insignificant
- Most demographics also indicate slight trends, but statistically insignificant
- Results hold when eliminating last round from play
- Possibly explanations include:
  - Symmetric capabilities and peer states
  - At some level, presence/quantity of nuclear capabilities is a larger driver than qualities

Reddie, Andrew W., and Bethany L. Goldblum. "Evidence of the unthinkable: Experimental wargaming at the nuclear threshold." *Journal of Peace Research* (2022):

	Dependent variable:			
	Nuclear use			
	(1)	(2)	(3)	(4)
Treatment	0.039	0.038	0.024	0.022
	(0.218)	(0.218)	(0.219)	(0.219)
Female		−0.059		−0.055
		(0.120)		(0.130)
College Degree		0.004		−0.014
		(0.127)		(0.140)
Age >29		−0.073		−0.089
		(0.122)		(0.124)
National Security			0.076	0.097
			(0.143)	(0.150)
More Right			0.058	0.069
			(0.117)	(0.122)
Reported Knowledge			−0.062	−0.074
			(0.127)	(0.139)
Constant	0.963***	1.120***	0.915***	1.099***
	(0.154)	(0.308)	(0.256)	(0.396)
Observations	426	426	426	426
Log Likelihood	−249.421	−249.144	−249.027	−248.602
Note:				*p<0.1; **p<0.05; ***p<0.01

Process data from the game may be as valuable as outcome data.



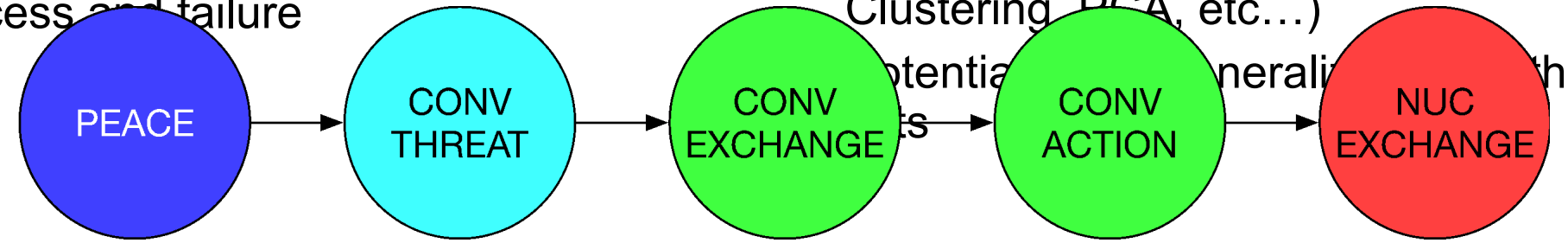
Overall strategic focus seems to be on:

- Contiguous borders
- Adjacent minor states
- Military and ‘Value’ targets

NOTE: While based on real data, these results are preliminary, non-conclusive, and for illustration only.

Process data can be broken down in to discrete steps or classes.

- Conflict Classes (More Flexible)
  - Categorizes rounds (made up of multiple moves by all actors) based on the combination of actions that were taken
  - Can be defined according to a wide range of analytic questions (e.g., how quickly did nuclear use occur?)
  - Tracks round-to-round conflict evolution and identifies macro trends
  - Can obscure fine detail of action-retaliation, dyad dynamics, deterrence success and failure
- Action-Dyad Simulation (More Powerful)
  - Simulates the game at a blow-by-blow level allowing for event detection and nuanced analysis
  - Can examine correlations between preceding actions, detect the attempt to deter, and the success/failure of deterrence attempts
  - Provides a coded set of specific event indicators that can be used in standard data science techniques (Random Forests, Clustering, PCA, etc…)

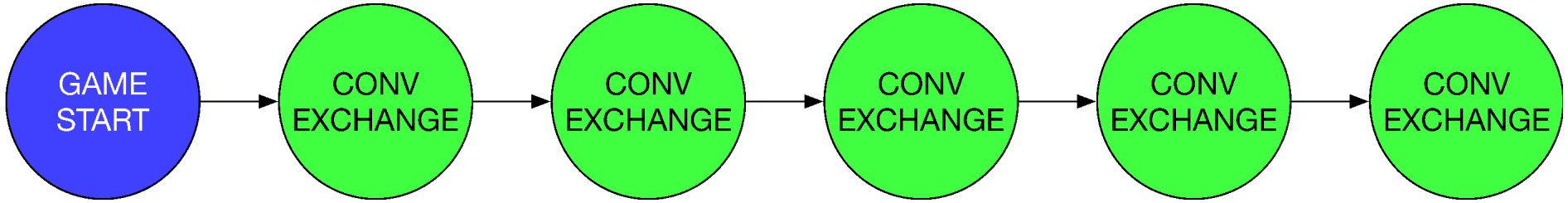


We can identify most likely escalation paths within SIGNAL.

