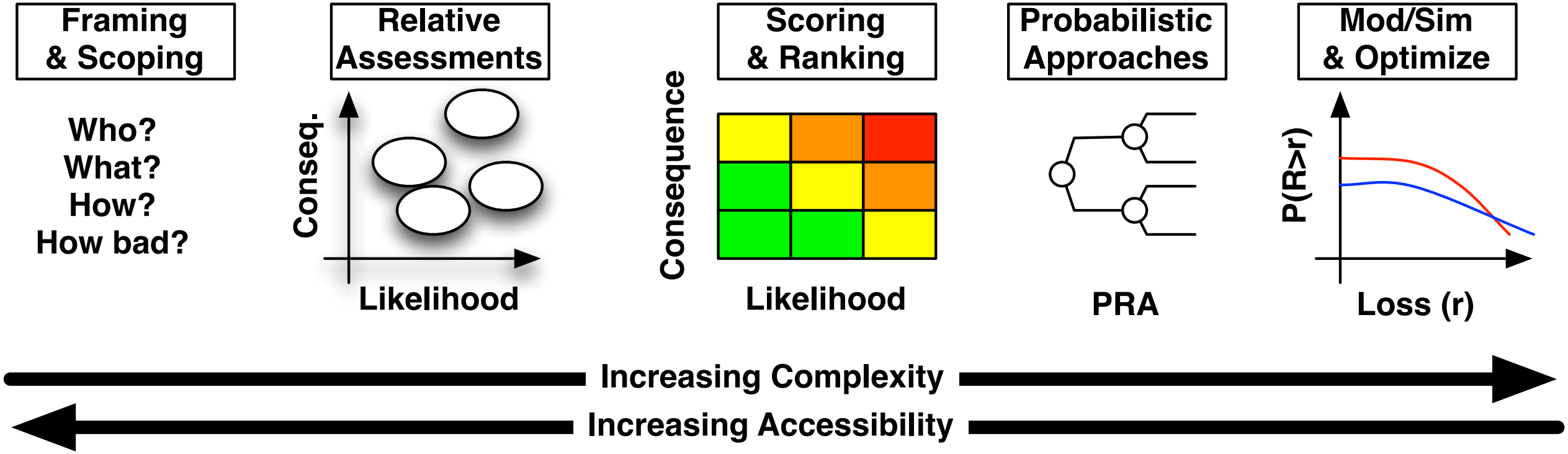
SAND2017-2408C

Modeling Adversaries and Strategic Stability

Jason Reinhardt
jcreinh@sandia.gov

15th PIIC Conference
Suzhou, China
November 2-5, 2016

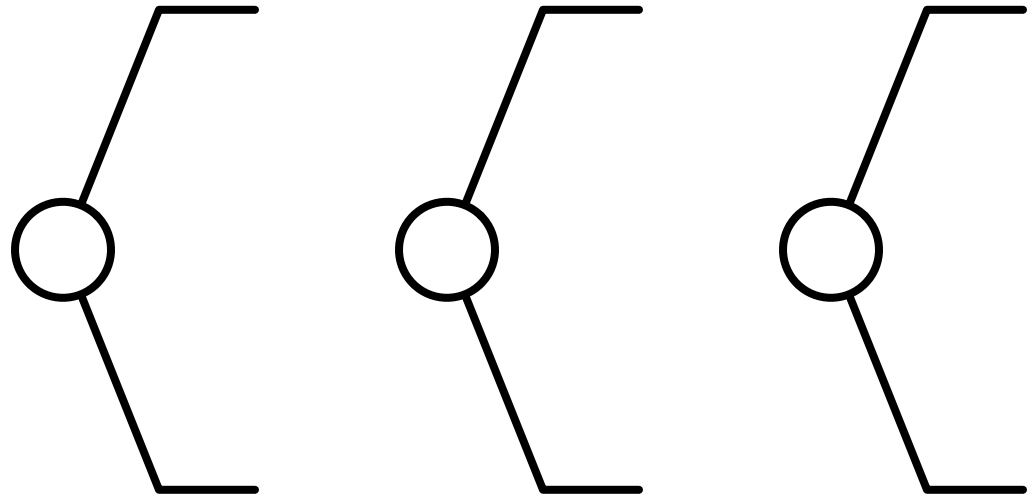
There is a broad family of risk analysis tools.



