SAND2021-10517PE

Tracing House Advisory Committee Meeting

Joshua Letchford

Ruby Booth, Gabriel Kelvin, Nick Blanchette

August 31, 2021

PEGASIS

PROGRAM for EXPERIMENTAL GAMING & ANALYSIS of STRATEGIC INTERACTION SCENARIOS

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.

Welcome!

Agenda

- 9:00am – 9:15am: Introduction and charge to advisory committee (Kiran Lakkaraju)
- 9:15am – 9:45am: Introduction to the project and experimental wargaming. (Kiran Lakkaraju)
- 9:45am – 10:30am: Research Design (Andrew Reddie)
- 10:30am – 10:45am: Discussion and Questions.
- 10:45am – 11:00am: Break
- 11:00am – 12:00pm: Game Design (Josh Letchford)
- 12:00pm – 12:30pm: Discussion

Feedback requested around three areas.

- **Question 1:** How well does our research question contribute to the academic discussion around (cyber) deterrence policy?
- **Question 2:** Do the proposed game mechanics of Tantalus address the key elements of the research question?
  - What game mechanics would you recommend be adapted or modified to better address the RQ, and how?
- **Question 3:** Is the level of abstraction for Tantalus appropriate for addressing the research question?
  - What changes would you recommend, if at all, to better balance fidelity and playability?

Context

- Tracing House is at the end of it’s first year (out of three).
  - Year 1:
    - Research design
      - Identifying the Independent Variables (IVs), Dependent Variables (DVs), and hypotheses.
    - Game design.
      - Define an early storyboard for the game.
  - Year 2 & 3:
    - Complete game design.
    - Implement online game.
    - Collect data.
    - Analysis.
- Requesting feedback at an early stage.

Three presenters for today

Kiran Lakkaraju
PI for Tracing House
Org. 8716

Andrew Reddie
Research Design
Org. 8716

Josh Letchford
Game Design
Org. 8762

Tracing House has a large and interdisciplinary team.

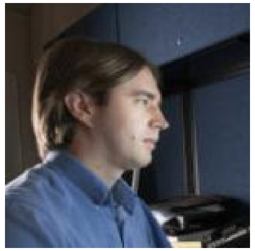
Sandia Staff

Kiran Lakkaraju
PI for TH
8716

Jason Reinhardt
Technical PM for
PEGASIS, Framing
8714

Andrew Reddie
Framing
8716

Ruby Booth
Game Design
8716



Josh Letchford
Game Design
8762

Chris Mairs
Cyber SME
5966

Natalie Prittinen
Cyber SME
8716

Jon Whetzel
Game
Implementation
6535

Nathan Fabian
Game
Implementation
6535

Representation from:

- Systems analysis
- Political Science/International Relations
- Computer Science
- Software engineering
- Cognitive psychology
- Cybersecurity
- Business

Management

Lynn Yang
PM for TH
8716

External Collaborators

Dr. Bethany Goldblum
UC-Berkeley

Gabriel Kelvin
UC-Berkeley – Poli. Sci
(Undergrad)

Interns

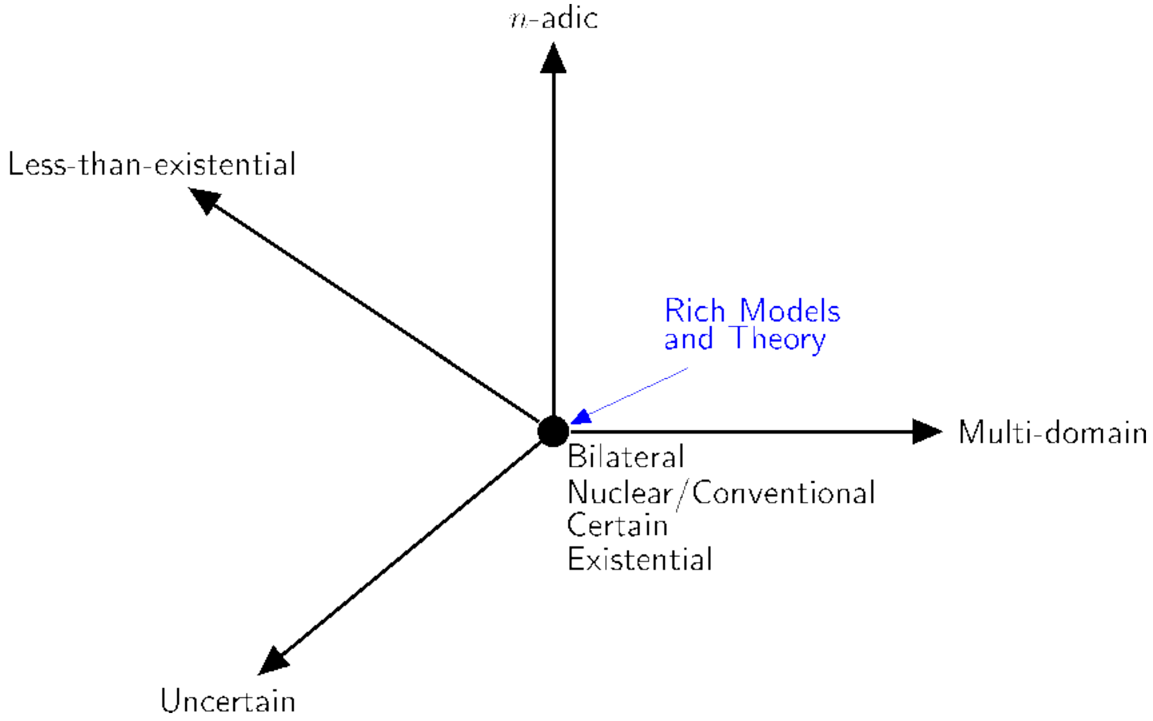
Mika Armenta
6672
U. Chicago – Psychology
(Grad)

Nicholas Blanchette
8716
MIT – Poli. Sci
(Grad)

Introduction to Tracing House and Experimental Wargaming

The problem: a complex conflict space.

1. Conflict has grown more complex.
2. Impact and reach of weapons has grown.
3. Uncertainty, of impact and of provenance, has grown.
4. Existing models do not capture this well.
5. Limited data.



Complexity-scarcity gap hinders our ability to study conflict spaces

Future conflict research exists in a "complexity-scarcity" gap, limiting strategic studies effectiveness and impact.

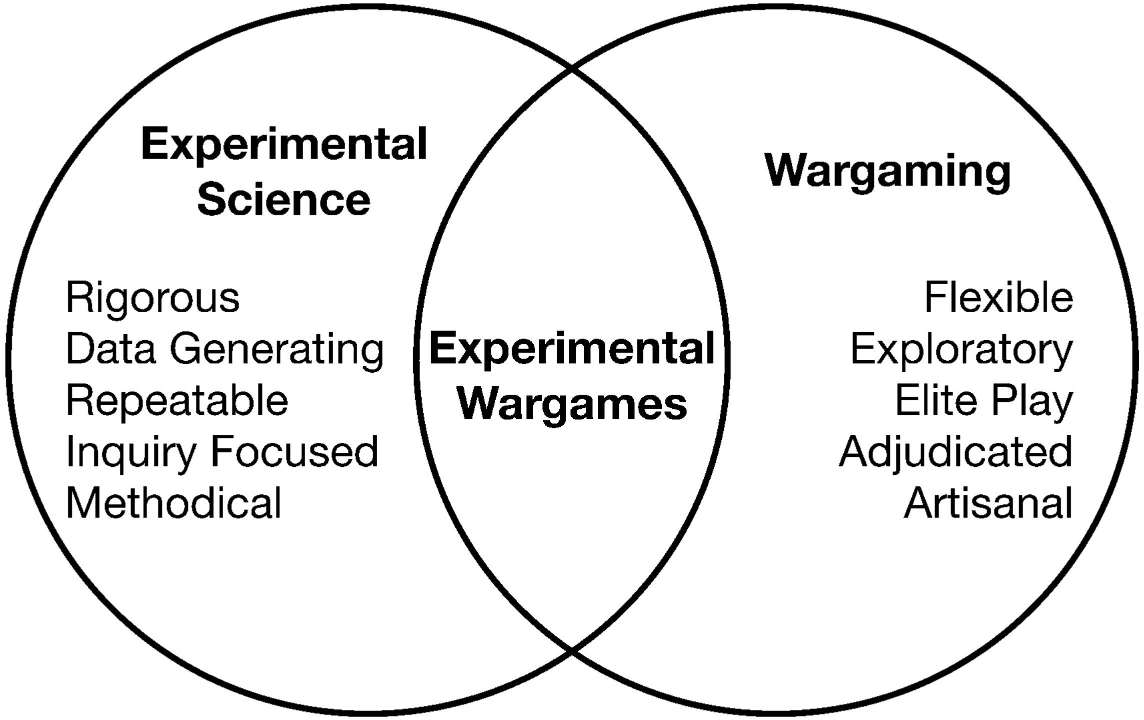
- **Complexity**
- Models from first principles difficult to construct.
- Complex interaction between numerous elements:
  - Individual characteristics
  - Cultural norms
  - Geopolitical considerations
  - Capabilities

- **Scarcity**
- Limited data on past conflicts.
- Existing data sparsely recorded
- Potentially biased collections of data.
- Limited "process data" – mainly focused on outcome.
- Limits the use of data driven algorithms.

How do we bridge this gap?

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



Experimental Science

Rigorous
Data Generating
Repeatable
Inquiry Focused
Methodical

Experimental Wargames

Wargaming

Flexible
Exploratory
Elite Play
Adjudicated
Artisanal

Traditional wargames are built for insight generation.

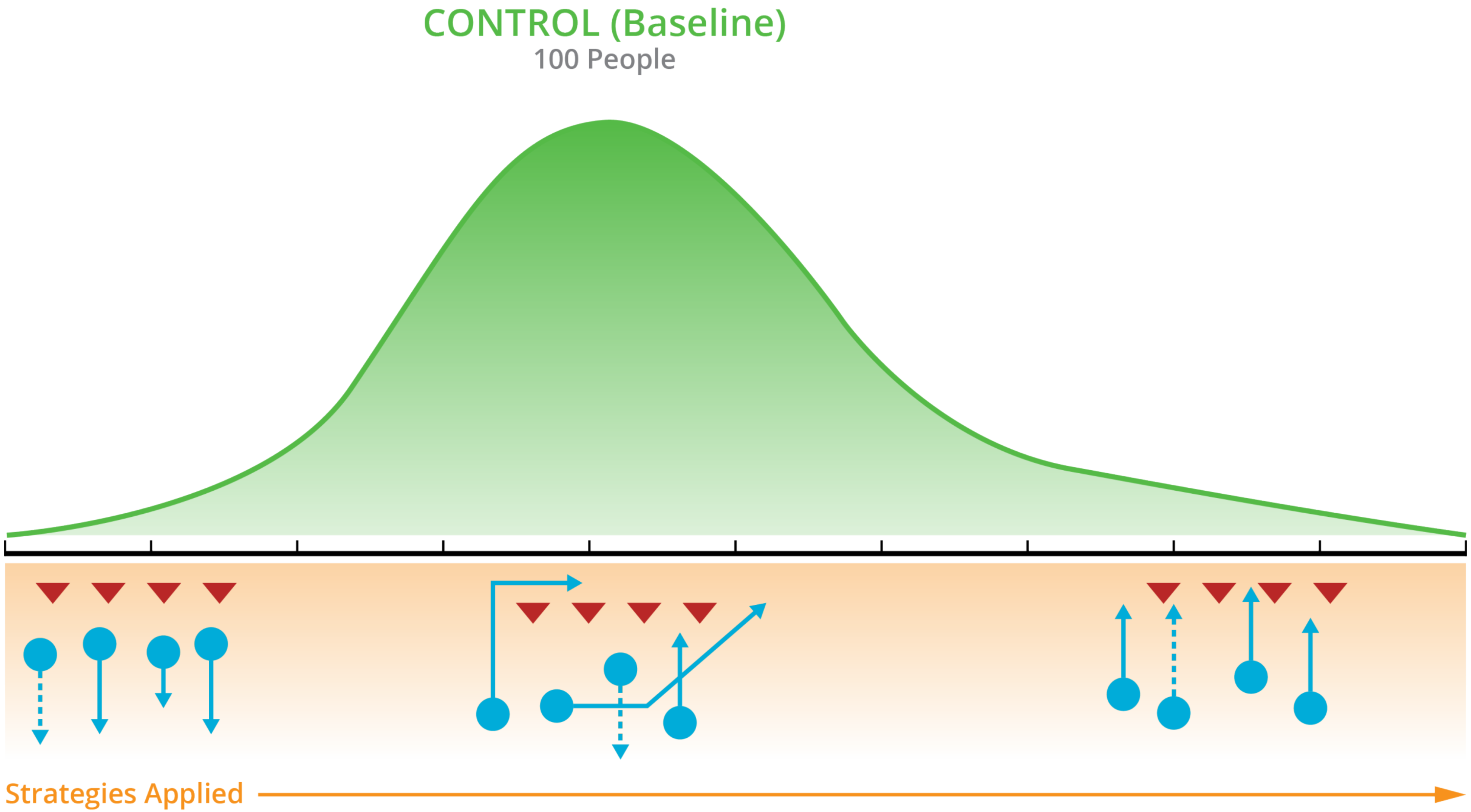
A wargame is a dynamic representation of conflict or competition in a synthetic environment, in which people make decisions and respond to the consequences of those decisions.[1]

- Traditional wargames focus on insight generation and exploration of strategies for players.
  - Seminar games.
  - Tactical games.
  - Matrix games.
- Traditional wargames are often:
  - Small numbers of players.
  - Difficult to replicate
  - Include adjudicators
- Wargames are rarely, if ever, used for data collection and experiments.

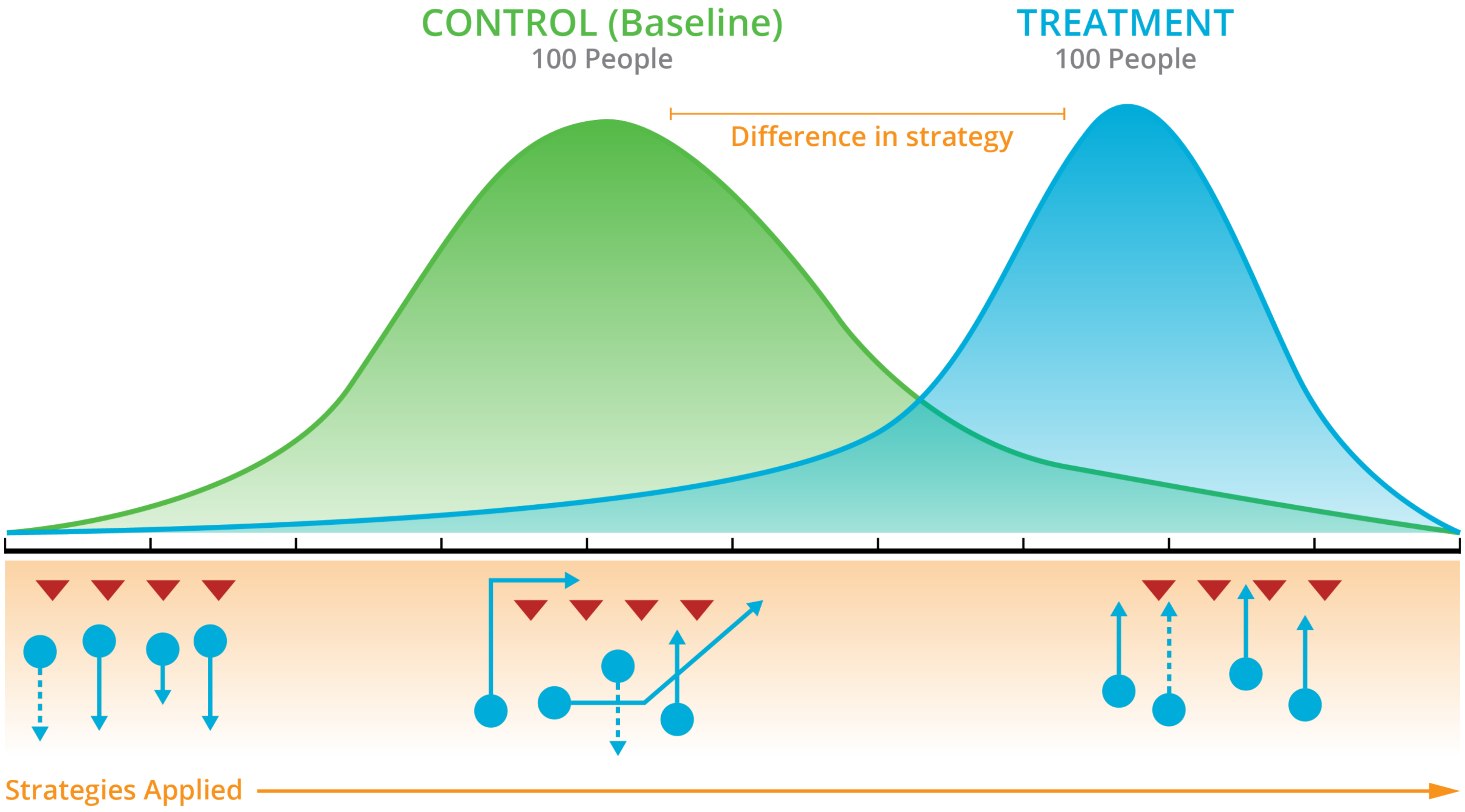
https://warontherocks.com/2019/08/wargaming-has-a-place-but-is-no-panacea-for-professional-military-education/

[1] Perla, "Terms of Reference", MORS Special Meeting, Wargaming Workshop, October, 17-21 October 2016.

We want to quantitatively understand behavioral distributions…



…and how those distributions change in controlled conditions.



SIGNAL was our first experimental wargame

- Project on Nuclear Gaming (PoNG) led by UC-Berkeley (PI: Michael Nacht) and partnered with Sandia National Labs and Lawrence Livermore National Labs.
- Studied the impact of tailored effect NW on conflict escalation.
- Hosted 10 events, gave 35 talks, awarded best student game at the SG&C.
- Collected data from more than 1000 players from around the world.