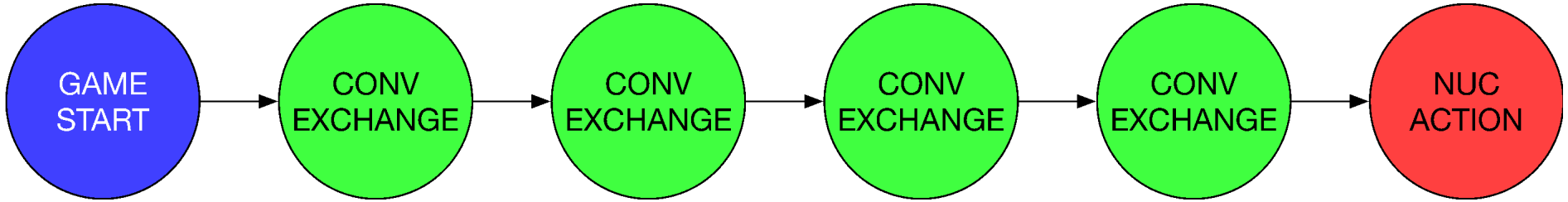
Coded data was rendered into transition probability matrices, conditioned on player actions, game state, round number, etc., which facilitated survival and next action/state analyses

Provides a foundation for probabilistic ML methods, such as Interactive Partially Observable Markov Decision Processes, which can be powerful tools for examining strategies and optimal conflict management policies.

Traditional Effect Games Most Likely Escalation Path

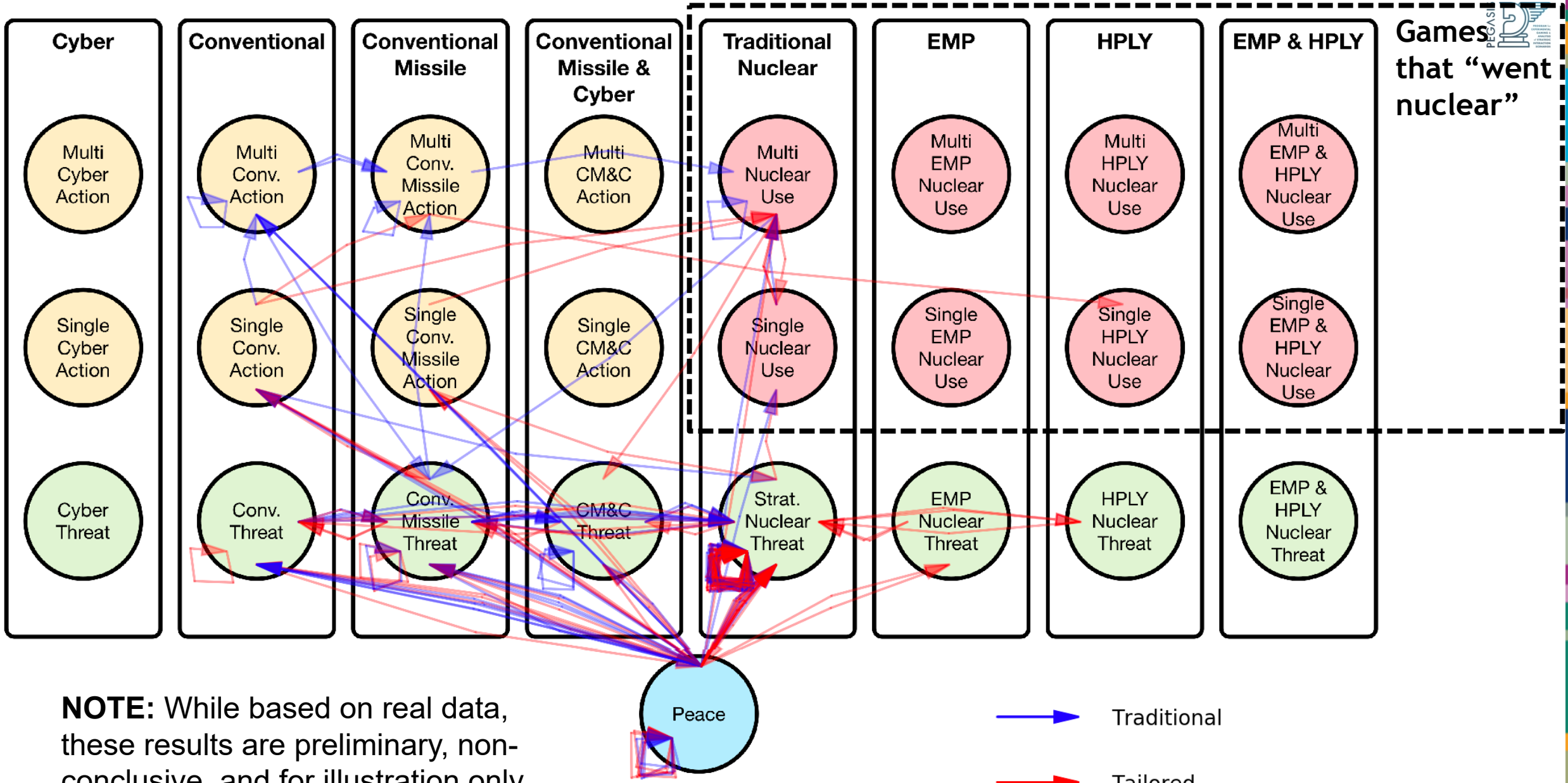


Tailored Effect Games Most Likely Escalation Path



NOTE: While based on real data, these results are preliminary, non-conclusive, and for