For an alternate treatment of the right end of this spectrum, see Paté-Cornell, "*Uncertainties in Risk Analysis: Six Levels of Treatment.*"

The Most Common “Risk” Approach

“Risk” =

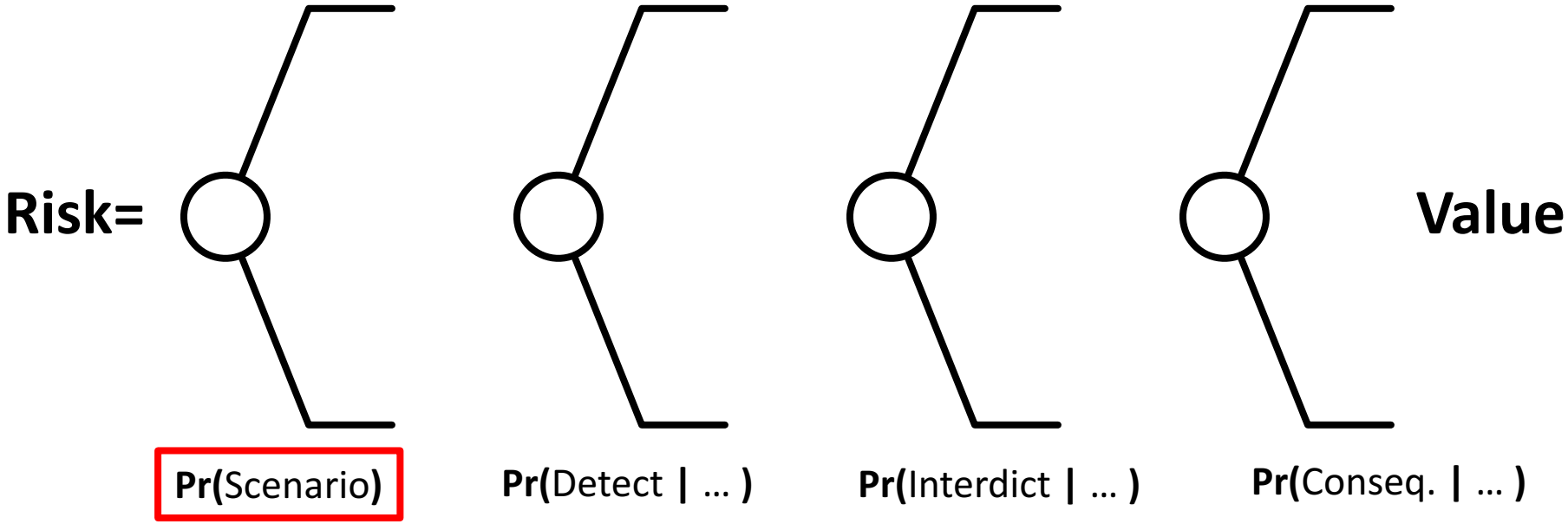


Value

Pr(Detect **| …)** **Pr(**Interdict **| …)** **Pr(**Conseq. **| …)**

- Can be done qualitatively or quantitatively
- Only provides an estimate of risk conditioned on the assumption that the adversary WILL attack
- Does not allow for deterrence, threat shifting, or other effects of strategic interaction

A full risk analytic approach focuses the adversary modeling problem.



- The adversary chooses the scenario:
  - Whether to attack, When to attack, How to attack
- Adversary choices influence the whole model structure
- Requires to estimate some form of probability distribution over the scenario space

The Common Methods

- Non-Probability Based
  - ***Red Team***
  - Adversary Capability Levels
  - Systematic Difficulty

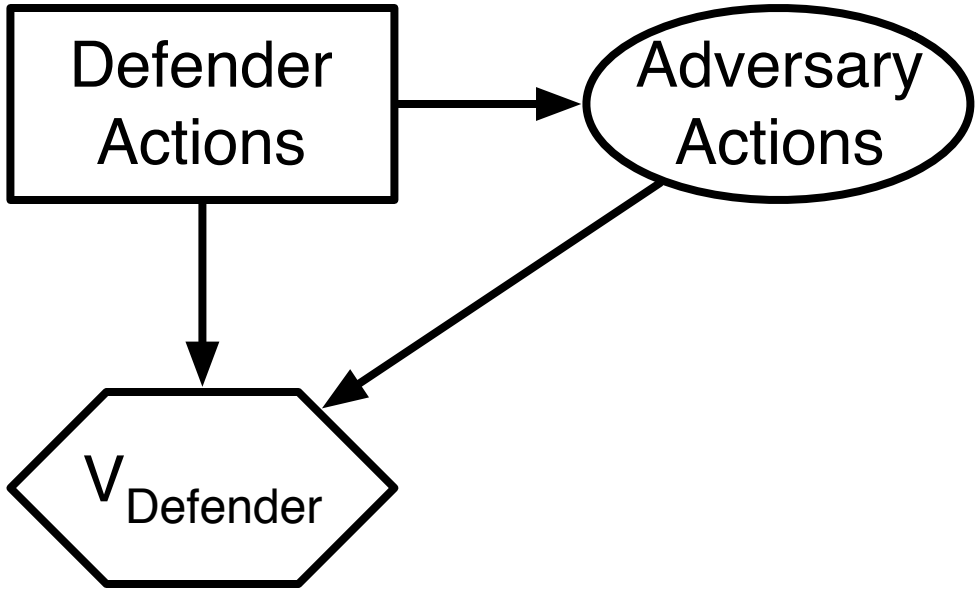
- Probability Based
  - ***Direct Elicitation***
  - Empirical Assessment
  - ***Game Theoretic***

The Red Team (or Blue Team?)