PROJECT ON NUCLEAR GAMING

NORTH SEA
SOUTH SEA
SIGNAL

SERIOUS GAMES SHOWCASE & CHALLENGE 2019 FINALIST

Gaming Research for Alliance Network Dynamics (GRAND)

- Platform to practice key elements of Alliance decision making in crisis scenarios
  - 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
- Data collection on crisis escalation
  - Supports development of models of crisis on & off ramps (precursors/de-escalatory factors)
- Configurable to multiple scenarios and varying number of players
- Online platform to allow players from across the world to participate from their locations
  - Better engages with important and busy personnel that have limited availability for standard wargames
- Funded by NATO-ACT

Understanding deterrence in a cyber context

- Deterrence concept historically applied in conventional and nuclear contexts.
- What kinds of actions or messages can influence adversary behavior?
- Cyber Deterrence:
  - How do we deter cyber attacks using cyber means? How do we deter state-sponsored cyber attacks using cyber means?
  - **How do we deter attacks (conventional/nuclear) using cyber means?**
- The possibility of cyber deterrence, and how it would operate, is part of a rich debate within the community.

> "Deterrence is a coercive strategy… the potential or actual application of force to influence the actions of a voluntary agent."
>
> – Lawrence Freedman
> *Deterrence*
> Policy Press, 2004, 130pp

Tracing House is a new LDRD project

PURDUE UNIVERSITY

Berkeley
UNIVERSITY OF CALIFORNIA

UNIVERSITY OF ILLINOIS
URBANA-CHAMPAIGN

Tracing House studies deterrence in a cyber context.

How does the communication-capability tradeoff impact the likelihood of making a deterrent threat and deterrence failure?

- Impact:
  - Better understanding of how to communicate to adversaries and impact cyber actor decision calculus
  - Develop a platform that will serve as a foundation to study additional strategic dilemmas

Tracing House will deliver a platform and an experimental wargame.

CASTLE

- **Platform** for studying strategic interaction scenarios.
- Allows us to:
  - Rapidly design and develop wargames.
  - Facilitate experimentation through scalable subject recruitment.
  - Facilitate analysis.

TANTALVS

- **Experimental wargame.**
- Designed to study the communication-capability tradeoff.
- Will address our research question/hypotheses.

University partners are enhancing core research of Tracing House.

PURDUE UNIVERSITY

Develop and apply data analytic methods and techniques to study strategic behavior in conflict scenarios.

Berkeley UNIVERSITY OF CALIFORNIA

Contribute to the development of Tantalus and CASTLE and informing data analysis methods.

UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN

Study the impact of laboratory effects on player behavior.

Agenda

- 9:00am – 9:15am: Introduction and charge to advisory committee (Kiran Lakkaraju)
- 9:15am – 9:45am: Introduction to the project and experimental wargaming. (Kiran Lakkaraju)
- **9:45am – 10:30am: Research Design (Andrew Reddie)**
- 10:30am – 10:45am: Discussion and Questions.
- 10:45am – 11:00am: Break
- 11:00am – 12:00pm: Game Design (Josh Letchford)
- 12:00pm – 12:30pm: Discussion

Sandia National Laboratories

Tracing House Research Design

Dr. Andrew Reddie

August 31, 2021

PEGASIS

PROGRAM for EXPERIMENTAL GAMING & ANALYSIS of STRATEGIC INTERACTION SCENARIOS

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.

BLUF

- Scholars and policy-makers have struggled to conceptualize what deterrence looks like within the cyber domain
  - “Drastically different”
  - “Far more complicated”
  - “Overrated”
- And what state actors might be using cyber deterrent threats against…
  - State actors using cyber means to attack military/government/civilian infrastructure
  - State actors using any means to attack military/government/civilian infrastructure
  - Nonstate actors using cyber means to attach military/government/civilian infrastructure
- As this debate has largely occurred in theoretical terms, it is ripe for empirical research
  - And where empirical study has occurred, there has been an over-reliance on formal methods to study these issues
- **Experimental methods, in general, and experimental wargaming methods, in particular, offer a means to engage with these questions and to examine the conditions under which cyber deterrent threats are made and are successful or fail**

Outline

- 1. Cyber Deterrence in Theory
- 2. The Communication-Capability Tradeoff
- 3. The RQs
- 4. Hypotheses
- 5. Experimental Design
- 6. Survey Design
- 7. Game Design
- 8. Findings and Future Research

TRACING HOUSE

TANTALVS

PEGASIS PROGRAM for EXPERIMENTAL GAMING & ANALYSIS of STRATEGIC INTERACTION SCENARIOS

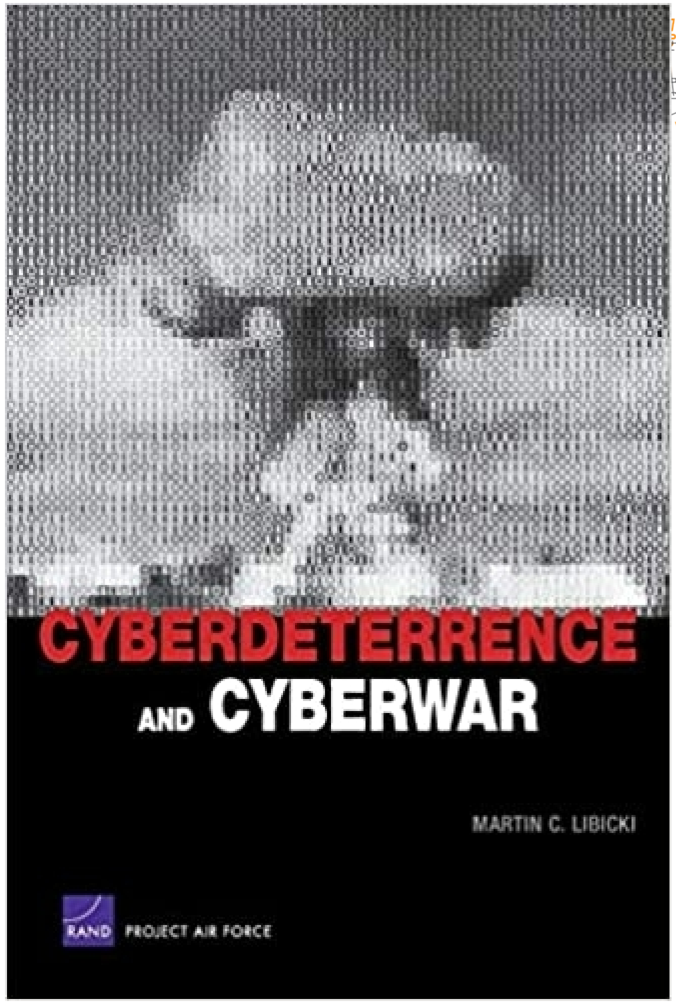
1. Cyber Deterrence in Theory

- The theoretical debate regarding the appropriateness of pursuing a cyber deterrence strategy remains robust (Nye 2016; Mandel 2017; Brantly 2018; Wilner 2020; Klimburg 2020)

Libicki (2009) points to three fundamental challenges associated with cyber deterrence:

- Do we know who did it?
- **Can we hold their assets at risk?**
- **Can we do so repeatedly?**

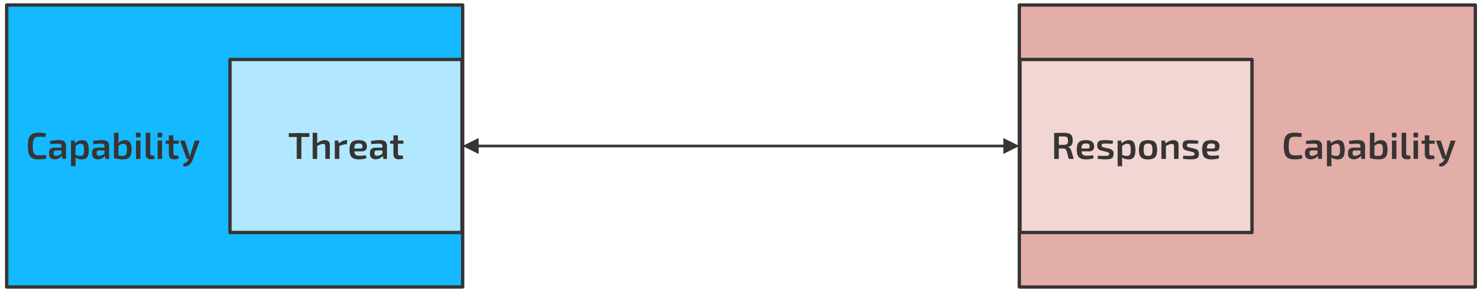
Most scholarship has been focused on the attribution problem (Borghard and Lonergan 2016)



Libicki, Martin C. *Cyberdeterrence and cyberwar*. RAND corporation, 2009.

2. The Capability-Communication Tradeoff (CCT) in Theory

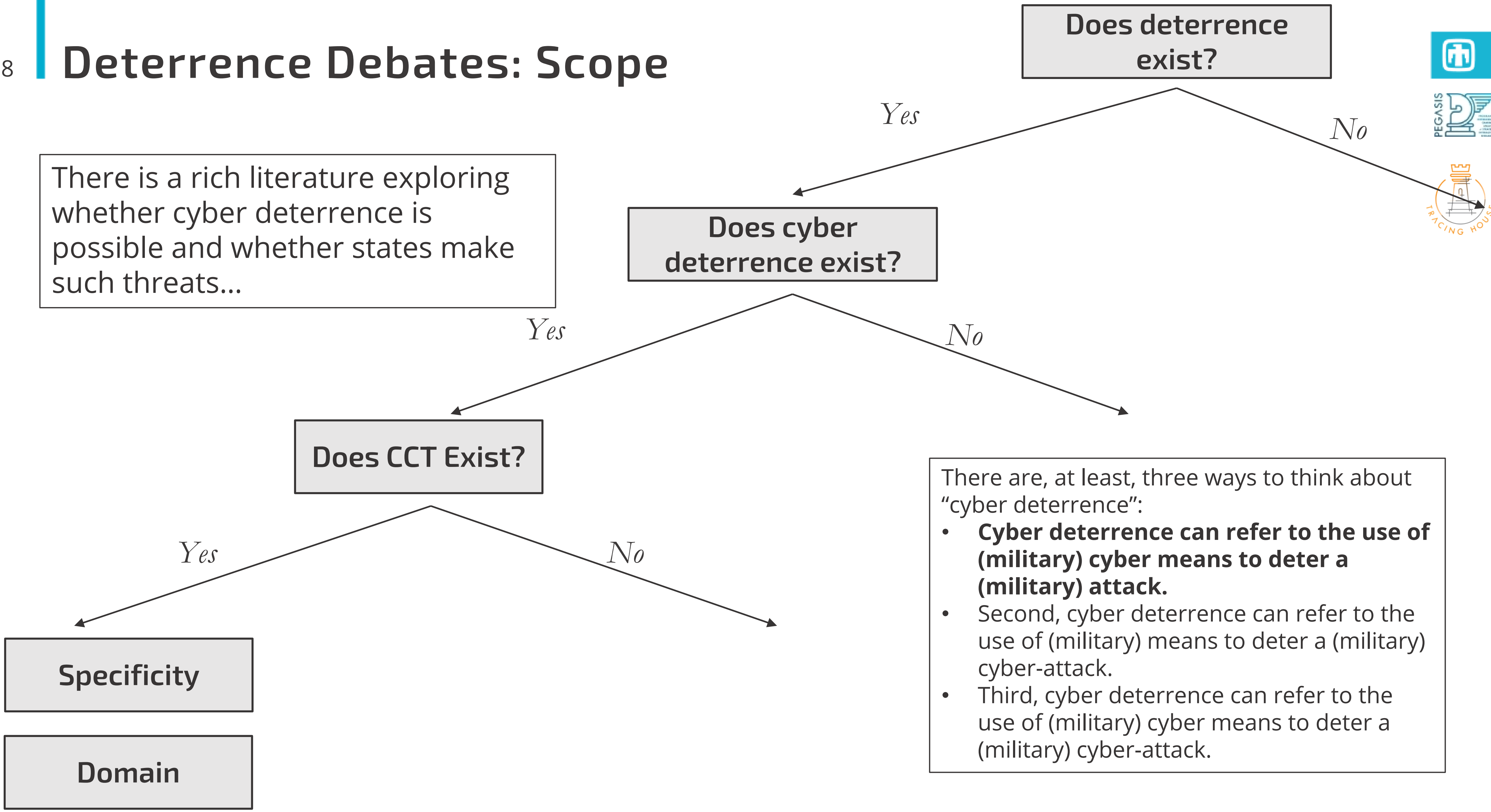
- Deterrence (by punishment) requires Blue's credible communication of a deterrent threat to a Red adversary…



- In the cyber domain, this is theorized to be challenge via two theorized mechanisms:
  - *Muting Driver:*
    - As Blue communicates a cyber deterrent threat, Red might **mute** its effects (defense)
  - *Mirroring Driver:*
    - As Blue communicates a cyber deterrent threat, Red might **mirror** the capability (offense)

This conundrum facing blue is analogous to the choice as to whether to "reveal or conceal" (Green & Long 2020)

Deterrence Debates: Scope



Does deterrence exist?

Yes

No

Does cyber deterrence exist?

Yes

No

Does CCT Exist?

Yes

No

Specificity

Domain

There is a rich literature exploring whether cyber deterrence is possible and whether states make such threats…

There are, at least, three ways to think about "cyber deterrence":

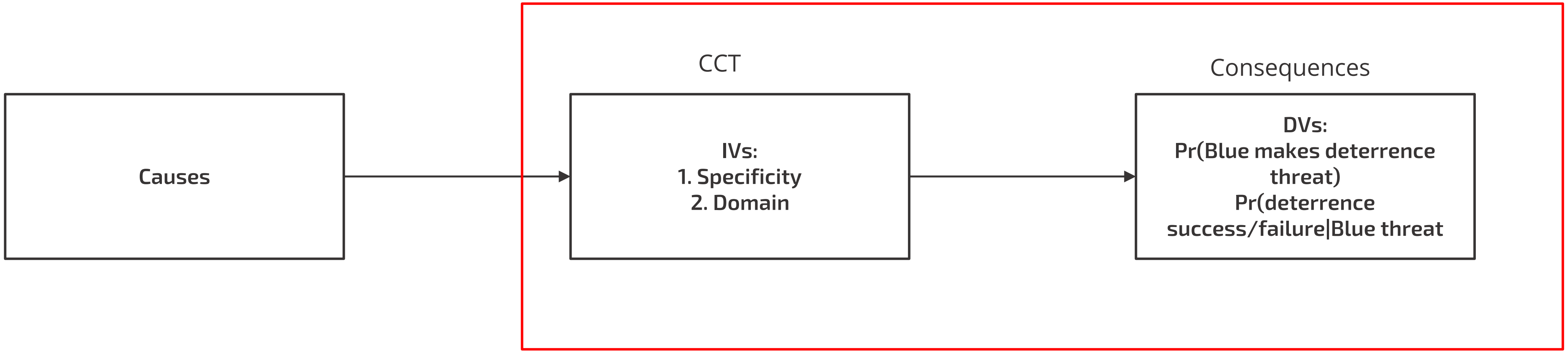
- **Cyber deterrence can refer to the use of (military) cyber means to deter a (military) attack.**
- Second, cyber deterrence can refer to the use of (military) means to deter a (military) cyber-attack.
- Third, cyber deterrence can refer to the use of (military) cyber means to deter a (military) cyber-attack.

3. Research Question(s)

We are primarily concerned about the use of cyber threats for deterrence ends. Thus,

- **How does the variation in the *domain* and *specificity* of a deterrent threat affect deterrence?**

1. Does this variation affect the likelihood of making a deterrent threat?
2. Does this variation affect the likelihood of deterrence failure?


CCT
Consequences
Causes
IVs:
1. Specificity
2. Domain
DVs:
Pr(Blue makes deterrence threat)
Pr(deterrence success/failure|Blue threat

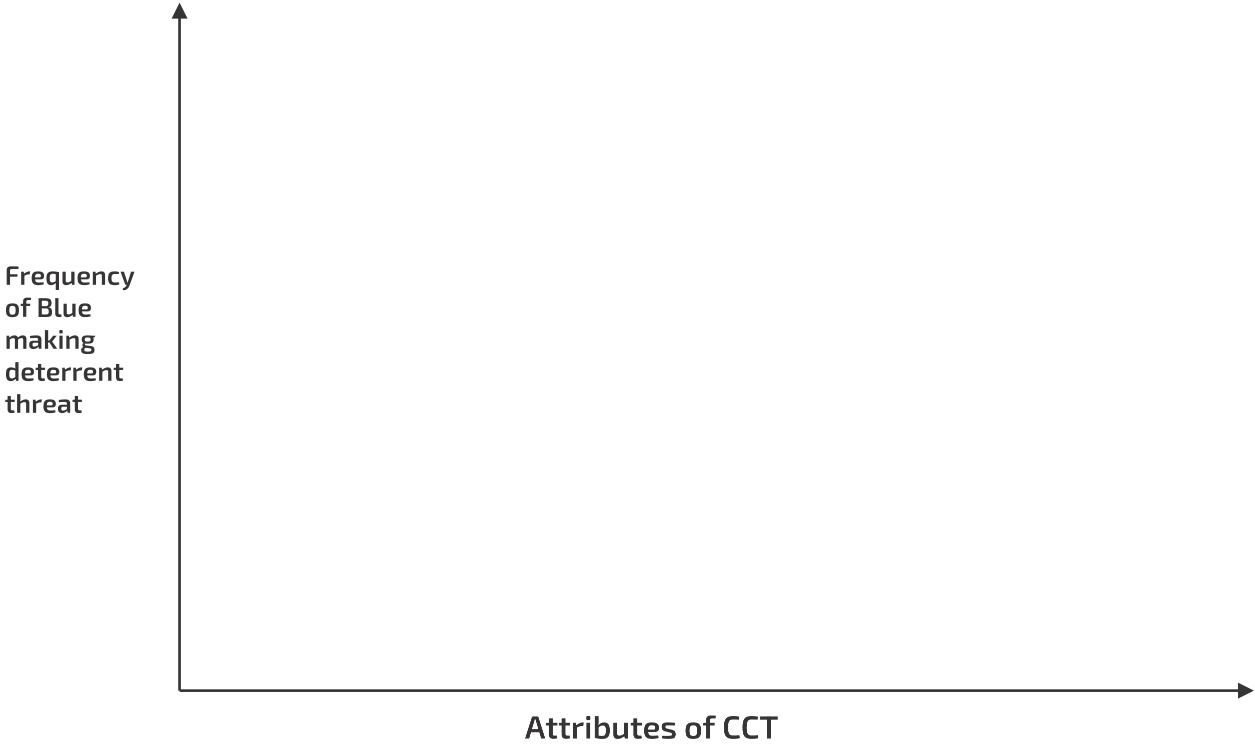
Independent Variable: Characteristics of the CCT