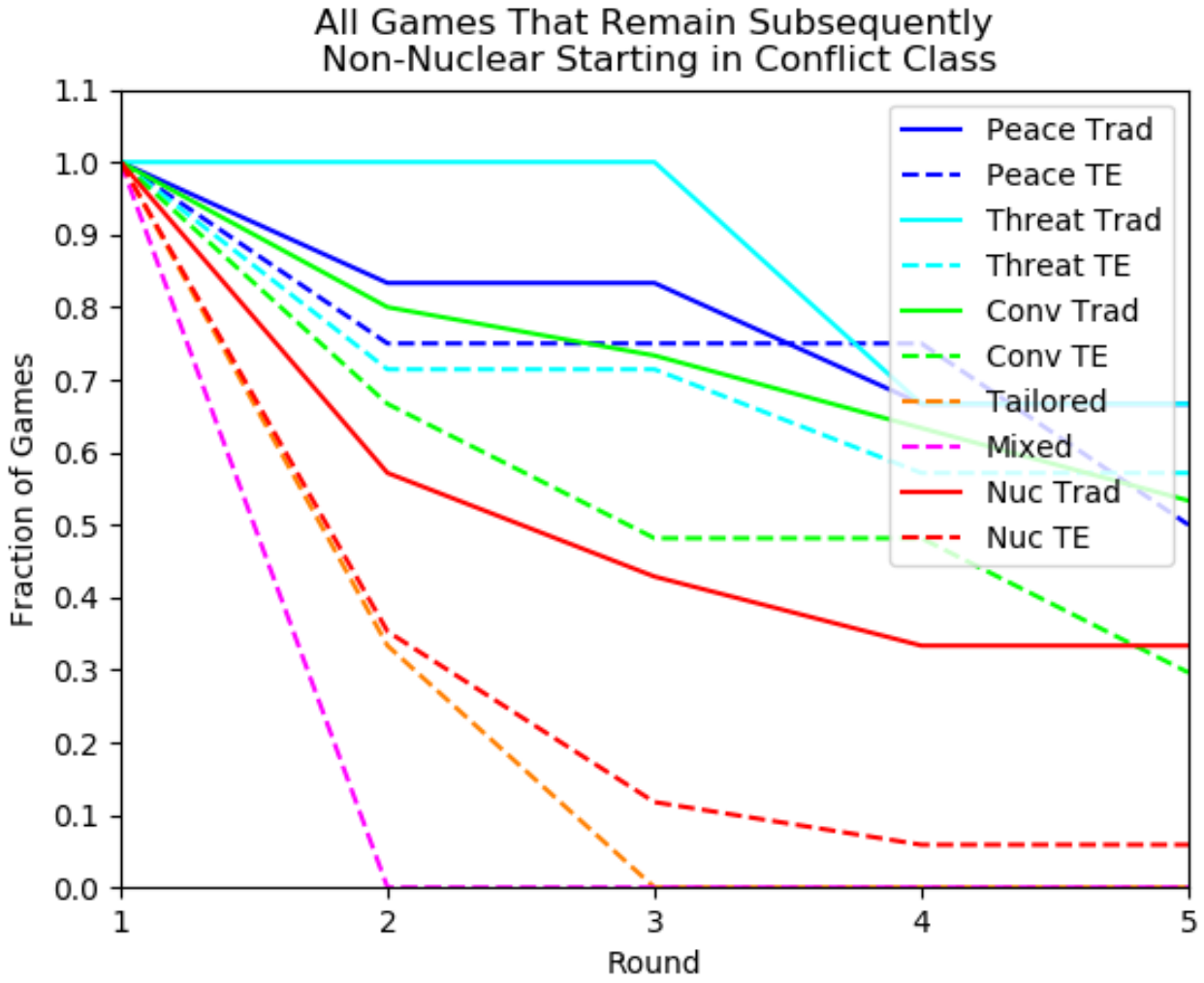
Conflict Class Analysis

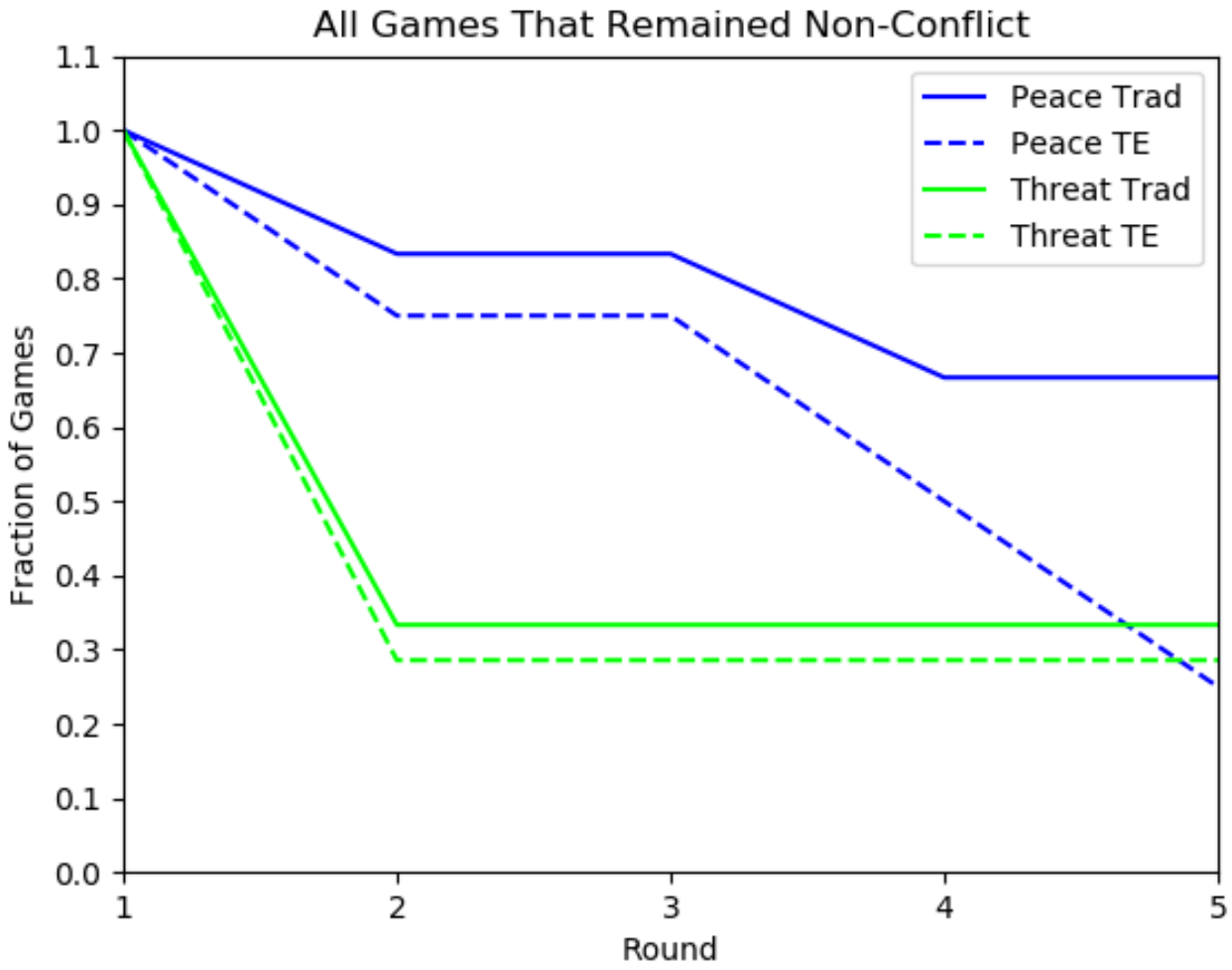


NOTE: While based on real data, these results are preliminary, non-conclusive, and for illustration only.

Conditional Survival Analysis

NOTE: While based on real data, these results are preliminary, non-conclusive, and for illustration only.





Games that started peacefully, tended to remain peaceful.
Games that started with nuclear use, tended to use nuclear weapons more.
Tailored effect weapons games, that included conflict, generally escalated faster.
Traditional effect weapons games, stayed non-nuclear longer.

PoNG achieved a lot in a short time.

- Hosted 10 events from California to London.
  - Includes a well attended event at Barksdale Air Force Base.
- High profile publications
  - Science.
  - Bulletin of Atomic Scientists
  - Journal of Peace Research
- Dozens of talks to Universities, Non-Profits, and National and International governmental organizations.
- 17 news articles spanning UNM press to Vice
- Awarded best student game at the 2019 Serious Games Showcase and Challenge (SGS&C)
- More than 1000 players from around the world who logged into our system
- Amassed the largest single data set for a wargame for analysis

Experimental Wargaming Sandia National Labs

Pre-10/2017 **October** **March 2020** **Present**

- Serious Games Focused on Homeland Security Applications
- Human Subjects and Computation Research
- Exploring Application to Deterrence Problems

Berkeley UNIVERSITY OF CALIFORNIA

Lawrence Livermore National Laboratory

PROJECT ON NUCLEAR GAMING

PEGASIS

PROGRAM for EXPERIMENTAL GAMING & ANALYSIS of STRATEGIC INTERACTION SCENARIOS

SIGNAL

TRACING HOUSE

CASTLE

GRAND GAMING RESEARCH for ALLIANCE NETWORK DYNAMICS

TANTALVS

ASSENT

Tracing House is an internally funded effort to better understand aspects of cyber deterrence using experimental wargaming.