- Assumptions
  - ‘Red Team’ thinks like adversary
  - Scenario examined is likely
- Pros
  - Rich adversary model, creative, adaptive, experienced
- Cons
  - Expensive to run (little to no sensitivity analysis)
  - Difficult to replicate results
  - Difficult to be systematic

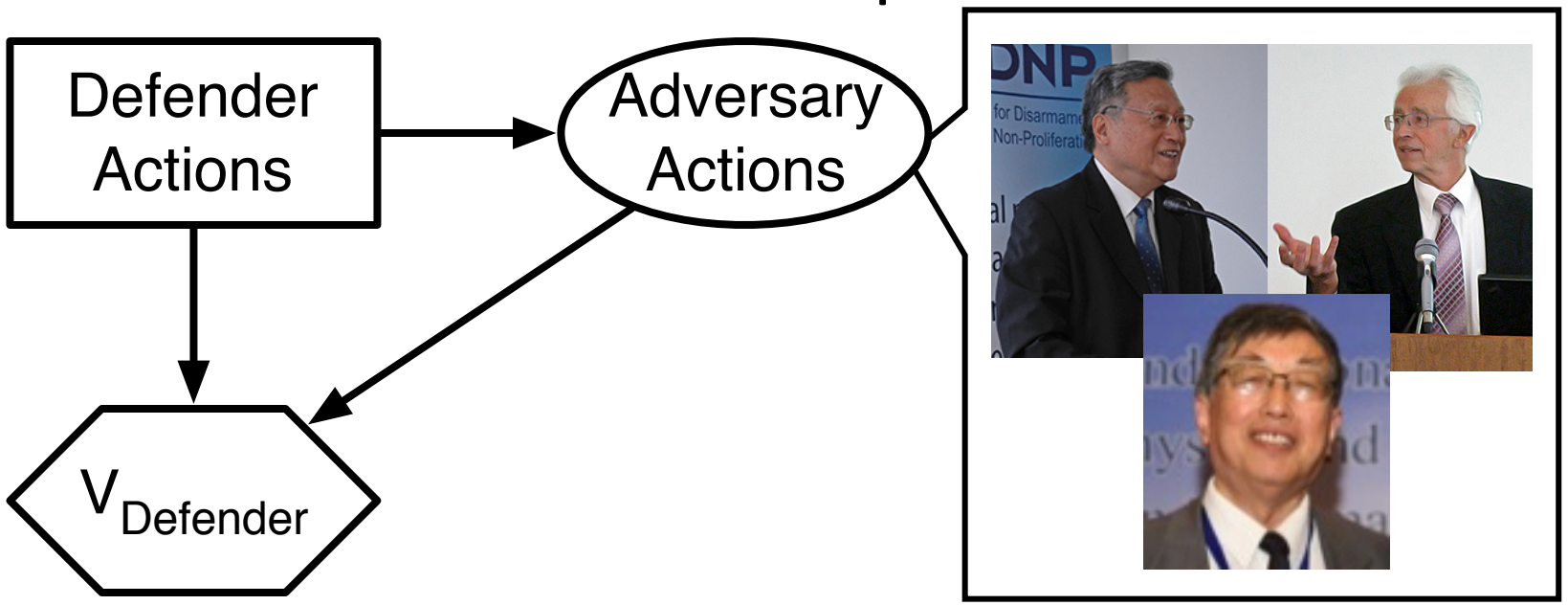
Can be useful for exposing issues and vulnerabilities, but conclusions regarding system performance are generally suggestive at best.

Direct Elicitation: Ask an Expert!



- Relies on expert inputs on adversary characteristics, reactions, and decisions
- Requires careful elicitation techniques to account for biases
- Formalizes expert inputs in the framework of probability
- Utilizes risk and decision analysis approaches to inform architecture decisions

Direct Elicitation: Ask an Expert!



- Relies on expert inputs on adversary characteristics, reactions, and decisions
- Requires careful elicitation techniques to account for biases
- Formalizes expert inputs in the framework of probability
- Utilizes risk and decision analysis approaches to inform architecture decisions

Direct Elicitation: Assessment

- Assumptions
  - Experts understand adversaries well
  - Adversaries do not adapt, or adapt in simple ways
- Pros
  - Allows for the formulation of quantitative model
  - Enables early phase sensitivity and trade-off analysis
  - Most useful when combined with physics models
- Cons
  - Highly dependent on expert opinions, and can include their biases
  - Repeated elicitations may be required as analysis proceeds
  - Costly to perform assessments over large sets of possible scenarios

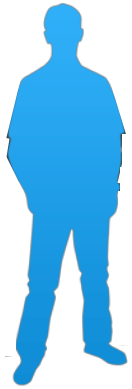
It is important to remember the difference between two things:

Behavior of Groups

- Aggregate behaviors
- Markets, crime, politics
- “What is the probability that SOME person will…?”