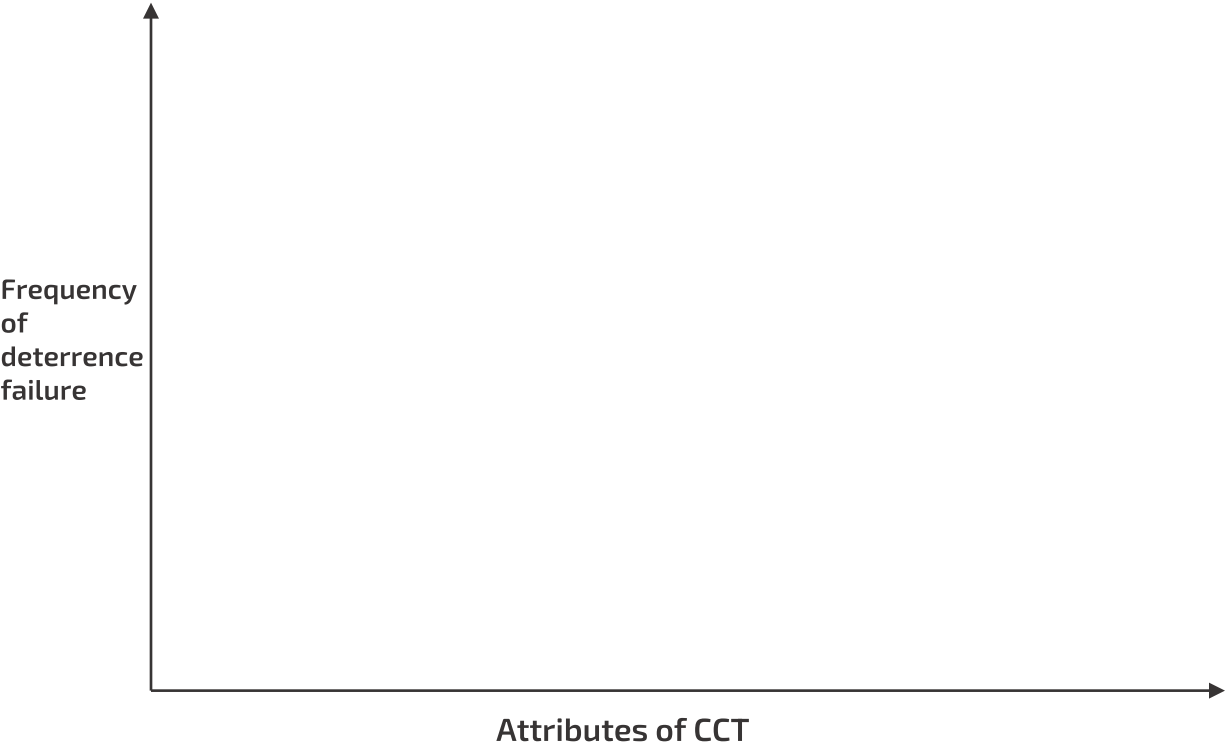
- We vary the characteristics capability communication trade-off as the instrument in our experiment:
  - Varying the **domain** of threat
    - Cyber
    - Conventional
    - Nuclear*
  - AND varying the **specificity** of the threat
    - Attack Vector
      - E.g. missile strike, malware
    - Target
      - Variation by type
        - Military, government, civilian
      - Variation by facility*

	No Domain	Conventional	Cyber
Vague Threat	1.	2. Low CCT	4. Med CCT
Specific Threat	Null	3. Low CCT	5. High CCT

Dependent Variable: Deterrer and Deterree

- We measure deterrence outcomes related to CCT in two ways:
  - The likelihood of a player making a deterrent threat | attributes of the deterrence threat
  - The likelihood of deterrence failure | attributes of the deterrent threat | deterrent threat is made





4. Hypotheses: Specificity of Deterrent Threat

H_{1A}: If Blue’s available deterrent threats are **specific**, then attempts to deter occur **less**, all else equal.
H_{1B}: If Blue’s available deterrent threats are **specific**, then there is **no effect** on attempts to deter.
H_{1C}: If Blue’s available deterrent threats are **specific**, then attempts to deter occur **more**, all else equal.

If A is observed:

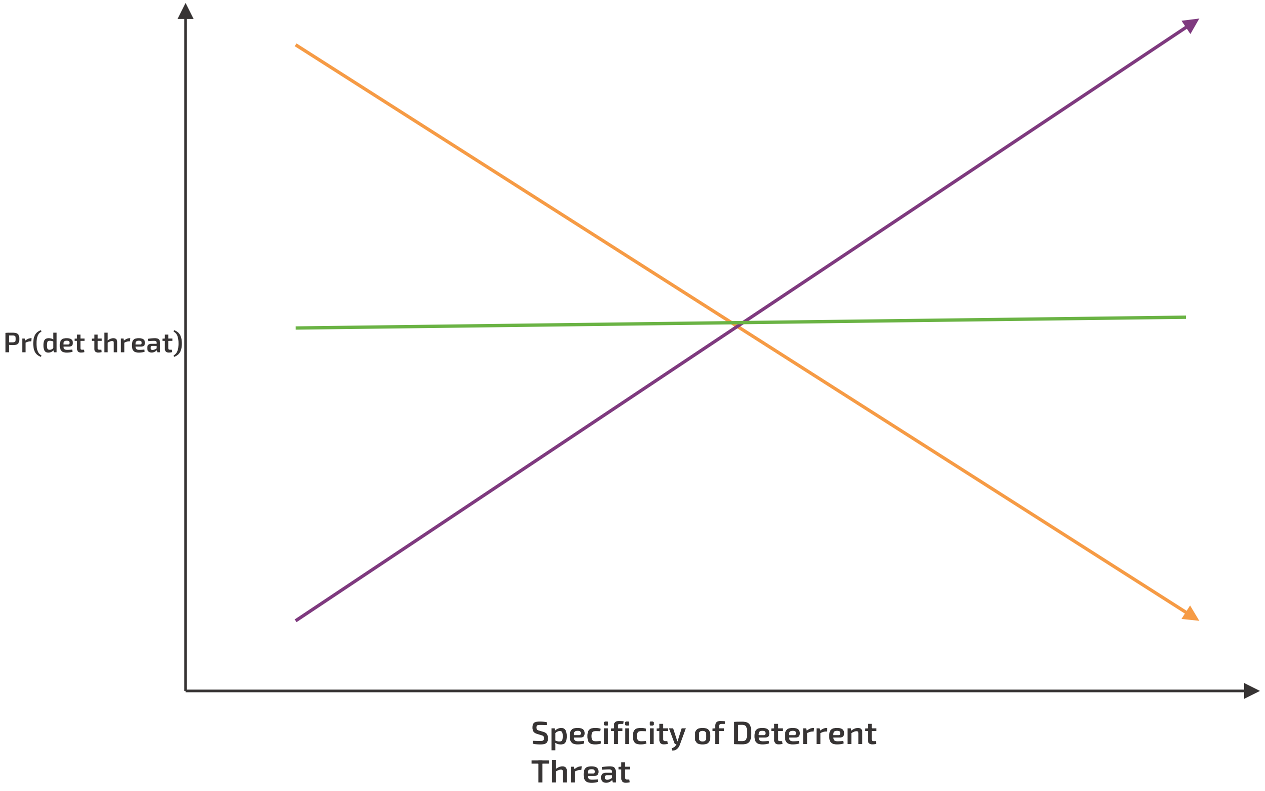
- Blue is attempting deterrence less
- We infer that the CCT is driving behavior that reflects existing theory

If B is observed:

- Blue does not change their deterrence behavior
- We infer,
    - 1) CCT does not exist or 2) CCT is not captured

If C is observed:

- Blue is attempting deterrence more
- We infer that the CCT is driving behavior, but in the opposite direction of existing theory

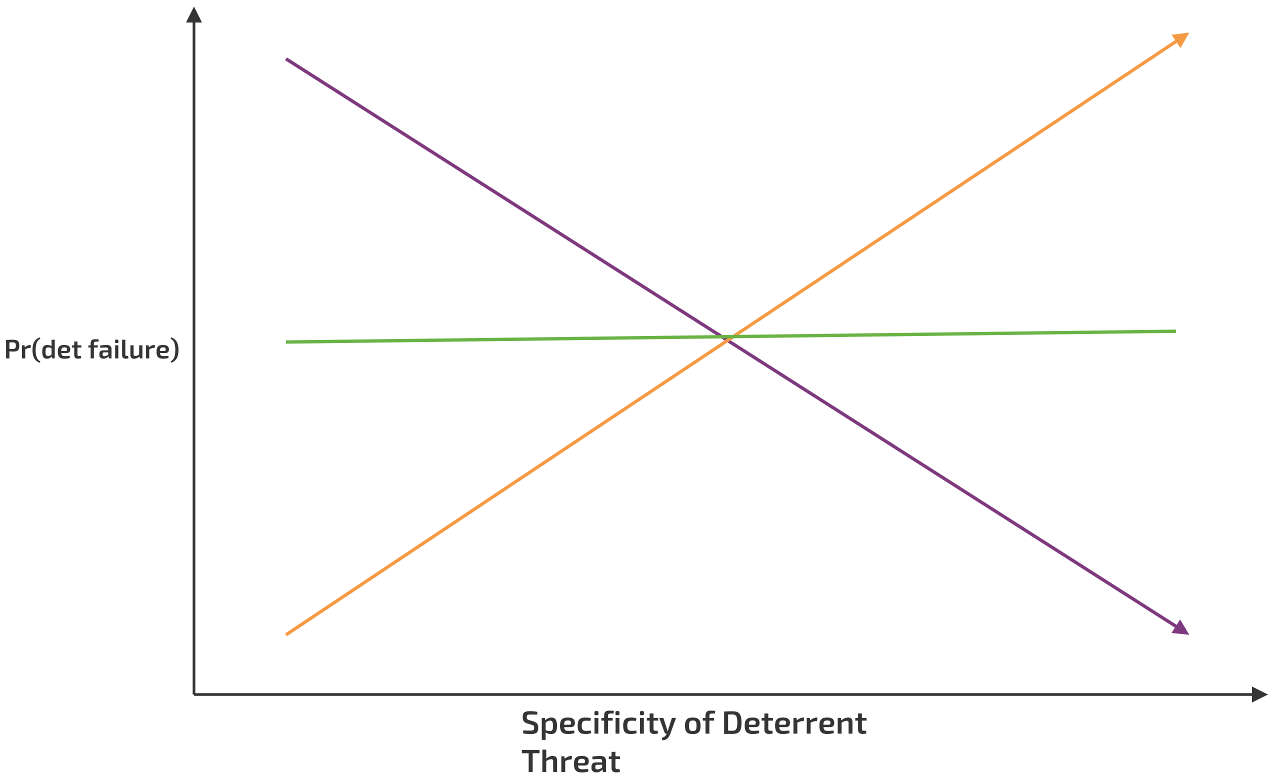


4. Hypotheses: Specificity of Deterrent Threat

H_{2A}: If Blue’s available deterrent threats are **specific**, then deterrence failure is **more** likely, all else equal.

H_{2B}: If Blue’s available deterrent threats are **specific**, then deterrence failure is **no more** likely, all else equal.

H_{2C}: If Blue’s available deterrent threats are **specific**, then deterrence failure is **less** likely, all else equal.



If A is observed:

- Red is not internalizing Blue’s deterrent threat
- We infer that the CCT is driving behavior that reflects existing theory

If B is observed:

- Red does not change their deterrence behavior
- We infer,
  - 1) CCT does not exist or 2) CCT is not captured

If C is observed:

- Red is internalizing Blue’s deterrent threat
- We infer that the CCT is driving behavior, but in the opposite direction of existing theory

4. Hypotheses: Domain of Deterrent Threat (Cyber vs. Other)

H_{3A}: If Blue’s available deterrent threats are **cyber**, then attempts to deter occur **less**, all else equal.

H_{3B}: There is no delta in the probability of deterrence attempt between cyber and non-cyber deterrent threats

H_{3C}: If Blue’s available deterrent threats are **cyber**, then attempts to deter occur **more**, all else equal.

If A is observed:

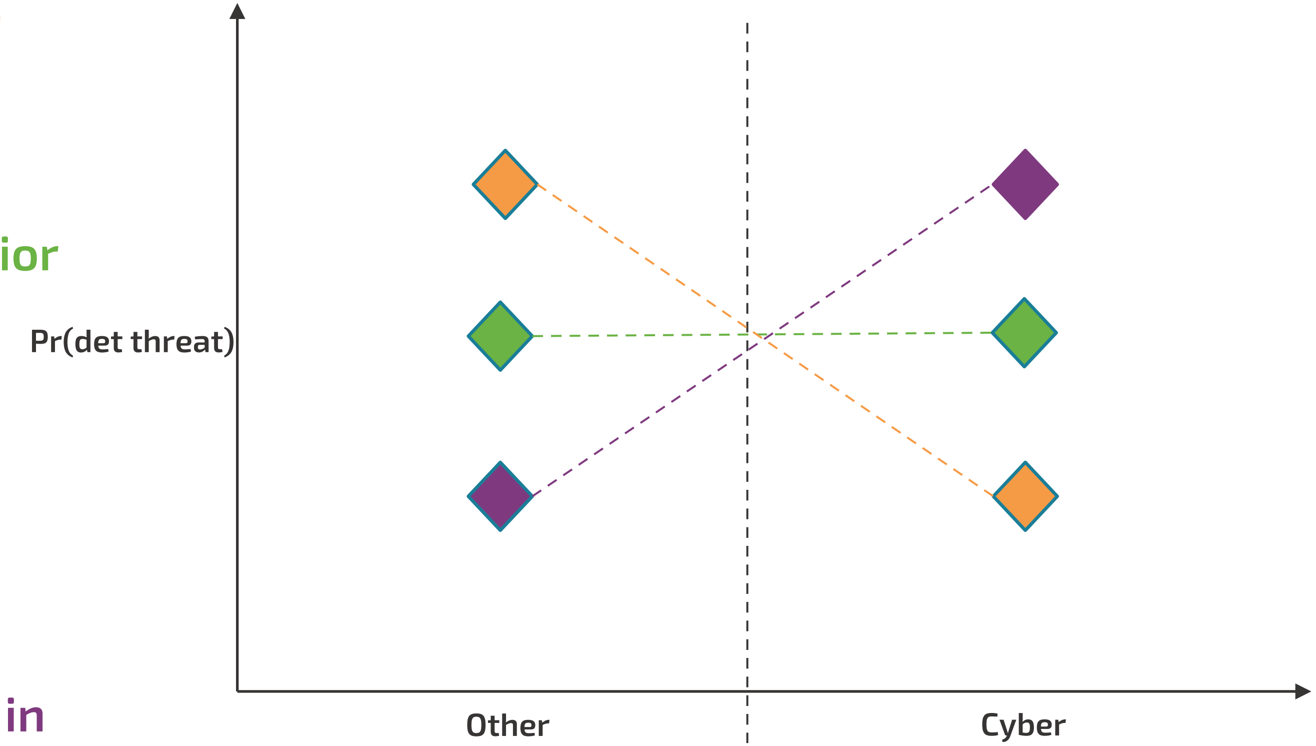
- **Blue is attempting cyber deterrence less**
- **We infer that the CCT is driving behavior that reflects existing theory**

If B is observed:

- **Blue does not change their deterrence behavior across domains**
- **We infer,**
  - **1) CCT does not exist or 2) CCT is not captured**

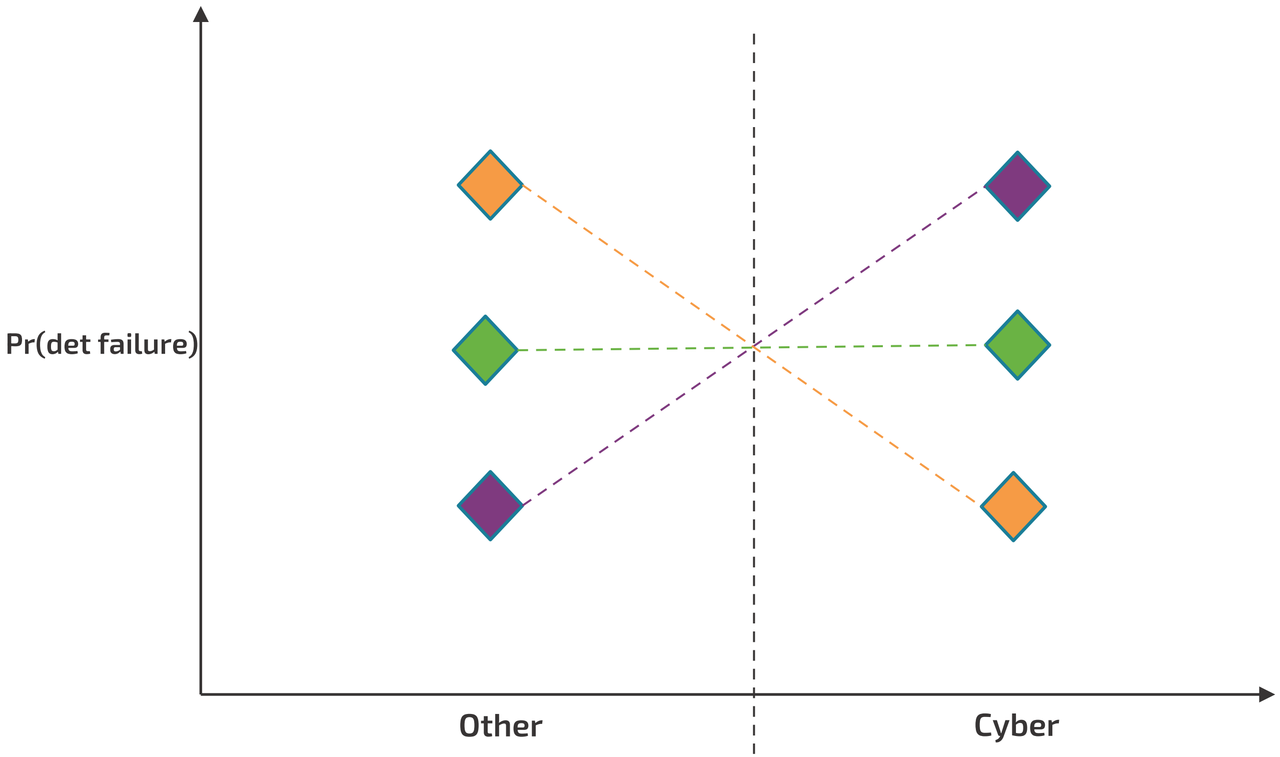
If C is observed:

- **Blue is attempting cyber deterrence more**
- **We infer that the CCT is driving behavior, but in the opposite direction of existing theory**



4. Hypotheses: Domain of Deterrent Threat (Cyber vs. Other)

H_{4A}: If Blue's available deterrent threats are **cyber**, then deterrence failure is **more** likely, all else equal.
H_{4B}: There is no delta in the probability of deterrence failure between cyber and non-cyber deterrent threats
H_{4C}: If Blue's available deterrent threats are **cyber**, then deterrence failure is **less** likely, all else equal.