Communication-capability tradeoff

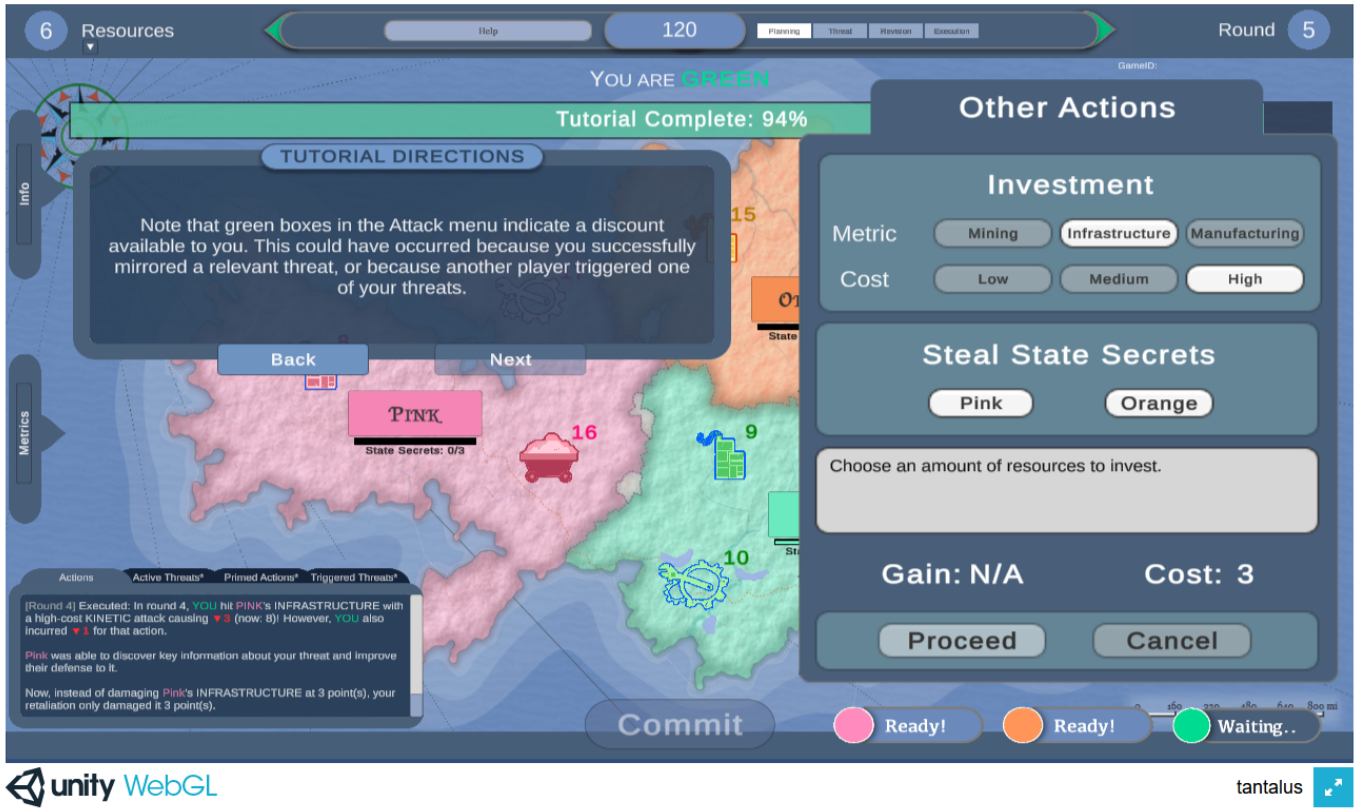
- Defenders need to communicate cyber threats to be effective.
- In the cyber domain, increased communication can decrease the effectiveness of the capability.

Research Question: How does the communication-capability tradeoff impact the techniques and strategies used by actors to manage conflict in a cyber context and the effectiveness of cyber threats?

Initial play testing is underway and will be deploying soon!

TANTALVS

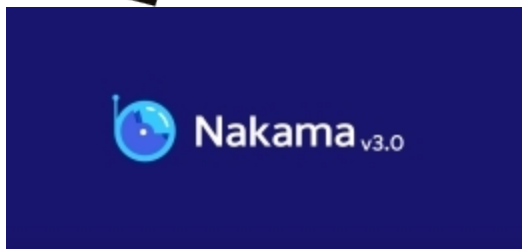
TRACING HOUSE



Tracing House is also developing CASTLE to make experimental wargaming research easier.

- Platform for building games to studying strategic interaction scenarios
- Software libraries built within Unity, designed for cloud deployment
- Allows us to:
  - Rapidly design and develop wargames
    - Focusing on round-structured games
  - Facilitate experimentation through scalable subject recruitment
    - Simplifies common player management tasks using the Nakama engine
  - Facilitate analysis
    - Cloud-based data collection backend
    - Analytic tools for standard tasks

CASTLE



Experimental Wargaming Sandia National Labs

Pre-10/2017 **October** **March 2020** **Present**

- Serious Games Focused on Homeland Security Applications
- Human Subjects and Computation Research
- Exploring Application to Deterrence Problems

Berkeley UNIVERSITY OF CALIFORNIA

Lawrence Livermore National Laboratory

PROJECT ON NUCLEAR GAMING

PEGASIS

PROGRAM for EXPERIMENTAL GAMING & ANALYSIS of STRATEGIC INTERACTION SCENARIOS

SIGNAL

TRACING HOUSE

CASTLE

GRAND GAMING RESEARCH for ALLIANCE NETWORK DYNAMICS

TANTALVS

ASSENT

Experimental Wargaming for Consensus Decision Making

The Gaming Research for Alliance Network Dynamics (GRAND) project is building a game to support NATO alliance decision makers.

ASSENT is a platform to practice key elements of Alliance decision making in crisis scenarios facilitate small, frequent, group interaction.