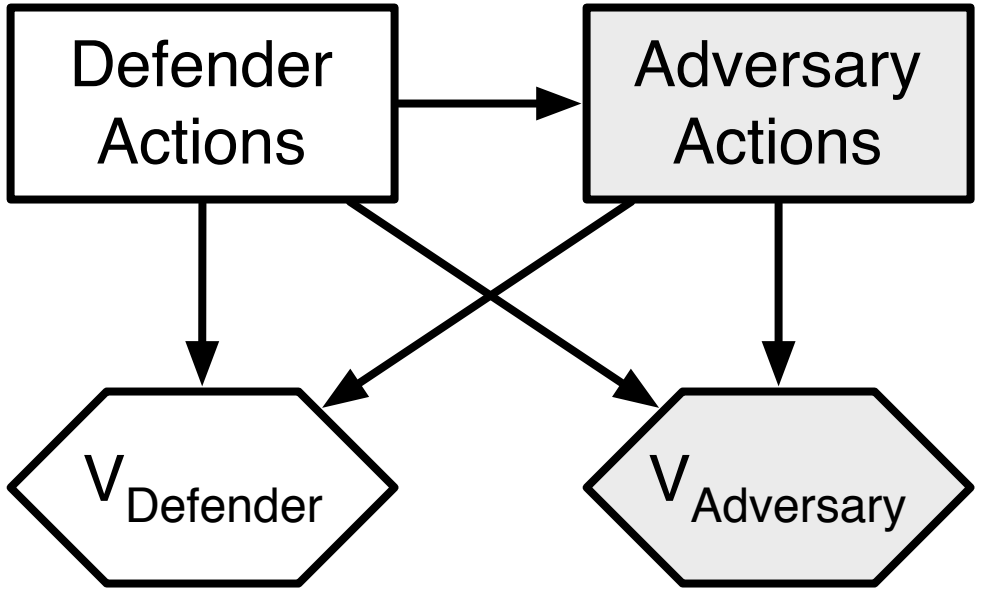
Behavior of Individuals

- Choices in strategic interaction and conflict
- “What is the probability that THIS person will…?”



Game Theoretic Methods

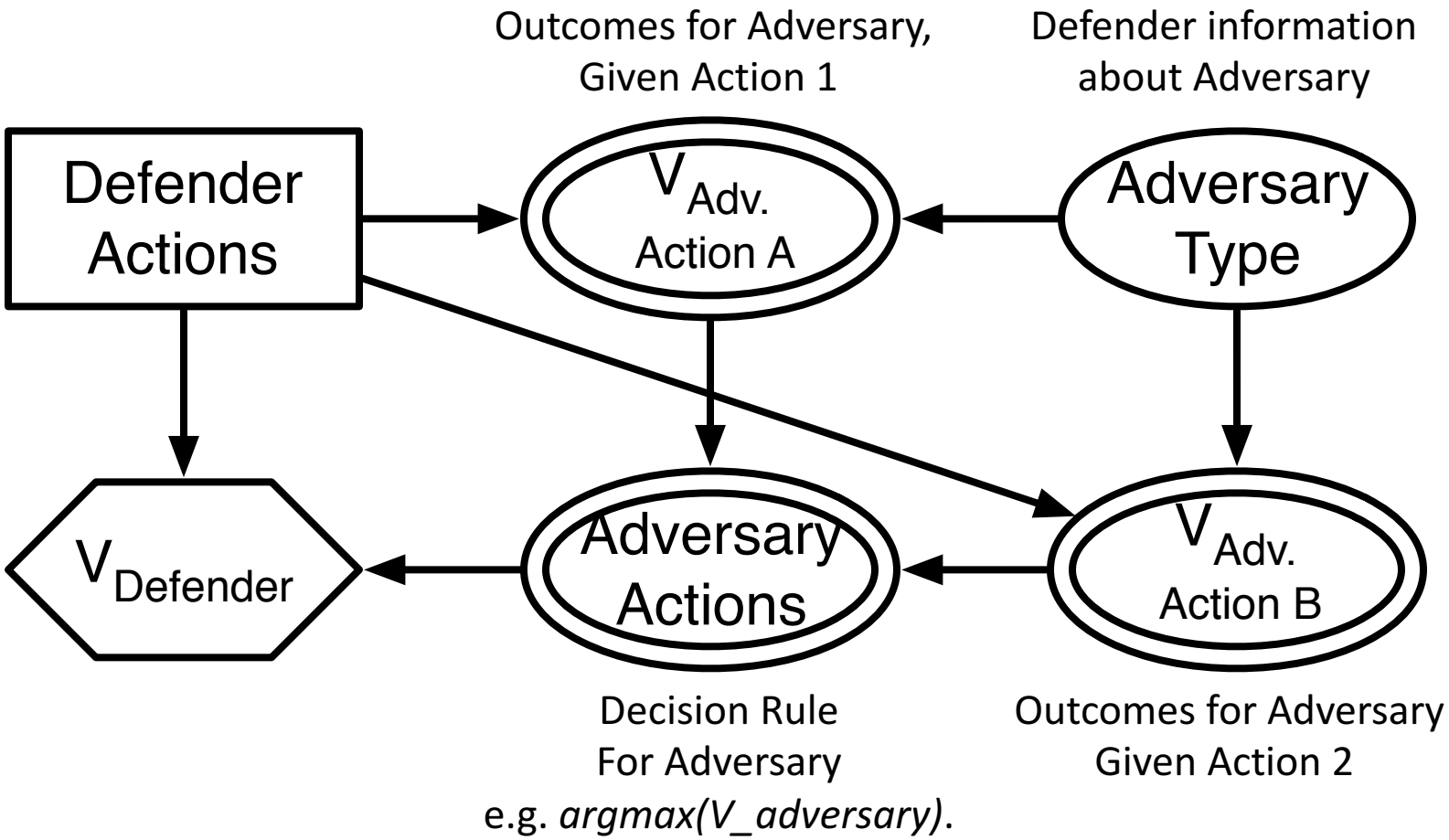
Model the decisions of the defender AND the decisions of the adversary.