If A is observed:

- Blue experiences cyber deterrence failure more
- We infer that the CCT is driving behavior that reflects existing theory

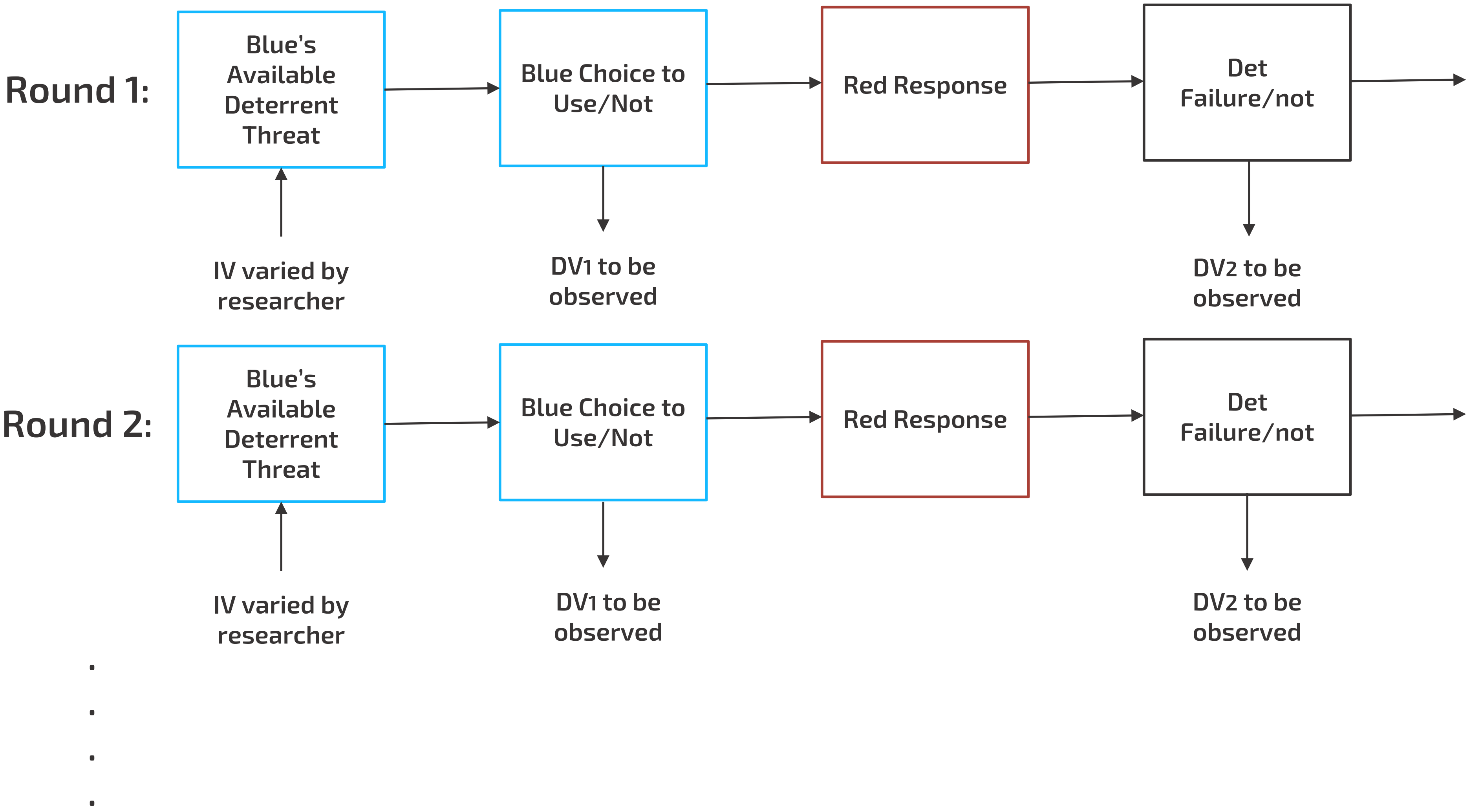
If B is observed:

- Blue does not change their deterrence behavior
- We infer,
  - 1) CCT does not exist or 2) CCT is not captured

If C is observed:

- Blue experiences cyber deterrence failure less
- We infer that the CCT is driving behavior, but in the opposite direction of existing theory

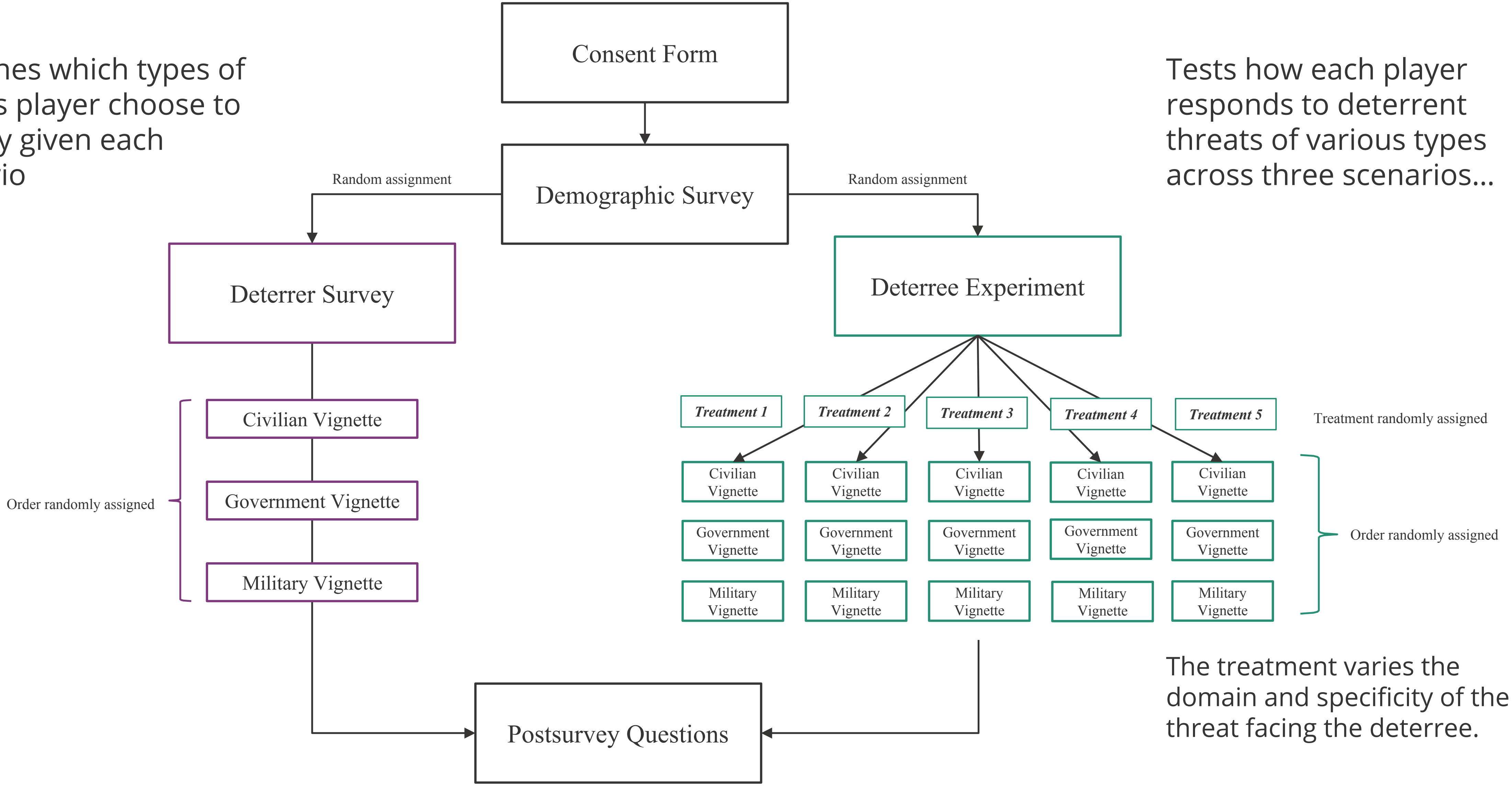
5. Generalized Experimental Design



6. Survey(s) Experimental Design

Examines which types of threats player choose to employ given each scenario

Tests how each player responds to deterrent threats of various types across three scenarios...



Consent Form

Demographic Survey

Random assignment

Random assignment

Deterrer Survey

Civilian Vignette

Government Vignette

Military Vignette

Order randomly assigned

Deterree Experiment

Treatment 1

Treatment 2

Treatment 3

Treatment 4

Treatment 5

Treatment randomly assigned

Civilian Vignette

Government Vignette

Military Vignette

Order randomly assigned

Postsurvey Questions

The treatment varies the domain and specificity of the threat facing the deterree.

7. The Case for a Wargaming Approach

- Offers a synthetic data-generating process for a research problem that remains largely theoretical
  - E.g. Where is the empirical example of a cyber threat being made?
- Deterrence represents a strategic interaction
  - Both players “get a vote”
- It is an intrinsically human process
  - The conditions under which rational actor models apply or not largely remains untested

7. Game Design Considerations:

Challenging requirements for game design:

- Implementing the treatment conditions.
  - Allow for the variation of the deterrent threat available to players in the game.
    - Domain
    - Specificity
      - Attack Vector
      - Target
      - Behavior
- Linking treatment conditions to dependent variables of interest.
  - Allowing for the measurement of the two dependent variables.
- The creation of conditions under which there is a Behavior to be deterred
  - i.e. There is something that Red would otherwise do that Blue must decide whether to deter or not.

8. Potential Findings and Future Research

- The proposed research design adjudicates whether the domain and the specificity of a deterrent threat has an influence on the incidence of a deterrent threat being made.
- It also adjudicates whether the domain and the specificity of a deterrent threat has an influence on the likelihood of deterrence success or failure.
- As such, **it directly contributes to research concerning the viability of using cyber threats (and, potentially, other domains) to deter an adversary.**
- The capability developed for this project also represents a testbed for further examination of threat behavior, the types of actions that states might be interested in deterring, and the actions that stem from deterrence failure vis a vis punishment.
  - There may be a fairly easy extension to considering non-state actors.
    - Traditionally treated as being difficult to deter given the limited to hold what they value at risk.

Sources

Goodman, Will. "Cyber deterrence: Tougher in theory than in practice?." *Strategic Studies Quarterly* 4, no. 3 (2010): 102-135.

Tor, Uri. "'Cumulative deterrence'as a new paradigm for cyber deterrence." *Journal of Strategic Studies* 40, no. 1-2 (2017): 92-117.

Brantly, Aaron F. "The cyber deterrence problem." In *2018 10th International Conference on Cyber Conflict (CyCon)*, pp. 31-54. IEEE, 2018.

Crosston, Matthew D. "World gone cyber MAD: How "mutually assured debilitation" is the best hope for cyber deterrence." *Strategic studies quarterly* 5, no. 1 (2011): 100-116.

Nye Jr, Joseph S. "Deterrence and dissuasion in cyberspace." *International security* 41, no. 3 (2016): 44-71.

Wilner, Alex S. "US cyber deterrence: Practice guiding theory." *Journal of Strategic Studies* 43, no. 2 (2020): 245-280.

Harknett, Richard J., and Joseph S. Nye Jr. "Is deterrence possible in cyberspace?." *International Security* 42, no. 2 (2017): 196-199.

Klimburg, Alexander. "Mixed signals: A flawed approach to cyber deterrence." *Survival* 62, no. 1 (2020): 107-130.

Libicki, Martin C. "Expectations of cyber deterrence." *Strategic Studies Quarterly* 12, no. 4 (2018): 44-57.

Questions?

areddie@sandia.gov

Agenda

- 9:00am – 9:15am: Introduction and charge to advisory committee (Kiran Lakkaraju)
- 9:15am – 9:45am: Introduction to the project and experimental wargaming. (Kiran Lakkaraju)
- 9:45am – 10:30am: Research Design (Andrew Reddie)
- 10:30am – 10:45am: Discussion and Questions.
- 10:45am – 11:00am: Break
- **11:00am – 12:00pm: Game Design (Josh Letchford)**
- 12:00pm – 12:30pm: Discussion

TANTALUS Game Design

Joshua Letchford

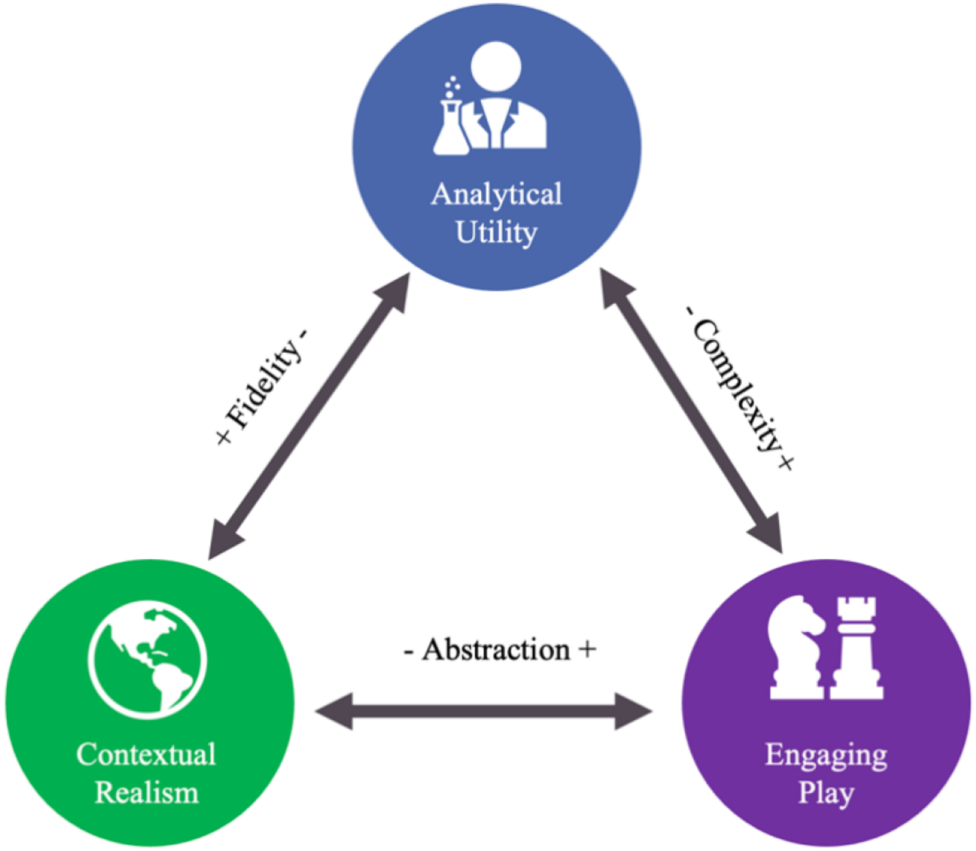
Ruby Booth, Gabriel Kelvin, Nick Blanchette

August 31, 2021

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.

TRILEMMA

- What aspects of **contextual reality** can't we live without?
- What elements are required for **analytical utility**?
- What do we need for **player engagement**?



OVERVIEW

- Assumptions and basic elements
- Modeling of capabilities
- Implementation of threats
- Number of players & victory conditions

ASSUMPTIONS

- Three player game with near peer capabilities
- High capability nations
  - Proxies not used for capability enhancement
  - Capability development is not critical to model
- Sub-strategic level of conflict (no existential level of threat), no NW
- 90 minute playtime + tutorial
- Online is primary focus for data collection
  - Board and TTX planned to be used for prototyping and engagement

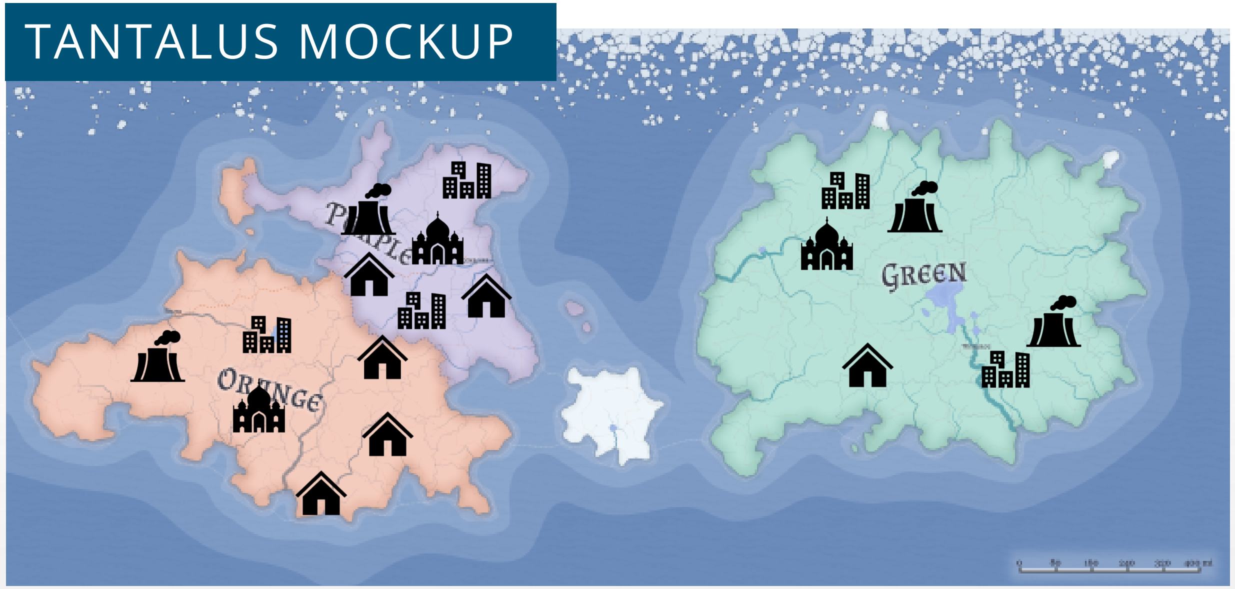
BASIC GAME ELEMENTS

- Competitive game
  - Achieve this via creating conflicting goals
- Simultaneous action selection followed by simultaneous action execution
- Rounds are timed
- Limited asymmetry
  - No significant asymmetry in capability
  - Limited asymmetry in terms of goal availability or difficulty
- End of game
  - Players have some uncertainty on when the game will end
    - Avoids some backwards induction issues