- Leverages SIGNAL platform, technology stack, and analysis tools
- Abstract design that distills key aspects of competition and cooperation
- Key element of Alliance consensus making protocols built into mechanics
  - Allows new staff to familiarize themselves with elements of Alliance consensus making protocols
- Configurable to multiple scenarios and varying number of players
- Online platform to allow players from across the world to participate from their locations

ASSENT

NATO Operations

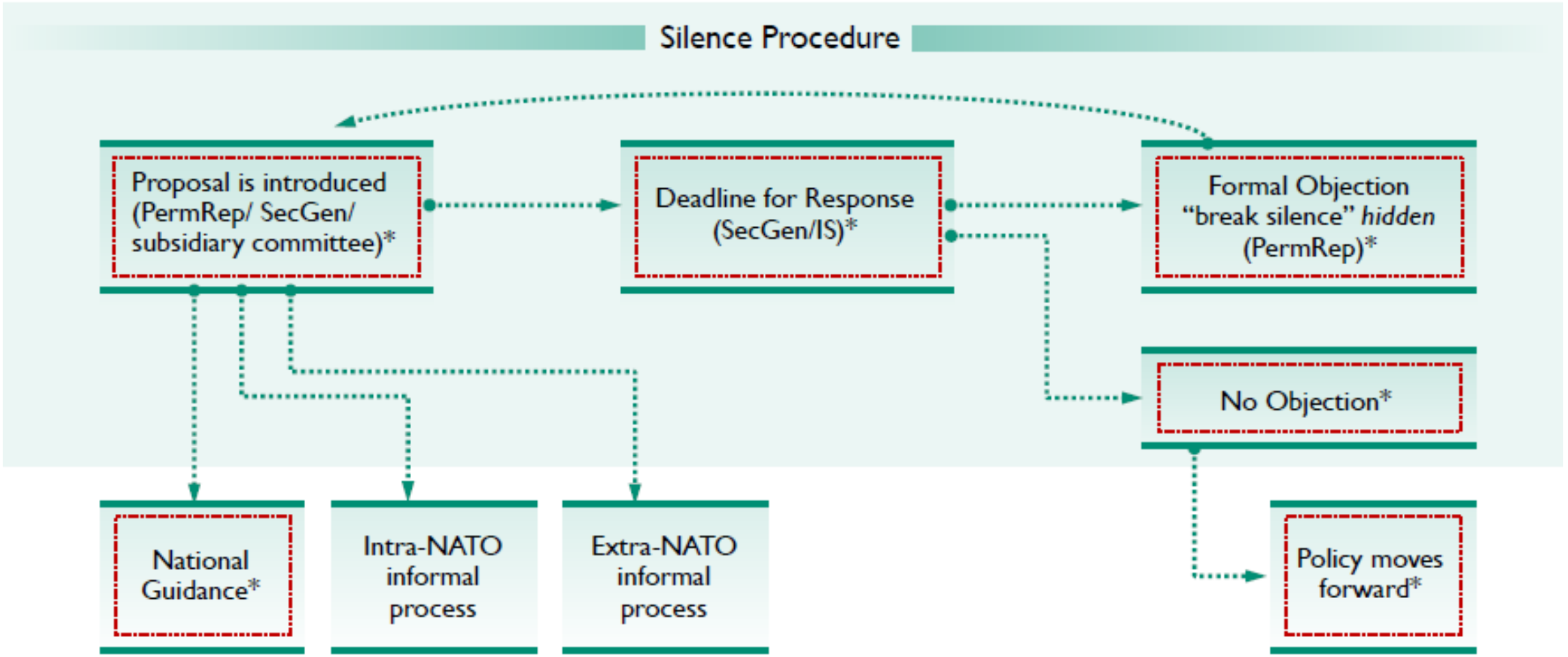
Preparations for Article 5 operations
- Integrated Air Defense Systems
- Air Policing
- e.g.…