But, how do we model the adversary's decisions?

Game Theoretic Method Models

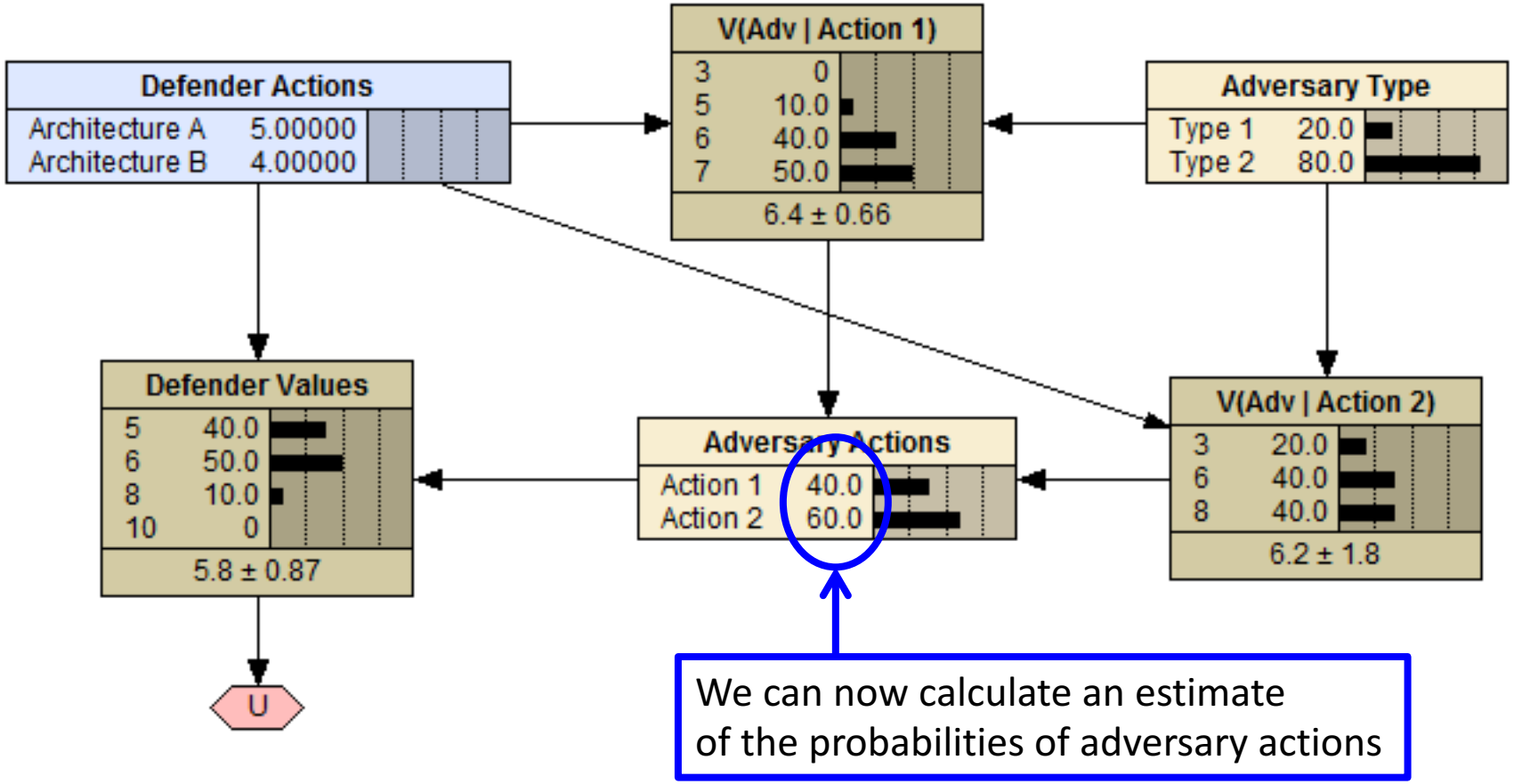
The “Adversarial Risk Analysis” technique systematically constructs an adversary model.