EXAMPLE GAME REPRESENTATION

- Two main elements in the play area
  - Top half provides geographical context
    - Highlight critical **assets**
    - Potential **targets** for attacks
  - Bottom half captures important **metrics** (M1..M5) for each nation
    - E.G. Economic Strength
- Players change the state of the game (metrics) via **actions**
- **Goals** exist that players attempt to achieve that are defined in terms of these metrics
- **Victory** is determined by achieved goals
- Players can make **threats** to other players to attempt to shape the strategies and goals other players are pursuing

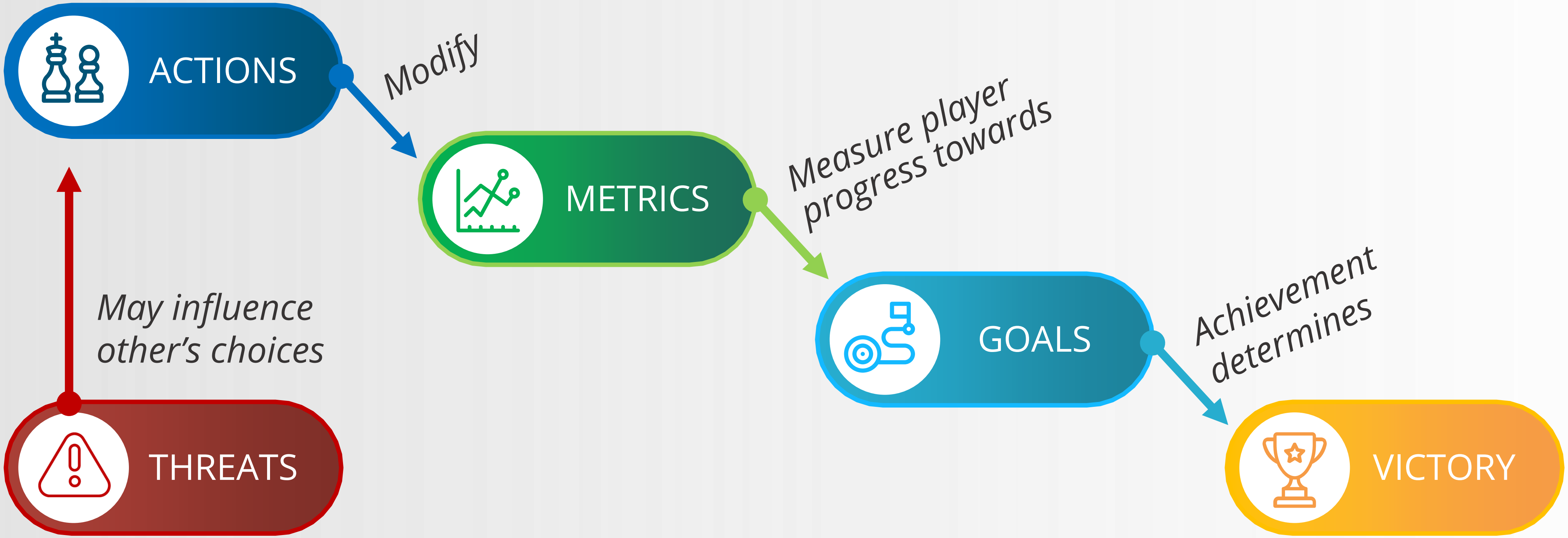
TANTALUS MOCKUP



	0	1	2	3	4	5	6	7	8	9
M1		🟢				🟣			🟠	
M2		🟠				🟢			🟣	
M3		🟣				🟠			🟢	
M4						🟠		🟣		🟢
M5	🟢		🟣		🟠					

Note: this figure is intended as a mockup for illustrative purposes but will likely not resemble the final product

ACTIONS, METRICS, AND GOALS



ACTIONS

Modify

METRICS

Measure player progress towards

GOALS

Achievement determines

VICTORY

May influence other's choices

THREATS

MODELING CAPABILITIES

ACTIONS

- Primary domains of interest
  - Cyber
  - Kinetic
- Some level of parity between Cyber and Kinetic options for the players
  - Driven by RQs
- Three classes of actions
  - Punitive – actions primarily focused on hurting other players
  - Mixed – actions with beneficial consequences for the player executing the action and meaningful downsides for other players
  - Beneficial – actions with beneficial consequences for the player executing the action but are either beneficial or benign for the other players
- All players have all capabilities at all times

KEY DIFFERENCES BETWEEN CYBER AND KINETIC

ACTIONS

- **Attribution***
  - Proxies
- **Fragility of capabilities***
  - Muting
- Speed of capability development
  - Mirroring (proliferation)
- **Effect uncertainty***
  - Larger or smaller than expected impact
  - Possibility of friendly fire
- **Cost of attacks***
  - Lower monetary cost?
  - Lower reputational costs?

LEVEL OF ABSTRACTION

ACTIONS

- These key differences drive what is necessary to model in the game
  - And what we can abstract out
- Beyond these key differences, we need similar level of impact and use
  - Avoid framing issues in how we represent these capabilities to the players
- Challenging because means and outcome are often conflated in cyber

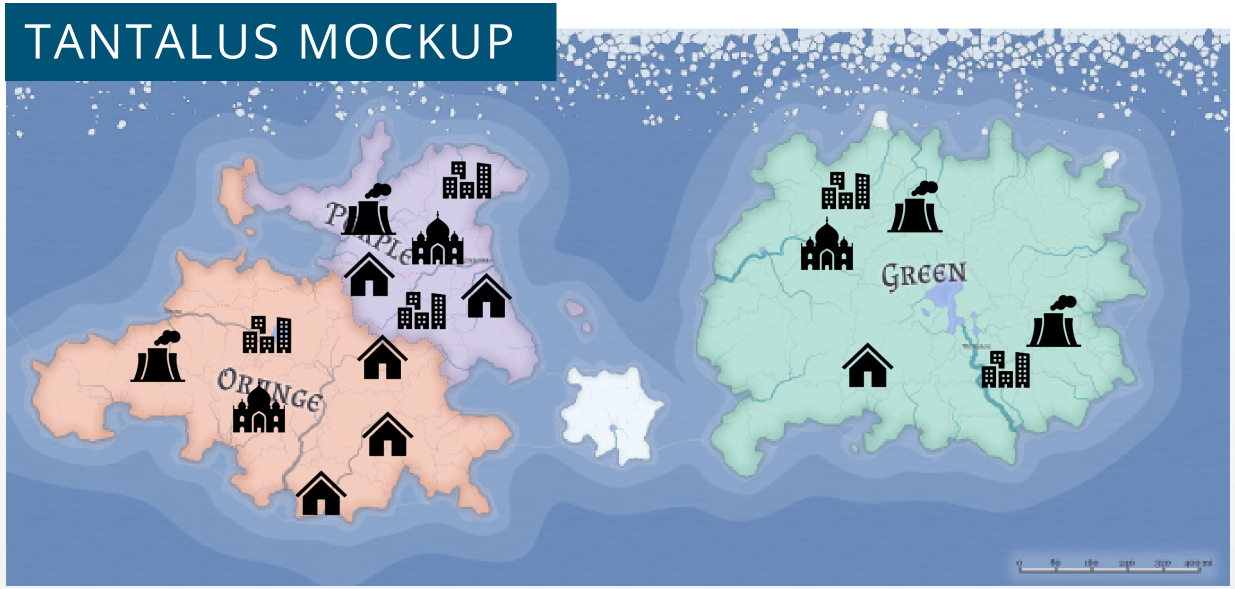
ROUND 1, TIME POINT 0

The first thing that players do each round is choose what action they want to take:

A1, A2, A3…

Green considers his options and decides to take action A3 this round.

TANTALUS MOCKUP



Note: this figure is intended as a mockup for illustrative purposes but will likely not resemble the final product

	0	1	2	3	4	5	6	7	8	9
M1		🟢				🟣			🟠	
M2		🟠				🟢			🟣	
M3		🟣				🟠			🟢	
M4						🟠		🟣		🟢
M5	🟢		🟣		🟠					

ANATOMY OF A THREAT

THREATS

- **Source:** Player who is issuing the threat (Deterrer)
  - P1, P2, P3
- **Recipient:** Player being threatened (Deterree)
  - P1, P2, P3
- **Behavior:** Adversary behavior (action) you are trying to prevent.
  - Constrained language to specify specific behaviors of interest.
- **Threated Action:** Threatened response
  - Attack Vector x Target
- **Attack Vector:** How threatened response will be achieved
  - Cyber: E.g. Ransomware, unspecified, or null/DNE.
  - Conventional: E.g. Missile, "Missile targeting critical infrastructure”, unspecified, or null/DNE.
- **Target:** Where threatened response will be achieved
  - E.g. economy, infrastructure, unspecified
- **Communicated to:** Set of players the threat was made in the presence of
  - Focusing on private threats (only source and recipient are aware of the threat)

WHAT DO THREATS LOOK LIKE?

THREATS

- **Threats as a structured event**
  - Provide the players with partially pre-built threats
    - "Behavior" that the players given the option to try to deter is pre-determined
    - Victory conditions and player goals are designed to make these threat options relevant
  - Player freedom
    - If they wish to make a threat
    - On how they want to threaten (within the constraint of the experimental condition they are in)
  - Players can take actions without making a corresponding threat
- **To make sure that all threats are capturable (and have the appropriate consequences) we are choosing to restrict communication**
  - Eliminate chat to avoid players making informal threats
- **Negative externalities for making a threat**
  - Muting
  - Mirroring
- **Positive externalities for making a threats**
  - Deterrence

Example threat:
If you take an action that attacks my economy I will:

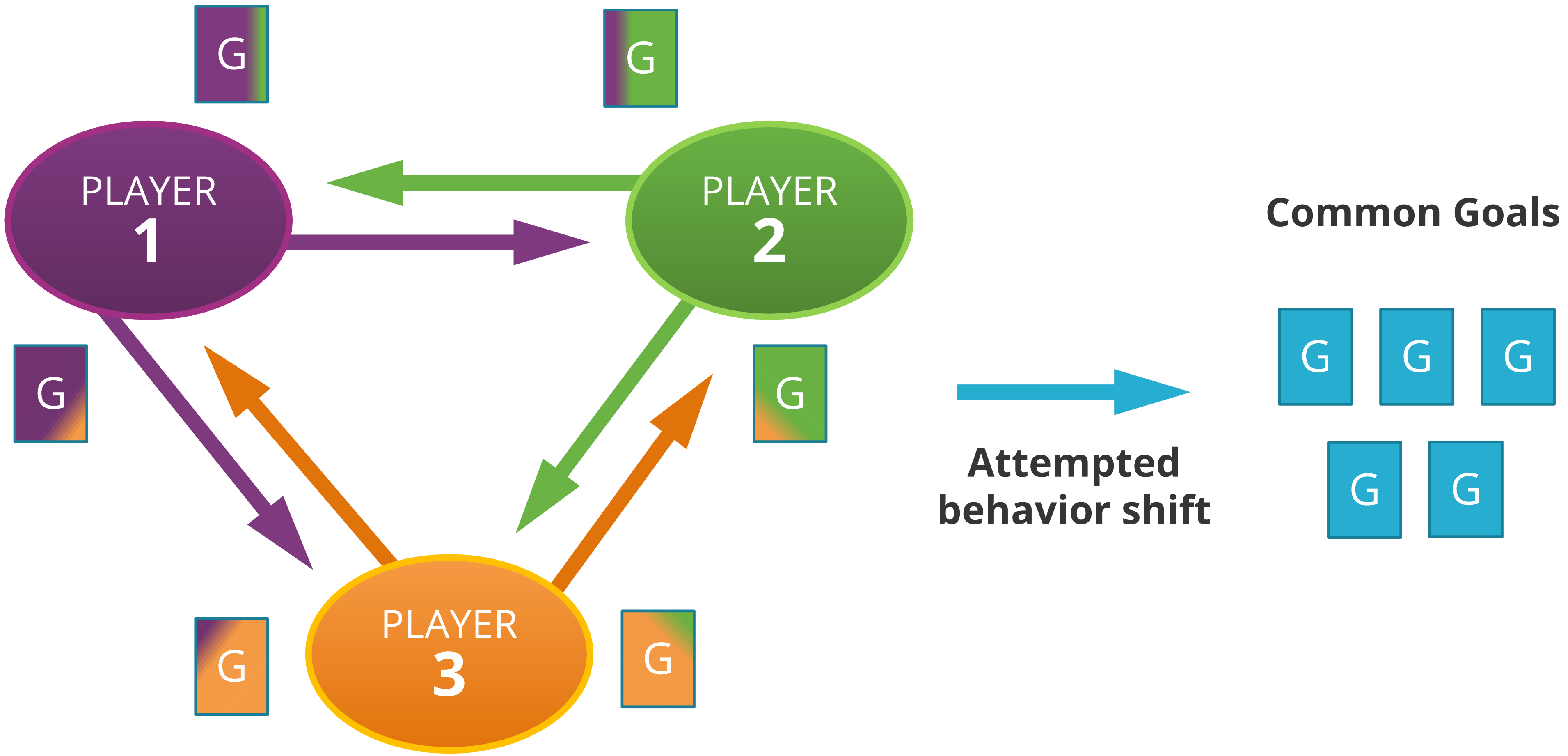
Attack you with my cyber capabilities **(Vague/Cyber)**

Attack your economic sector with a ransomware attack **(Specific/Cyber)**

USING GOALS TO INCENTIVIZE THREATS

GOALS

Player Assigned Goals



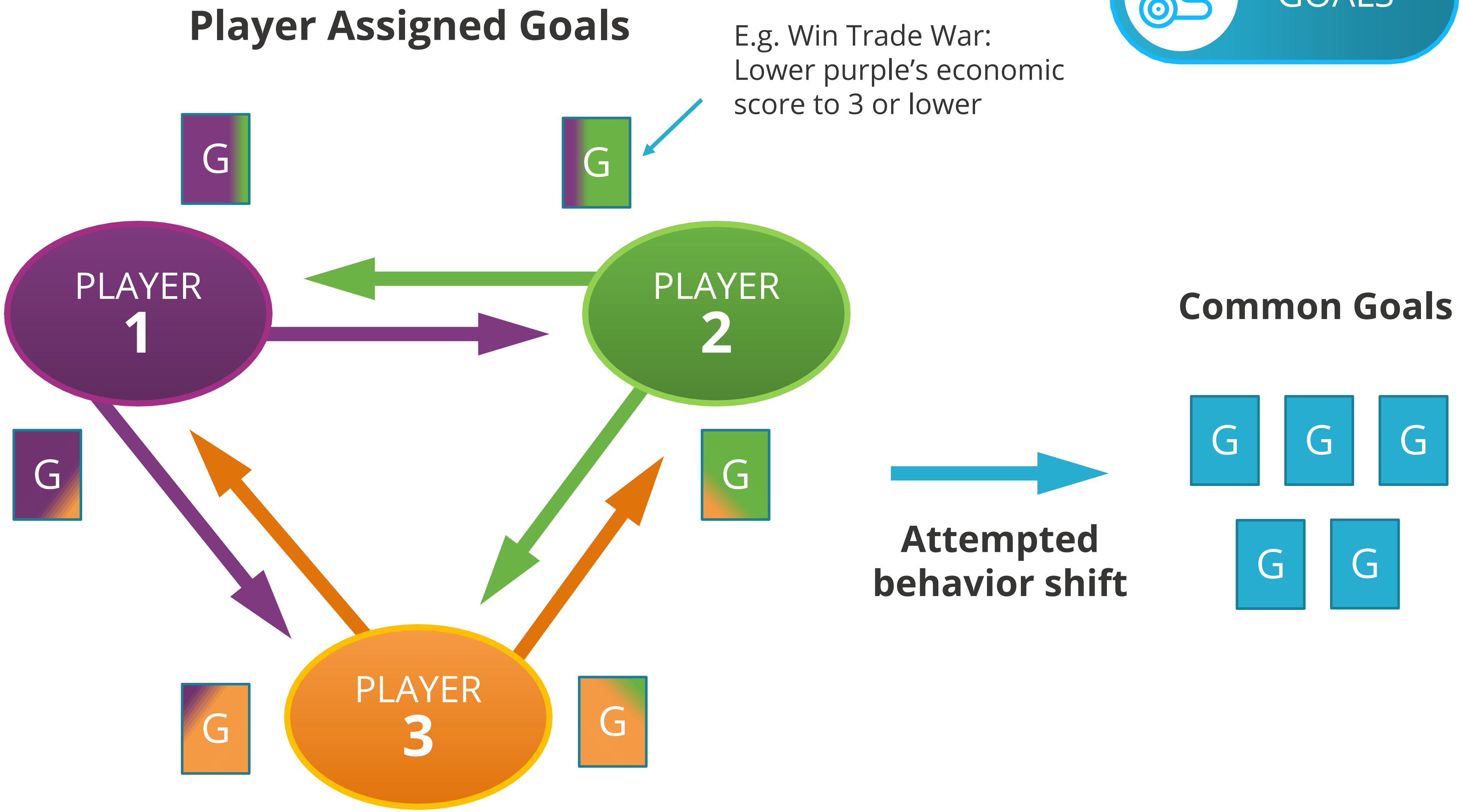
PERSONAL GOALS

GOALS

- Personal goals are achievable only by one player they are assigned to
  - This in some sense makes them more valuable as there is less competition for them
- Designed to shape the potential initial conflict
  - These goals will either:
    - Require the player to lower one or more metrics of another player
    - Require the player to take actions that have a side effect of lowering metrics of that player
- This provides a funneling effect for threats at the start of the game
  - Each player knows a small subset of actions each of their adversaries are incentivized to pursue against them
    - And due to the design, they know that allowing that player to achieve that personal goal will negatively impact their ability to compete for the public goals