Article 5 operations

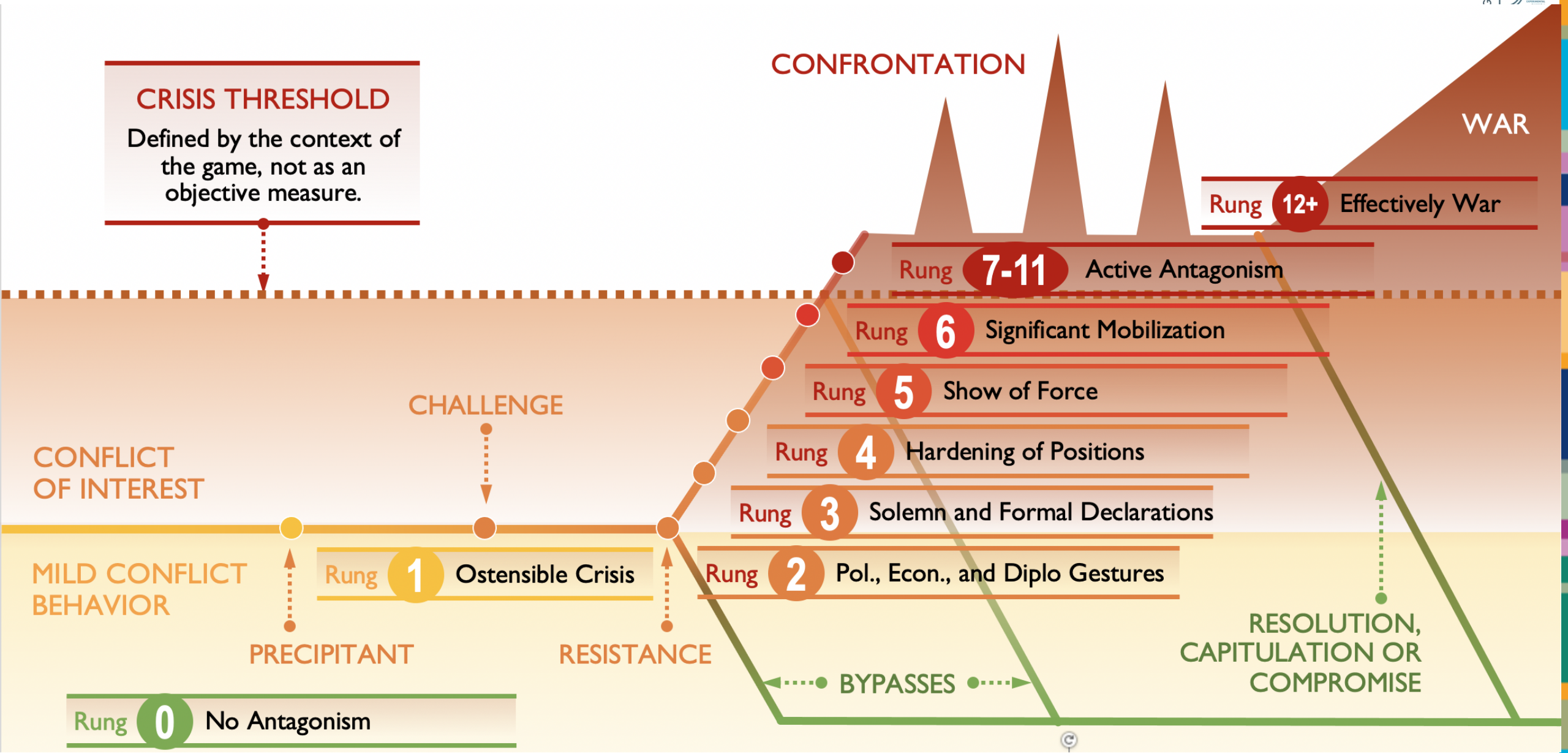
Article 4 operations
- Consultations

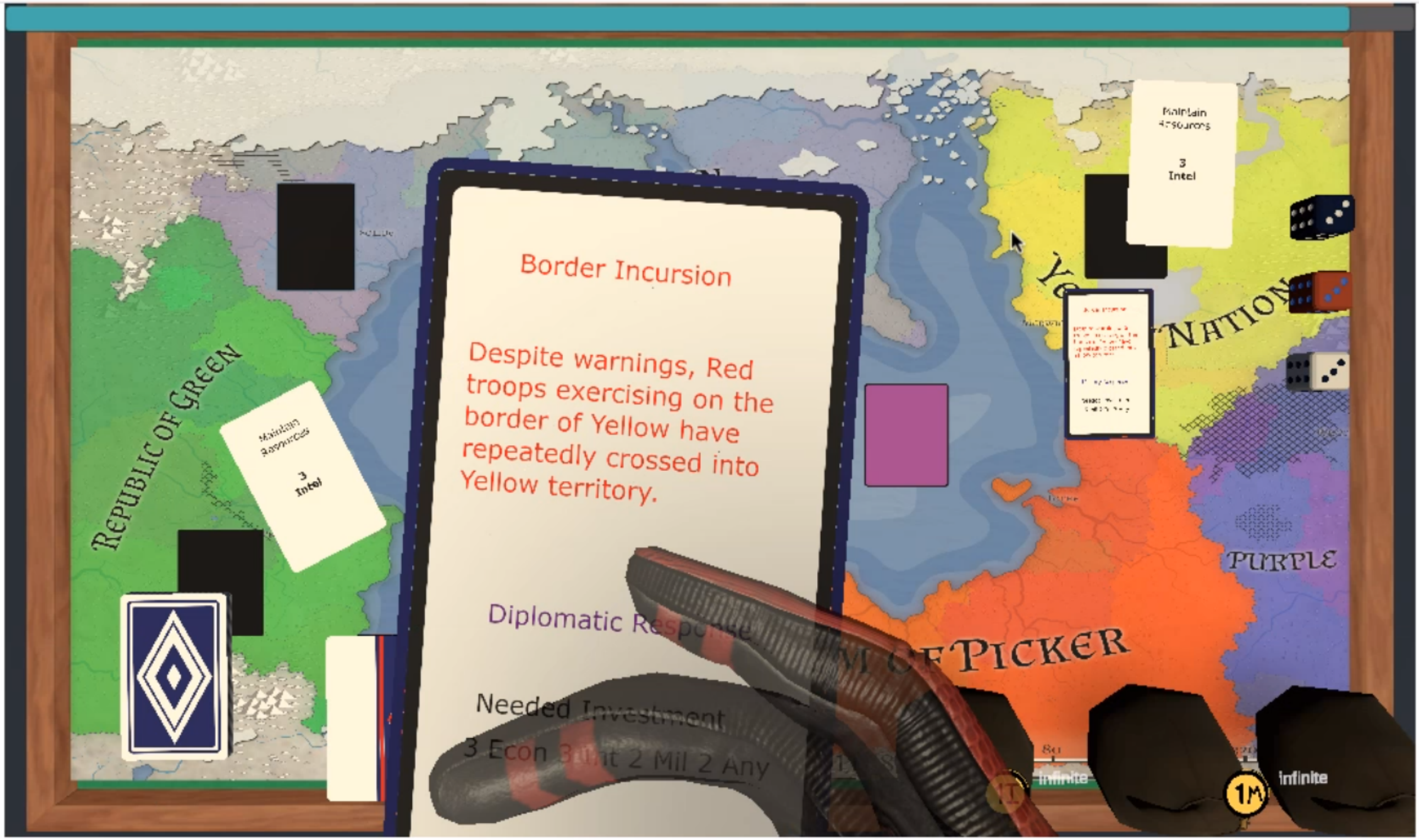
Non-Article 5 operations
- Eg. anti-piracy Horn of Africa/Med