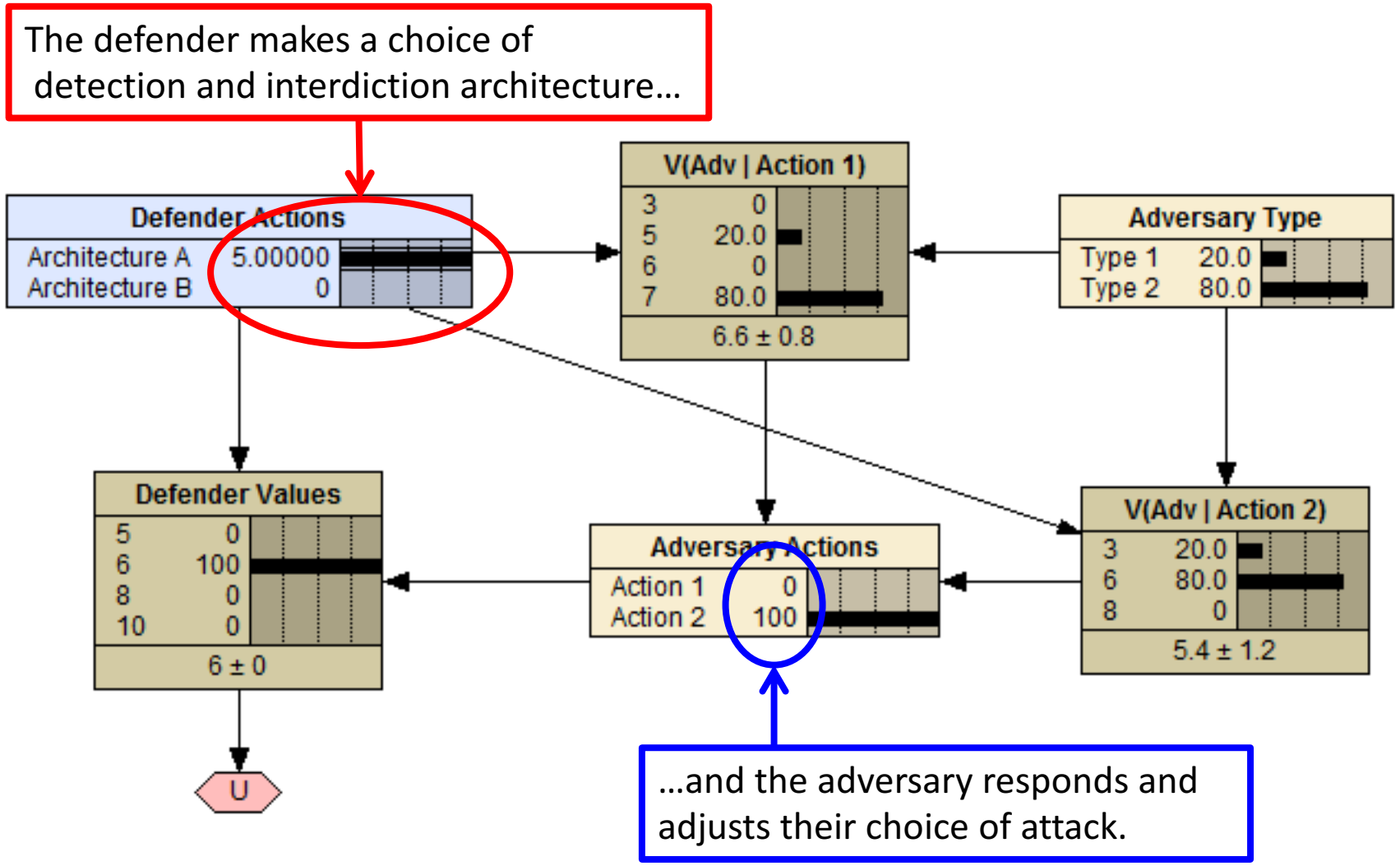
Game Theoretic Method Models

This model can be implemented using familiar tools.