USING GOALS TO INCENTIVIZE THREATS

GOALS



COMMON GOALS

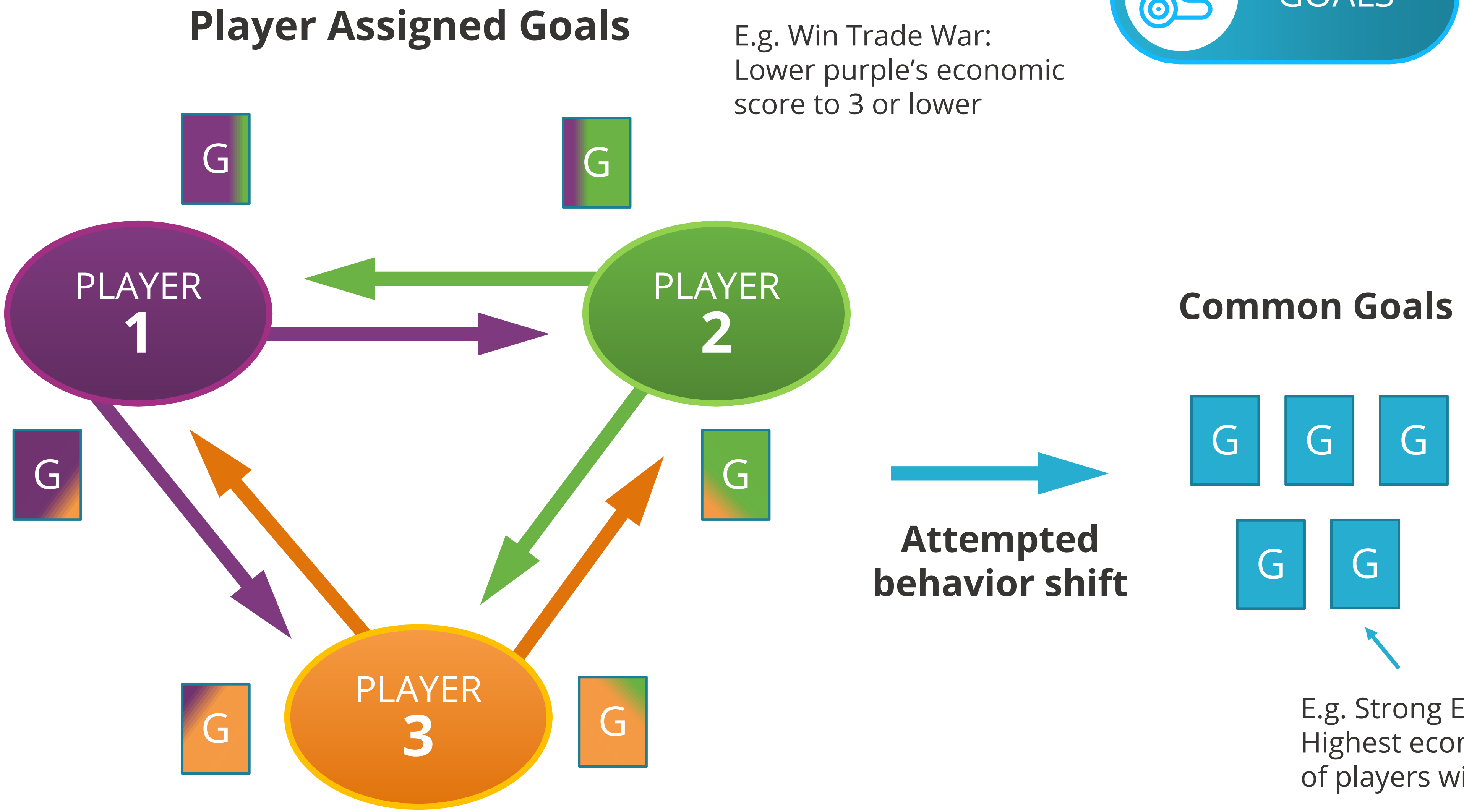
GOALS

- Common goals are designed to be less antagonistic but still competitive
  - Each has two requirements:
    - One which is a required static minimum score to qualify
      - E.g. have an economy metric of at least 7
    - The other is competitive
      - E.g. Out of all players who qualify, the one with the highest economy metric achieves this goal

- This allows us to push the players to compete, but not to destruction
  - If the players lean too heavily on the punishment strategies, then these common goals are designed such that none of the players should qualify
    - Since the personal goals are not sufficient for victory, this means that such games will likely have no winners
    - Avoids incentivizing “pyrrhic” victories

USING GOALS TO INCENTIVIZE THREATS

GOALS



E.g. Win Trade War:
Lower purple's economic score to 3 or lower

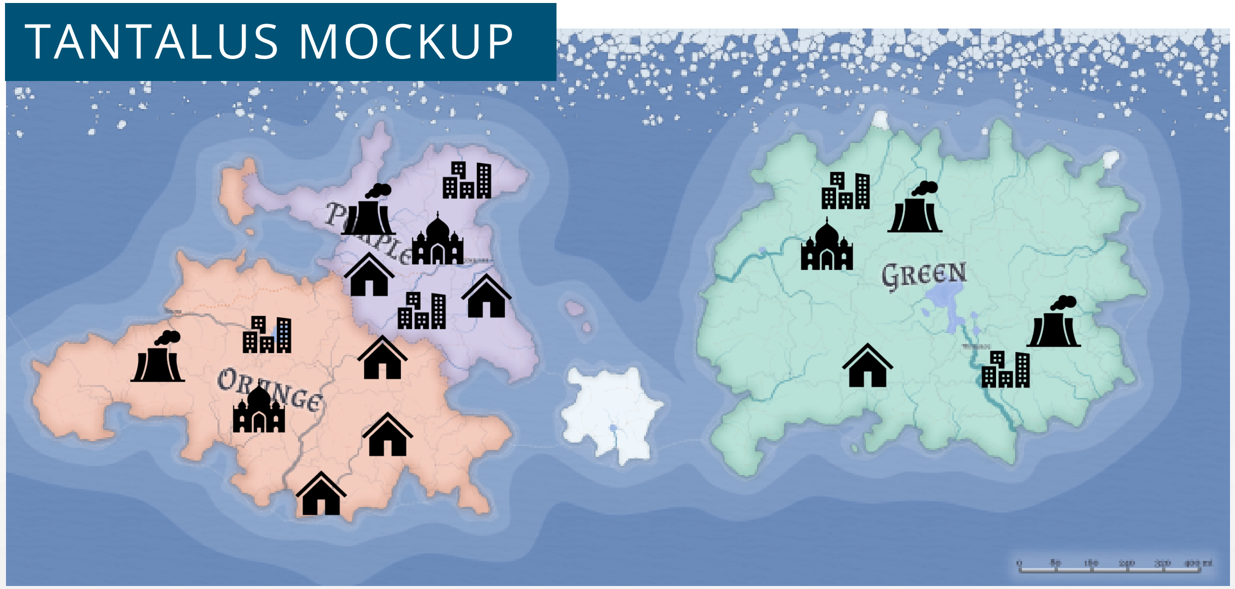
E.g. Strong Economy:
Highest economic score out of players with at least 7

ROUND 1, TIME POINT 1

After actions have been initially chosen, players are able choose if they wish to make threats:

T1, T2, T3…

Given what she knows about Green's goals, Orange chooses to make threat T2 to Green to try to prevent him from taking action A3.


TANTALUS MOCKUP

Note: this figure is intended as a mockup for illustrative purposes but will likely not resemble the final product

	0	1	2	3	4	5	6	7	8	9
M1		🟢				🟣			🟠	
M2		🟠				🟢			🟣	
M3		🟣				🟠			🟢	
M4						🟠		🟣		🟢
M5	🟢		🟣		🟠					

ACTION EXECUTION UNCERTAINTIES

ACTIONS

- Attribution
  - Requires us to introduce some level of background noise to hide player attacks in.
    - Need to be careful that this background level of attacks is not too impactful
- May want to include explicit actions that interact with the uncertainty (ability to investigate)
- May want to model attribution uncertainty as a delay before attribution
- ~~May want to misattribute some background attacks as originating from a player~~
- Effect uncertainty
  - This can be captured by adding randomization to both attack success and effect on attack success
  - We might also want to include randomization over when actions execute
    - Which also helps us with creating attribution uncertainty
- Online setting can allow us to provide information about potential outcomes

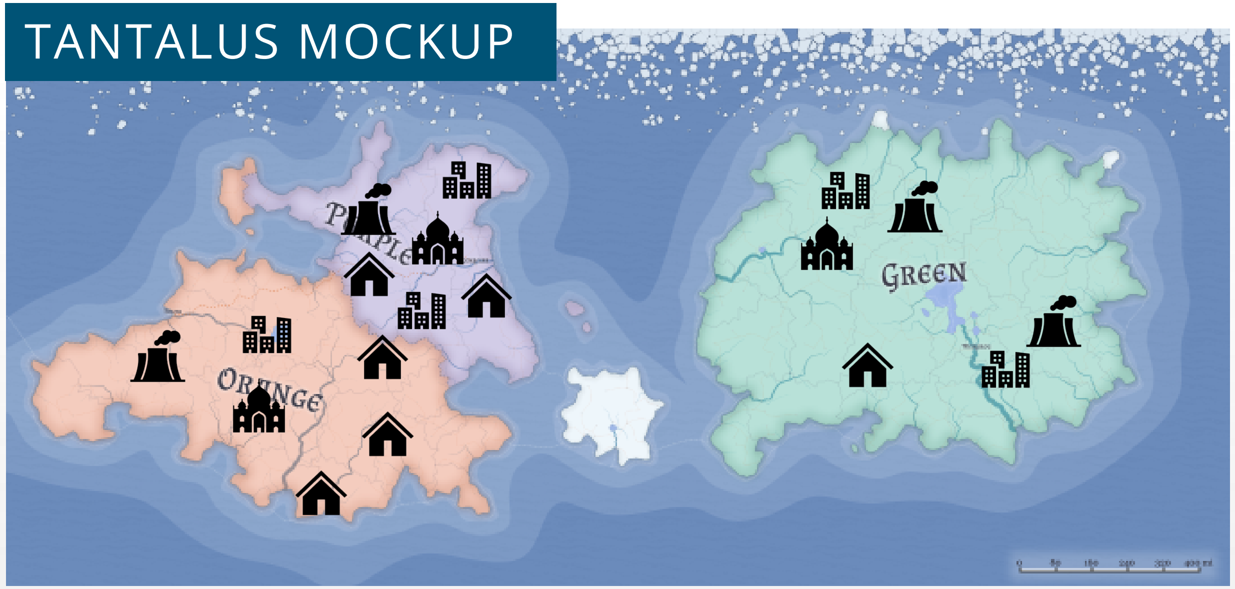
ROUND 1, POINT TIME 2

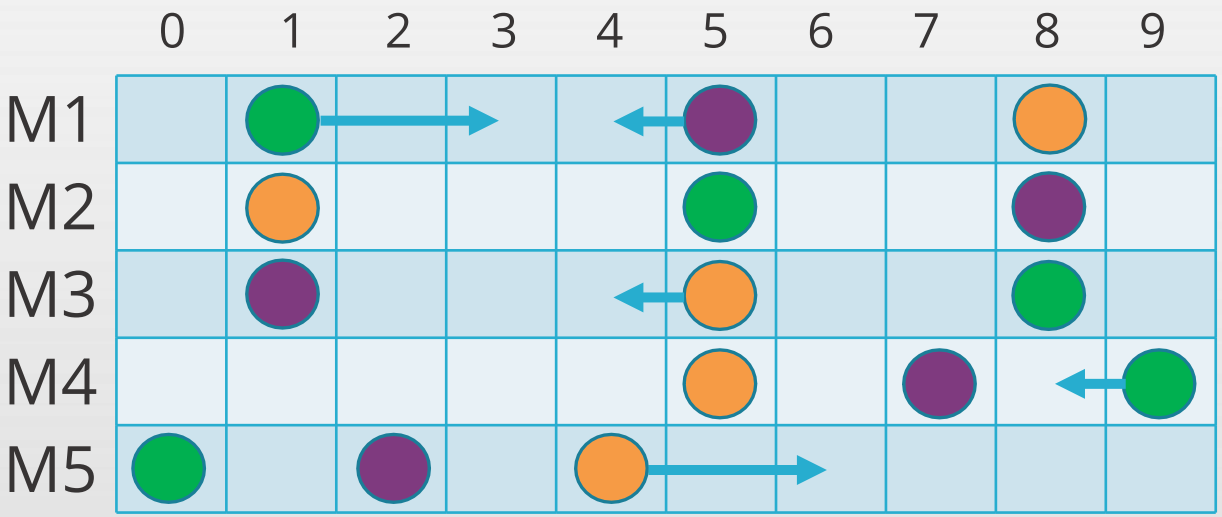
After the players have decided on their threats for the round, players are given a chance to re-evaluate their action choice for the round if they had chosen an action that another player has threatened they will respond to.

Next, after all players have finalized their actions for the round we execute these actions. This involves both calculating how the game state updates and what information is revealed to the players.

Green decides to go ahead with action A3 even with Orange's threat T2 and the game state updates. In the next round Orange will need to decide given the information received if they think that A3 happened and do they follow through on their threat?

TANTALUS MOCKUP





	0	1	2	3	4	5	6	7	8	9
M1		🟢 →				← 🟣			🟠	
M2		🟠				🟢			🟣	
M3		🟣				← 🟠			🟢	
M4						🟠		🟣		← 🟢
M5	🟢		🟣		🟠 →					

Note: this figure is intended as a mockup for illustrative purposes but will likely not resemble the final product

VICTORY CONDITIONS

VICTORY

- The RQs don’t care who wins
  - Luckily, most players do
- Thus, victory conditions are our best lever in shaping player behavior
  - Guide players to interact with and incentivize them to care about particular outcomes

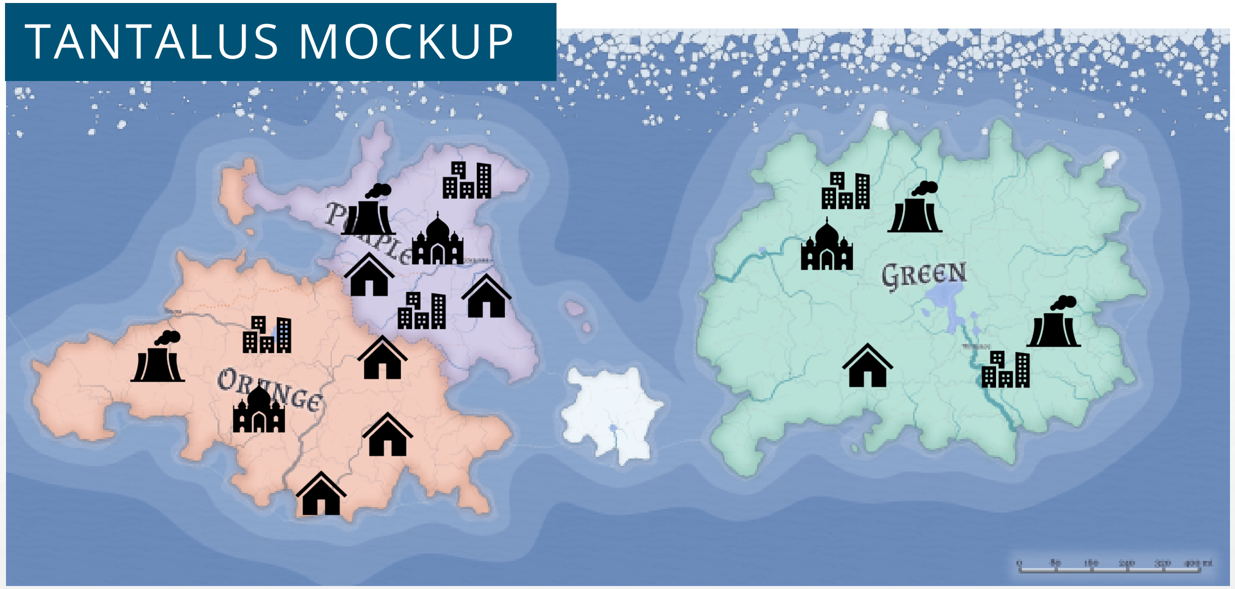
- Public victory conditions:
  - A mixed of shared and personal goals
  - Unique victory categories (player with the most in metric x wins this goal)
  - Possibility of multiple winners
    - E.g. all players who achieve victory in at least 4 goals
  - Possibility of no winners
    - Each goal has some minimal requirements to be achieved, if no player achieves victory in at least 4 categories….

ROUND 14, TIME POINT 2

After a number of rounds of play have taken place, the game is approaching its end.

After a final action execution, we will determine which players achieved which goals, and who (if any) won.

While it will depend on the exact victory conditions specified, Orange appears to be in a strong position to win this game.