We are embedding a model of the NATO Consensus Process



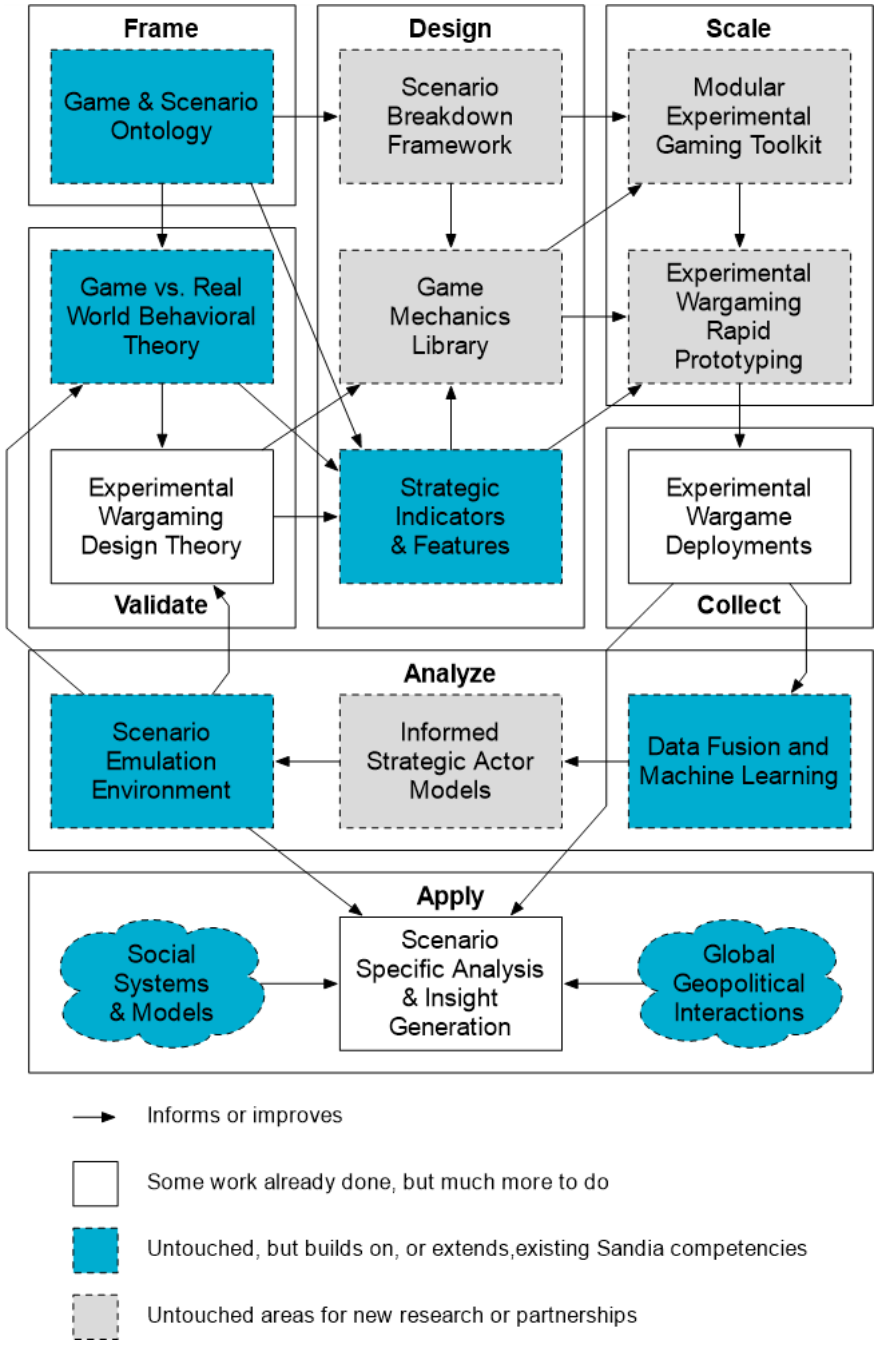
We are developing an ASSENT Analytic Model for Crises




Border Incursion
Despite warnings, Red troops exercising on the border of Yellow have repeatedly crossed into Yellow territory.
Diplomatic Response
Needed Investment
3 Econ 3 Mil 2 Mil 2 Any
Maintain Resources
3 Intel
Republic of Green
Purple
Picker
Infinite
Infinite

A roadmap for experimental wargaming

- Key questions:
  - How do you validate wargames and models that inform decision making?
  - How do you generalize from multiple “deep” scenarios to broad principles?
  - How do we transform wargame design from an artisan process to one that is more automated/streamlined?



Jason Reinhardt
jcreinh@sandia.gov

Kiran Lakkaraju
klakkar@sandia.gov

Mika Armenta
mlarmen@sandia.gov

Backup

PEGASIS
PROGRAM for EXPERIMENTAL GAMING & ANALYSIS of STRATEGIC INTERACTION SCENARIOS