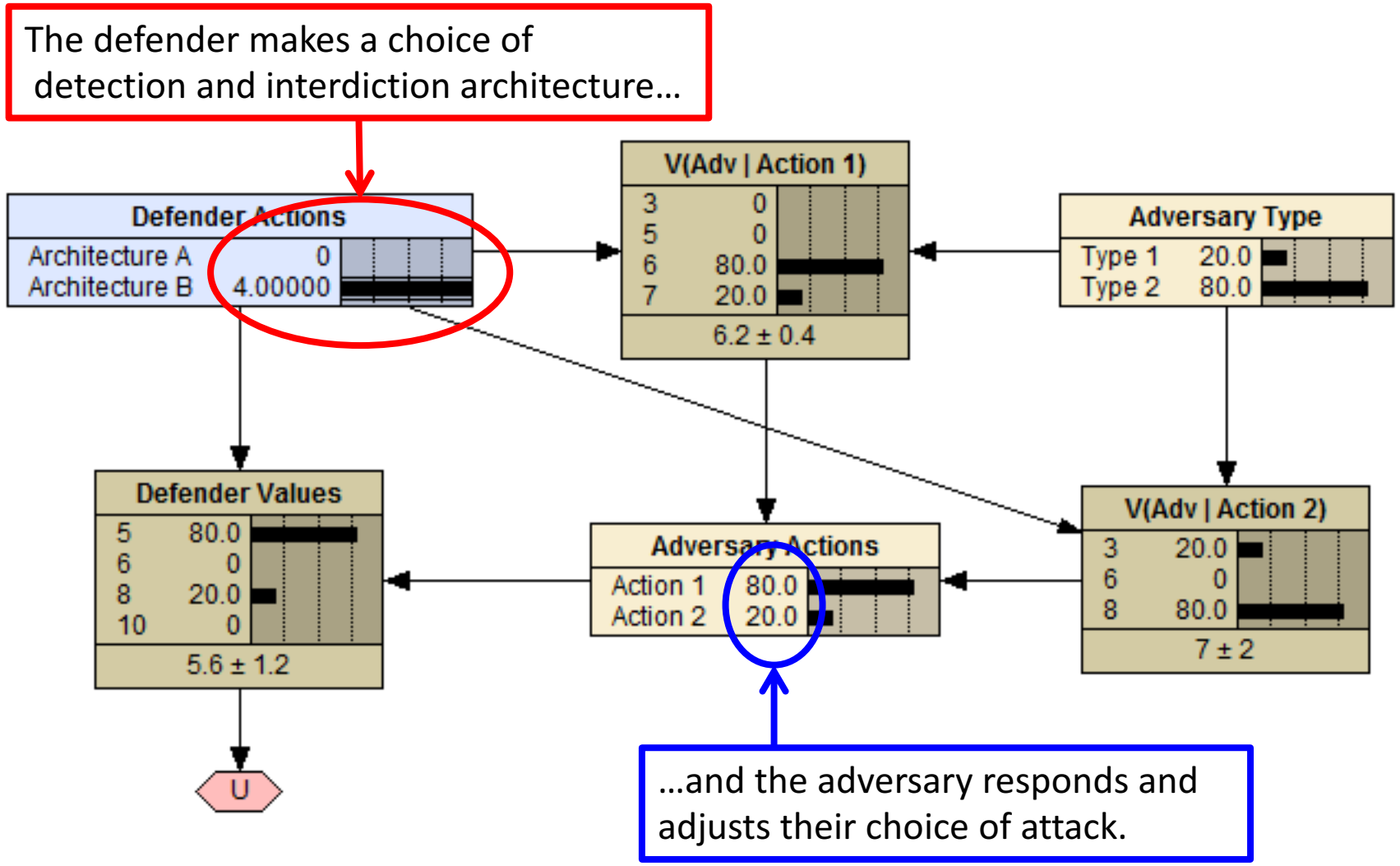
Analysis can then answer important questions.



Threat shifting behavior is captured by the game theoretic model.