TANTALUS MOCKUP

Note: this figure is intended as a mockup for illustrative purposes but will likely not resemble the final product

	0	1	2	3	4	5	6	7	8	9
M1					🟢	🟣			🟠	
M2				🟣		🟢		🟠		
M3		🟠	🟣				🟢			
M4			🟣		🟢	🟠				
M5			🟢		🟠				🟣	

CONSIDERATIONS ON NUMBER OF PLAYERS

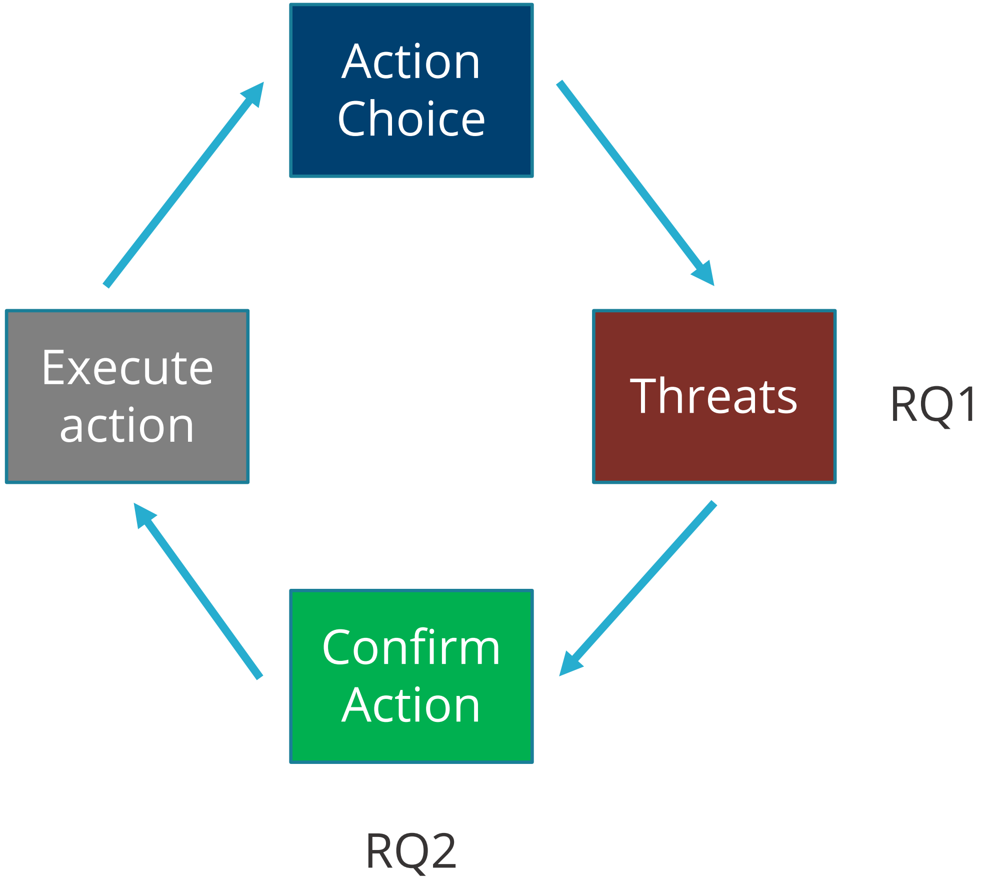
- While there are advantages to two-players
  - Data collection
    - Easier to fill games
    - More data points for the same population of players
- Recall that the primary focus here is on deterrence
  - When people attempt it
  - When it succeeds/fails
- True two-player zero sum settings are problematic
  - Deterrence relies on threat of punishment
  - In a zero sum setting, punishment strategies don't exist
  - Reality gets around this by never being truly zero-sum

- Expand the game to three players
  - While the overall game may still be zero-sum, dyadic relationships are no longer required to be zero-sum
- Allow for an all lose condition
  - E.g. MAD
- Allow for multiple winners
  - With the understanding that we don't want this to be a cooperative game where everyone can win

SUMMARY

- Threats:
  - Restricted domain on Behavior
    - Consistent across treatments
  - Restrictions on possible threats
    - Restrictions capture both domain and specificity
    - **Varies with experimental condition**
  - Personal goals used to create desire for Behaviors
    - **Creates a condition where there there is a behavior to deterred**

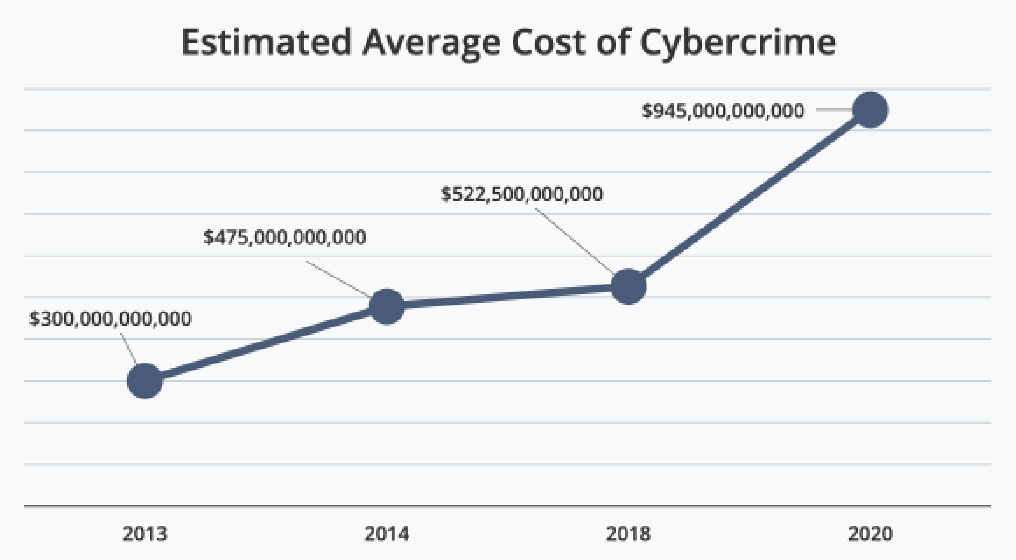
Main gameplay loop:



BACKUP SLIDES

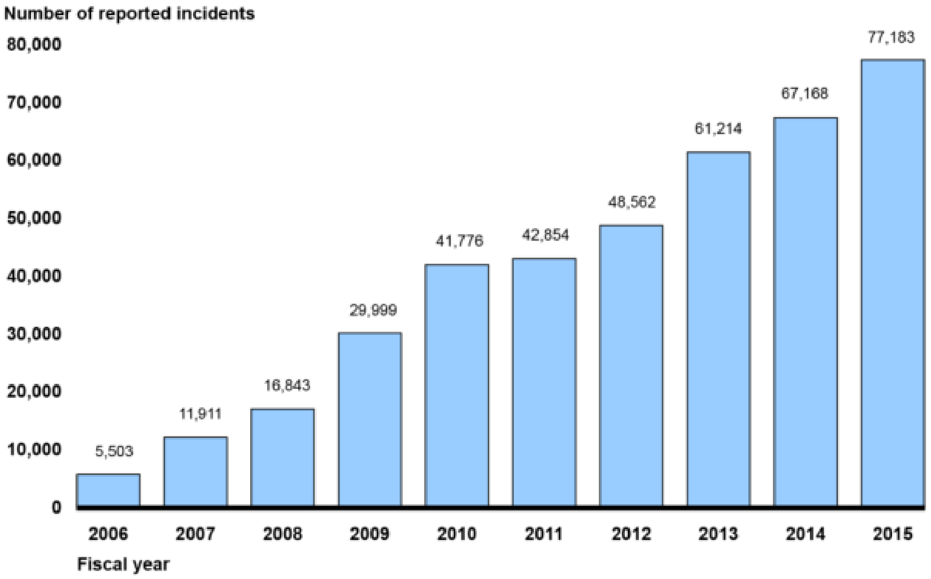
Cyber attacks are growing

- Globally, the cost of cybercrime is estimated to be nearly $1 trillion
- Most common threat actor motivation is financial gain
- The threat of cyber attacks is prevalent and growing in both the public and private sector in the US.
    - Between 2006-2015, US Federal Agencies saw an increase of about 1,303 percent of cyber incidents, based on reporting to US-CERT



https://www.mcafee.com/enterprise/en-us/assets/reports/rp-hidden-costs-of-cybercrime.pdf

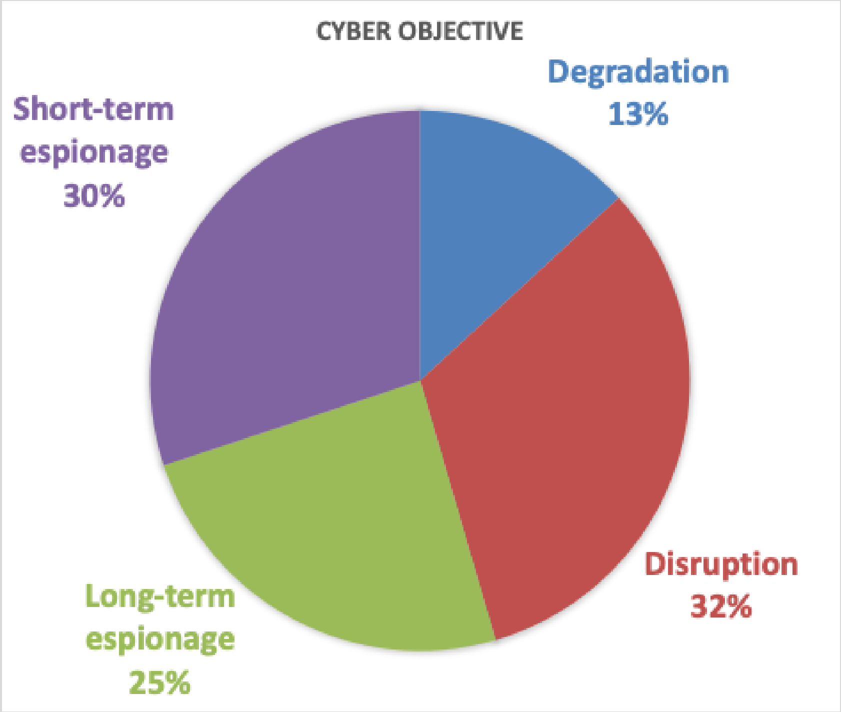
Figure 1: Incidents Reported by Federal Agencies, Fiscal Years 2006 through 2015



Source: GAO analysis of United States Computer Emergency Readiness Team and Office of Management and Budget data for fiscal years 2006-2015. | GAO-16-501

Cyber attacks can be state-sponsored

- Many states sponsor cyber operation targeting other states.
  - 34 countries suspected of sponsoring cyber operations[1].
- Attacks can vary in objective.
- Impact of state sponsored attacks can go beyond monetary value – impact national security.
- United States is a frequent target.
  - Of the 266 incidents in the DCID[2], 30% involved US as a target.



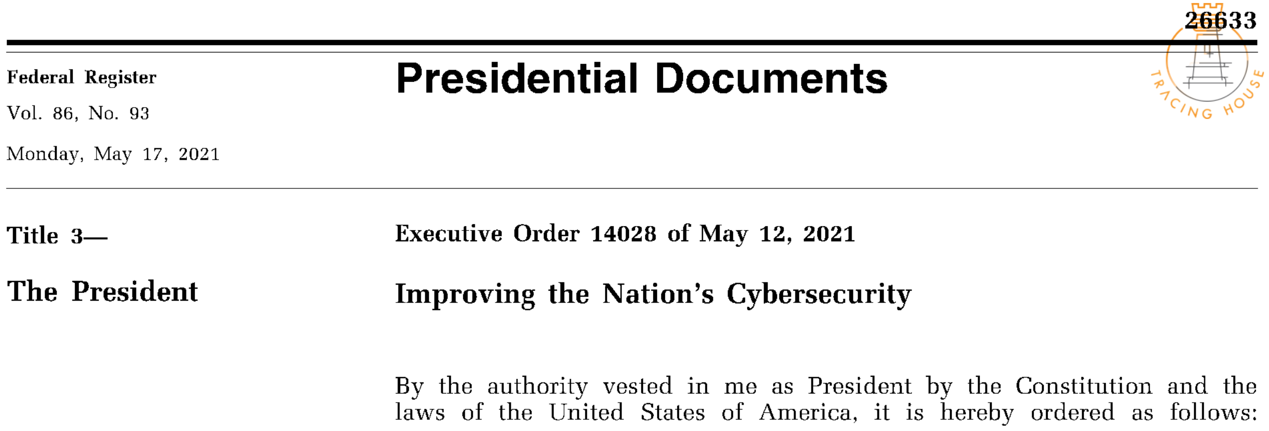
From the DCID dataset.

[1] https://www.cfr.org/cyber-operations/

[2] Valeriano, Brandon, and Ryan C Maness. 2014. "The Dynamics of Cyber Conflict between Rival Antagonists, 2001–11." *Journal of Peace Research* 51 (3): 347–60. https://doi.org/10.1177/0022343313518940 .

How do we stop these attacks?

- Cyber defenses
- Decades of research, technology, processes and policies to defend.
- But attacks still occur…
  - New technology – new vulnerabilities.
  - Addressing the human dimension.



26633

Federal Register
Vol. 86, No. 93
Monday, May 17, 2021

Presidential Documents

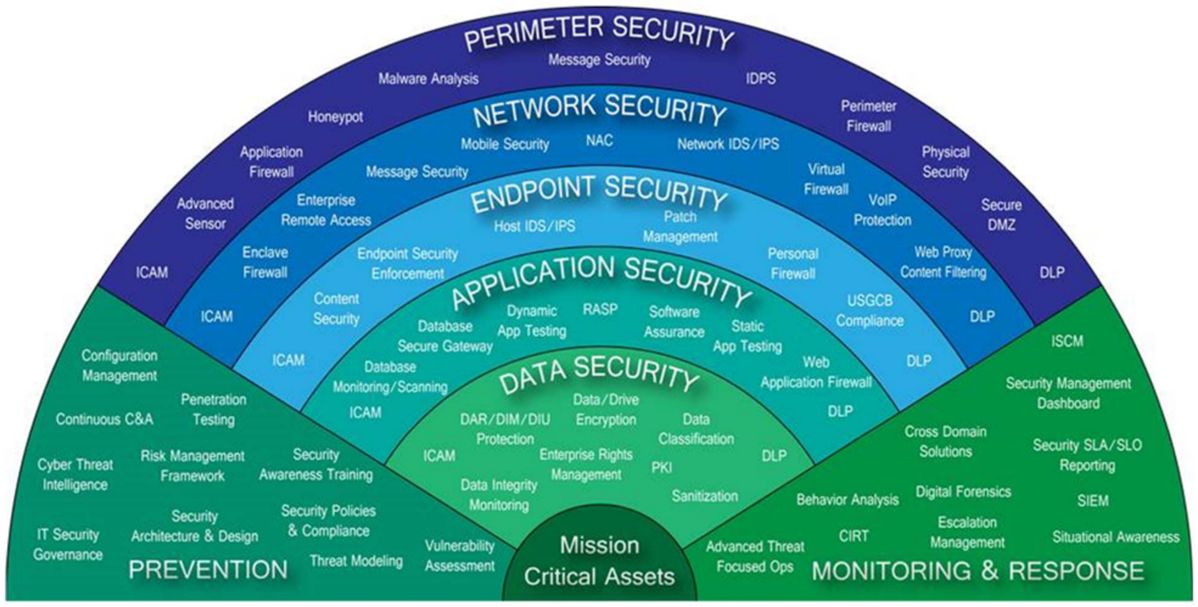
Title 3—

The President

Executive Order 14028 of May 12, 2021

Improving the Nation's Cybersecurity

By the authority vested in me as President by the Constitution and the laws of the United States of America, it is hereby ordered as follows:



Northrop Gruman Fan

Deterrer Survey

You and your colleagues are advisors to your country's national security council during an unfolding crisis. As a trusted advisor, your administration leans on you for advice.

Vignette 1: Threat to Adversary Civilian Systems [VIGNETTE ORDER RANDOMIZED]
[At the outset of the crisis,] there is increased tension between your country and a more and more aggressive rival. Your intelligence agencies tell you that your rival plans to infiltrate your country's **civilian infrastructure**. Infrastructure at risk includes the energy grid, systems that supply food, and communication networks. If your rival fulfills their plan, your **citizens** may be at risk. A number of your colleagues in government have argued that this development is a risk to your state's national security that cannot be tolerated.

Response Options: You have several tools that you can use, from conventional military actions (air power, naval assets, and soldiers) to cyber attack actions (malware, denial of service). [The options below are RANDOMIZED]

Option 1: Your country's leader can make it clear that there will be consequences if your rival carries out their attack.

Option 2: Your **conventional** forces can mobilize. These forces include air, naval, and land-based assets. Your country's leader can make it clear that these forces will be used if your rival carries out their attack.

Option 3: Your **conventional** forces can mobilize. These forces include air, naval, and land-based assets. Your country's leader can make it clear that these forces will be used against your rival's intelligence headquarters if your rival carries out their attack.

Option 4: Your **cyber** forces can mobilize. This involves preparing malware and denial of service attacks. Your country's leader can make it clear that these forces will be used if your rival carries out their attack.

Option 5: Your **cyber** forces can mobilize. This involves preparing malware and denial of service attacks. Your country's leader can make it clear that these forces will be used against your rival's intelligence headquarters if your rival carries out their attack.

Deterrer Survey

Ranking:

Of those options above, please rank them from the option that you're most likely to suggest (1) to the option that you're least likely to suggest (5)?

Logic Questions:

1. As you ranked your choices, how concerned were you that your rival might take measures to defend against your potential action? [SLIDER]

    Not concerned at all (1)

    …

    Extremely concerned (5)

2. As you ranked your choices, how concerned were you that your rival might take the same actions against your own country? [SLIDER]

    Not concerned at all (1)

    …

Extremely concerned (5)

Hypotheses (Collected)

Specificity matters

H1A: If Blue's available deterrent threats are **specific**, then attempts to deter occur **less**, all else equal.
H1B: If Blue's available deterrent threats are **specific**, then there is **no effect** on attempts to deter.
H1C: If Blue's available deterrent threats are **specific**, then attempts to deter occur **more**, all else equal.

H2A: If Blue's available deterrent threats are **specific**, then deterrence failure is **more** likely, all else equal.
H2B: If Blue's available deterrent threats are **specific**, then deterrence failure is **no more** likely, all else equal.
H2C: If Blue's available deterrent threats are **specific**, then deterrence failure is **less** likely, all else equal.

Domain matters

H3A: If Blue's available deterrent threats are **cyber**, then attempts to deter occur **less**, all else equal.
H3B: There is no delta in the probability of deterrence attempt between cyber and non-cyber deterrent threats
H3C: If Blue's available deterrent threats are **cyber**, then attempts to deter occur **more**, all else equal.

H4A: If Blue's available deterrent threats are **cyber**, then deterrence failure is **more** likely, all else equal.
H4B: There is no delta in the probability of deterrence failure between cyber and non-cyber deterrent threats
H4C: If Blue's available deterrent threats are **cyber**, then deterrence failure is **less** likely, all else equal.

Deterree Experiment

You are an advisor to your country's national security council during an unfolding crisis. As a trusted advisor, your administration leans on you for policy advice as it makes crisis decisions.

Vignette 1: Civilian Assets [RANDOMIZED VIGNETTE ORDER]
[At the outset of the crisis] there is tension between your country and a more and more aggressive rival, your intelligence agencies tell you that your rival has learned of your routine intelligence gathering activities focused on their **civilian** infrastructure. These activities focus on monitoring the energy grid, systems that supply food, and communication networks. Your intelligence agencies argue that these activities are essential to providing a clear picture of your rival's capabilities.

Please rank the following recommendations. [RANDOMIZED ORDER]

Suspend your intelligence gathering activities altogether.
Decrease your intelligence gathering activities.
Continue your intelligence gathering activities at the same level.
Increase your intelligence gathering activities.

Deterree Experimental Treatment

Respondent randomly receives ONE of the treatments below (this assignment is consistent across vignettes):

Treatment 1: Your rival's leader makes clear that there will be consequences if you continue your intelligence gathering activities.

Treatment 2: Your rival mobilizes **conventional** forces. These forces include air, naval, and land-based assets. Your rival's leader makes clear that these forces will be used if you continue your intelligence gathering activities.

Treatment 3: Your rival mobilizes **conventional** forces. These forces include air, naval, and land-based assets. Your rival's leader makes clear that these forces will be used against your intelligence headquarters if you continue your intelligence gathering activities.

Treatment 4: Your rival mobilizes **cyber** forces. Your intelligences agencies suggest that you prepare for your rival to use malware and zero-day vulnerabilities to target your systems. Your rival's leader makes clear that these forces will be used if you continue your intelligence gathering activities.

Treatment 5: Your rival mobilizes **cyber** forces. Your intelligences agencies suggest that you prepare for your rival to use malware and zero-day vulnerabilities to target your systems. Your rival's leader makes clear that these forces will be used against your intelligence headquarters if you continue your intelligence gathering activities.

Deterree Experiment DVs and Logic

Given the actions of your rival, please rank the following recommendations.
[RANDOMIZED ORDER]

- Suspend your intelligence gathering activities altogether.
- Decrease your intelligence gathering activities.
- Continue your intelligence gathering activities at the same level.
- Increase your intelligence gathering activities.

Logic Questions:

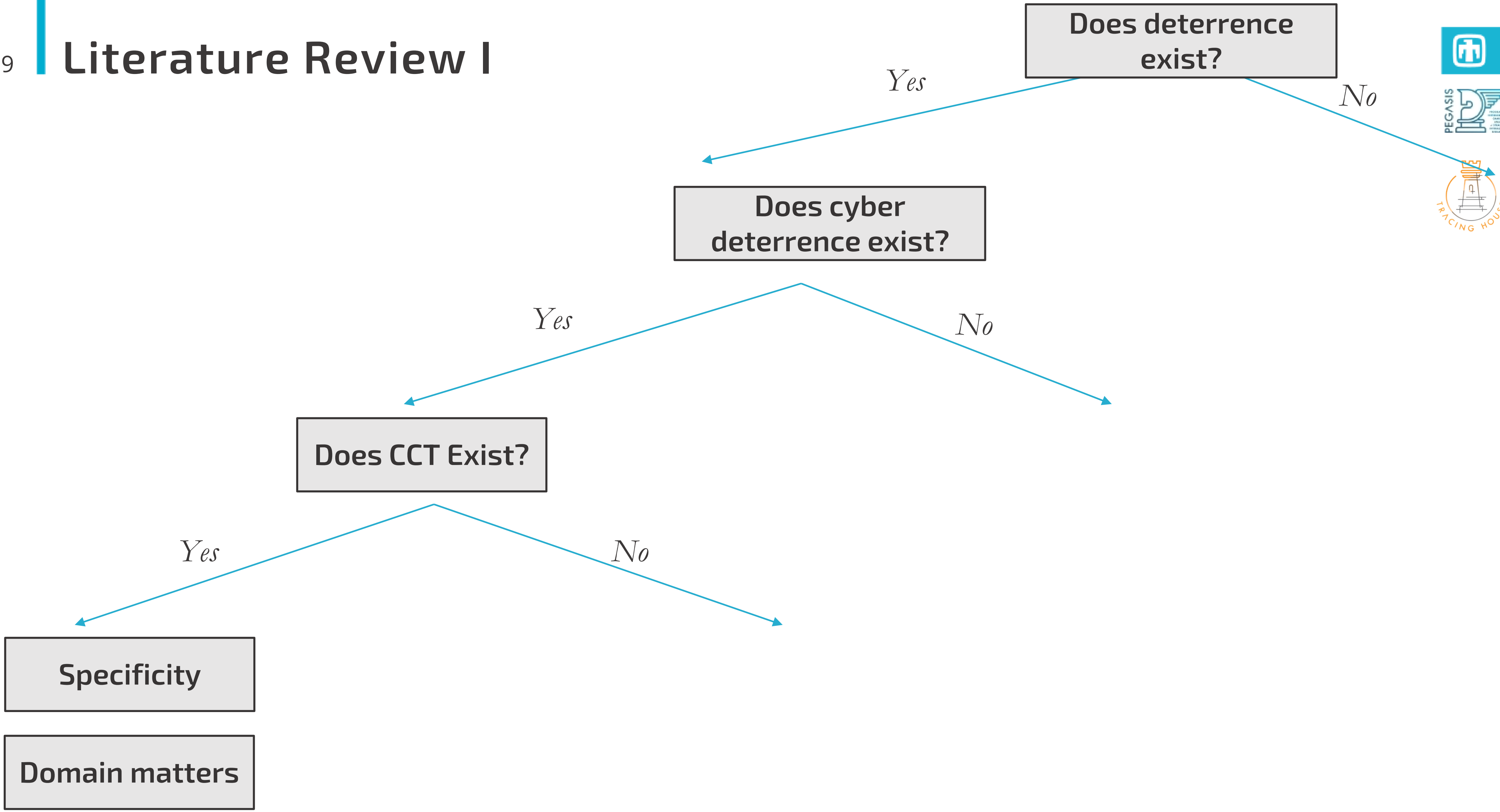
As you ranked your choices, how confident are you that your own forces can defend against the threatened action?

Not at all confident (1)
…
Extremely confident (5)

As you ranked your choices, how confident are you that your own forces can threaten the same type of action?

Not at all confident (1)
…
Extremely confident (5)

Literature Review I



Does deterrence exist?

Yes

No

Does cyber deterrence exist?

Yes

No

Does CCT Exist?

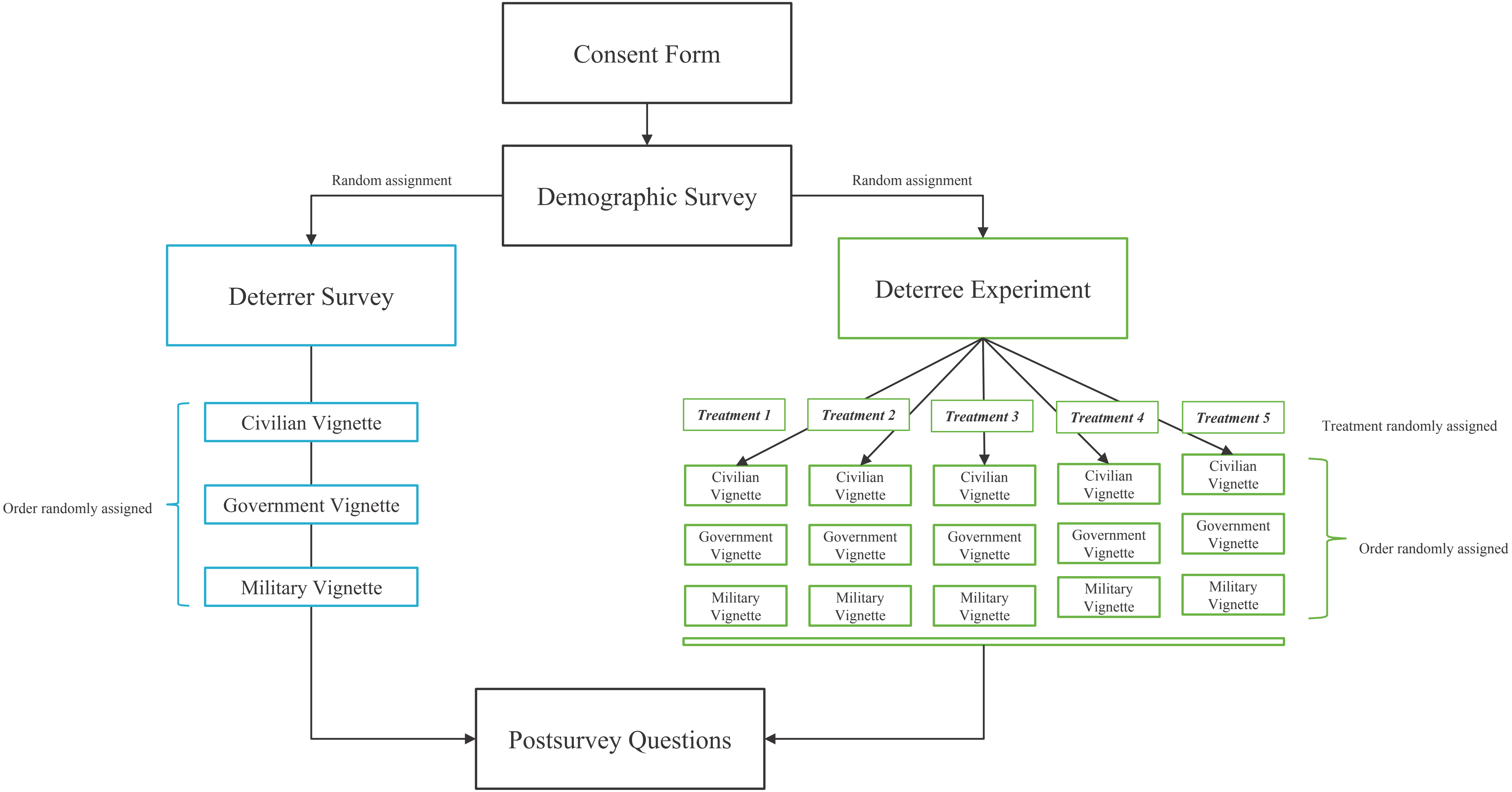
Yes

No

Specificity

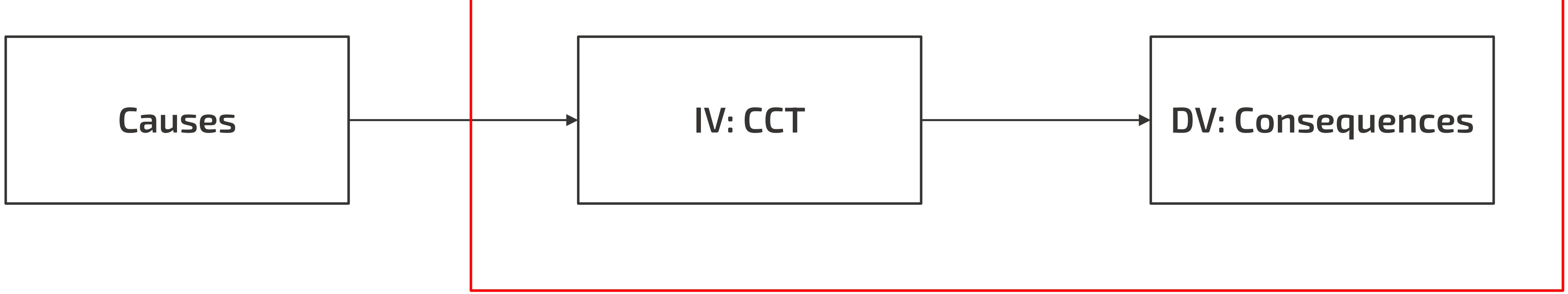
Domain matters

Survey Experiment Design



Research Question

- To get at the challenge posed by the theorized CCT, we ask:
- **How does the variation in the *domain* and *scope* of a deterrent threat affect deterrence?**
  - Does this variation affect the likelihood of making a deterrent threat?
  - Does this variation affect the likelihood of deterrence failure?

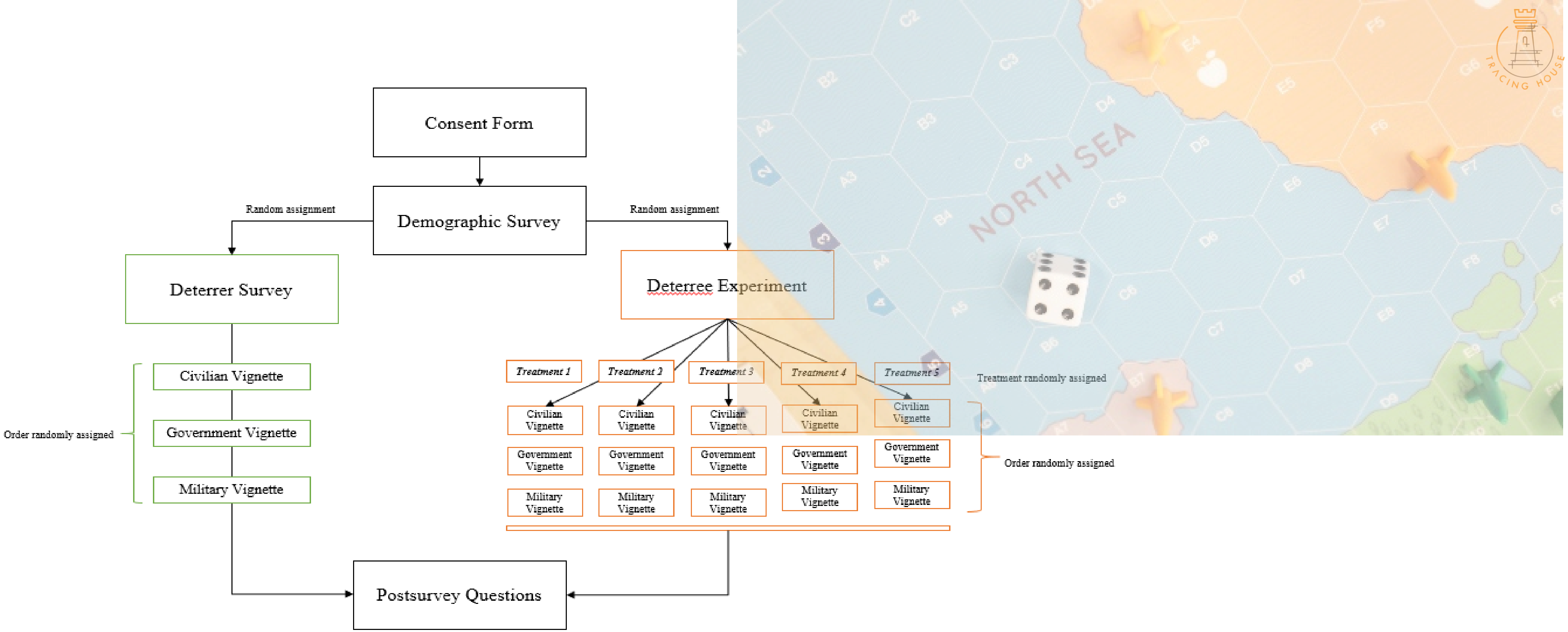


Causes

IV: CCT

DV: Consequences

Note: This RD seeks out the theorized effects of the CCT



Consent Form

Demographic Survey

Random assignment

Random assignment

Deterrer Survey

Deterree Experiment

Order randomly assigned

Civilian Vignette

Government Vignette

Military Vignette

Treatment 1

Treatment 2

Treatment 3

Treatment 4

Treatment 5

Treatment randomly assigned

Treatment 1	*Treatment 2*	*Treatment 3*	*Treatment 4*	*Treatment 5*
Civilian Vignette	Civilian Vignette	Civilian Vignette	Civilian Vignette	Civilian Vignette
Government Vignette	Government Vignette	Government Vignette	Government Vignette	Government Vignette
Military Vignette	Military Vignette	Military Vignette	Military Vignette	Military Vignette

Order randomly assigned

Postsurvey Questions

Cyber Proxy Actors: Conceptualizing the Problem

- **Cyber proxies**:
  - Non-state actor(s) that conduct offensive cyber operations to achieve objectives on behalf or with the blessing of a patron state, in exchange for political, financial, or logistical support. (Maurer, 2018)
- **Representative cases**:
  - Syrian Electronic Army; organized cyber crime toleration in China and Russia; Russo-Ukrainian War
- **Variation in state-proxy relationships:**
  - Contractualized Delegation
  - Active Orchestration
  - Passive Orchestration
  - Sanctioning
- **Why should we care?**
  - Non-state malevolent behavior in cyberspace remains widespread, with low barriers to entry and exit, presenting plenty of potential proxy actors for states.
  - Cyber proxies compound challenges of deterrence in cyberspace by introducing the complication of credibly threatening non-state actors and plausible deniability for states.
  - Cyber proxy activity and sophistication is on the rise.

Guiding Questions and Goals

1. **How and why do states use proxy actors in the cyber domain?**
   - Do state-proxy relationships vary in type? How so?
   - Why do some states build close relationships with some proxies and distant relationships with others? Are certain states more likely to pursue some proxy arrangements over others?
   - Are certain types of cyber proxies more or less likely or effective under certain state-proxy relationship structures?

2. **What is the strategic logic of employing proxies in cyberspace and how can we test it? Under what conditions are cyber proxies effective?**
   - What does an effective/ineffective state-cyber proxy relationship look like?
   - Does the use of cyber proxies influence the likelihood of escalation or deterrent failure?
   - Do proxies interact with the research questions outlined in TH?
   - Can proxies be captured in an experimental context (either through experimental surveys or war games)?

Cyber Proxies and Tracing House

- **While proxies may be relevant to decisions related to issuing deterrence threats, the problem of attribution, and the character of threat issued, the usage of proxies does not *matter* for the TH research question:**
  - The presence of proxies does not interact significantly with the two components of the CCT independent variable (specificity of deterrent threat and domain of the threat).
  - The use of proxies may contribute to attribution problems that complicate the decision to issue threats, but should not meaningfully influence a state's propensity to vary the specificity of deterrence threats once they decide to do so.
  - States that utilize cyber proxies may use these capabilities against civilian, government, or military infrastructure targets.
  - Cyber proxy use likely appears to be more frequent due to selection effects

- **Cyber proxies and the cyber deterrence problem**
  - Proxies are relevant to the general challenge of deterrence, resembling a category of offensive action in cyberspace that is, itself, difficult and unique as a deterrence problem.
  - More critical for questions relating to attribution (proxies are, by construction, a more complex attribution task), generalized deterrence problems, or escalation management.

Tradeoffs in Excluding Proxies

- **Core tradeoff between game simplicity and contextual realism.**
  - Incorporating cyber proxies to increase contextual realism would complicate the game design and mechanics.
- **What is lost by excluding proxy actors?**
  - Certain high-profile cases of offensive cyber campaigns involve the state use of proxy actors (e.g. Fancy Bear and Cozy Bear APT activity may not be approximated by this particular game design.)
  - Excluding proxy actors from the game and its discussion removes an important way in which non-state actors ‘matter’ and are used in cyberspace.
  - Excluding proxy actors removes a mechanism through which states seek to generate attribution ambiguity and leverage plausible deniability when exerting power in cyberspace.

Selected Bibliography

- Yaacov Bar-Siman-Tov, “The Strategy of War by Proxy,” Cooperation and Conflict (1984).
- Erica Borghard and Shawn Lonergan, “Can States Calculate the Risks of Using Cyber Proxies?,” *Foreign Policy Research institute*, May 7 2016.
- Erica Borghard, *Friends with benefits? Power and Influence in Proxy Warfare*, Columbia UP (2014).
- Daniel Byman, “Why Engage in Proxy War? A State’s Perspective,” Brookings, May 21, 2018.
- Jamie Collier, “Proxy Actors in the Cyber Domain,” St. Antony’s International Review (2013).
- Tim Maurer, “Cyber Proxies and their Implications for Liberal Democracies,” *The Washington Quarterly* (2018).
- Tim Maurer, *Cyber Mercenaries: The State, Hackers, and Power*, Cambridge UP (2018).
- Michael Poznansky, “Revisiting Plausible Deniability,” Journal of Strategic Studies (2020).
- Idean Salehyan, “The Delegation of War to Rebel Organizations,” Journal of Conflict Resolution (2010).