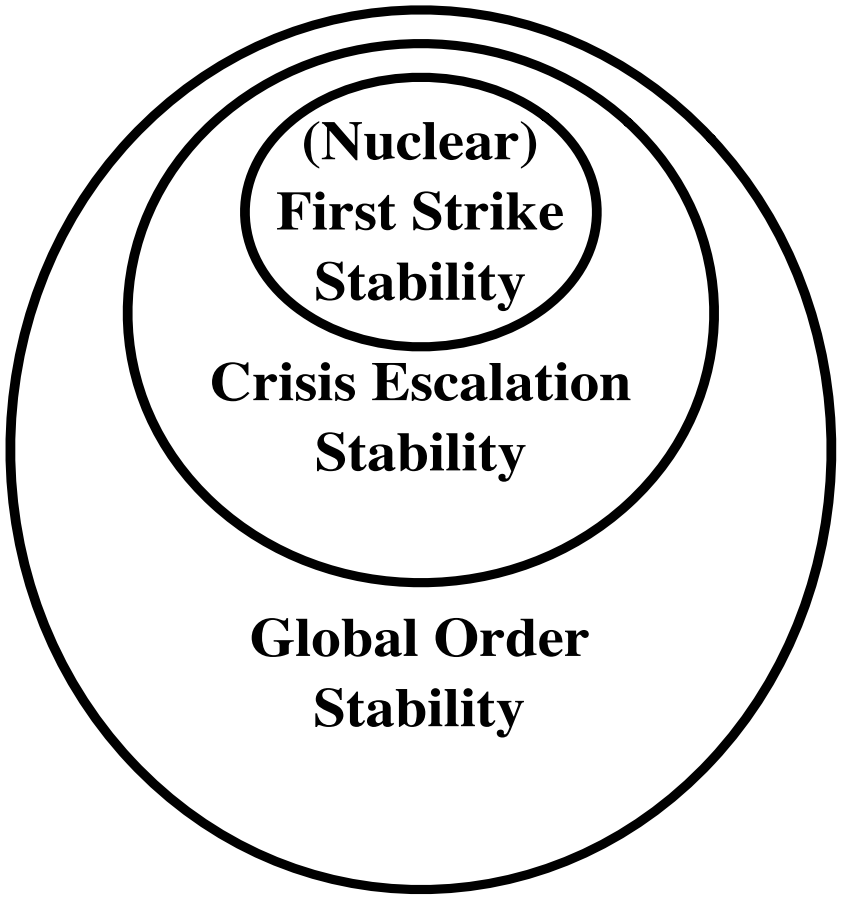


Threat shifting behavior is captured by the game theoretic model.

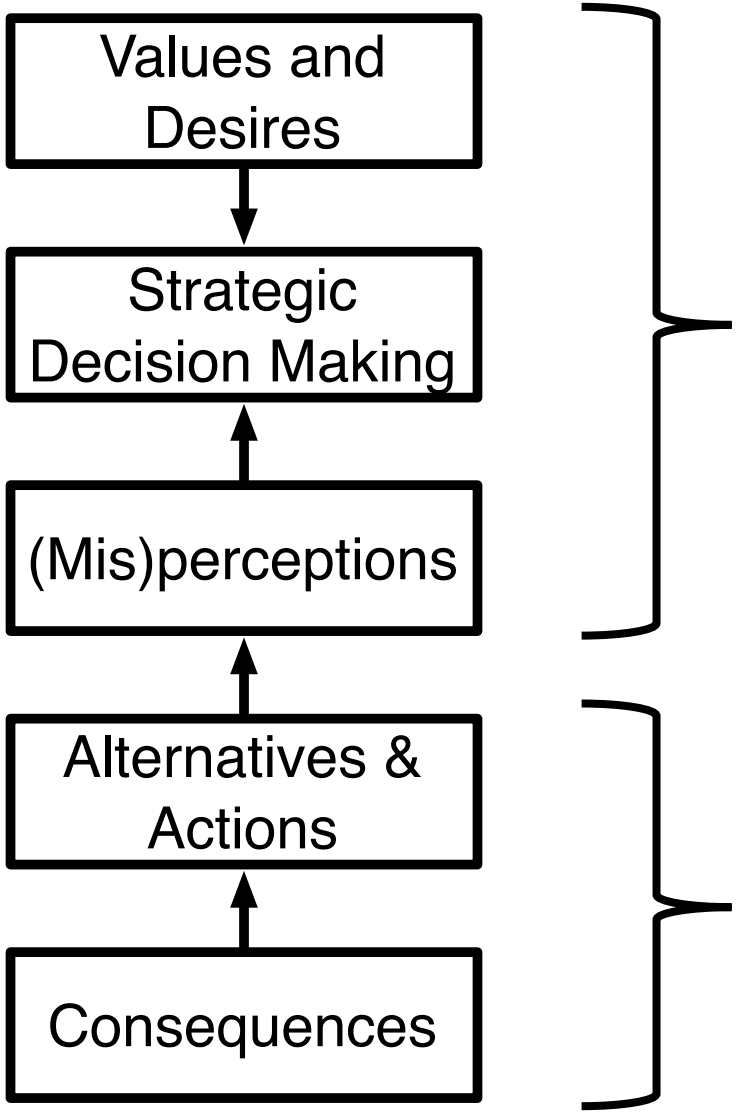


Strategic Stability

- At least three definitions of stability that require careful attention and thought
  - First strike stability is the classic definition
  - Crisis escalation stability is preventing violence
  - Global order stability is a “peaceful evolution”
- The central questions in all cases are:
  - How do incentives change in complex conflicts?
  - How likely and how bad are outcomes?
  - What are the trade-offs?
  - How can actions and policies affect those incentives, outcomes, and trade-offs?
- The focus should be on decisions that adversaries make, and the incentives that drive them.



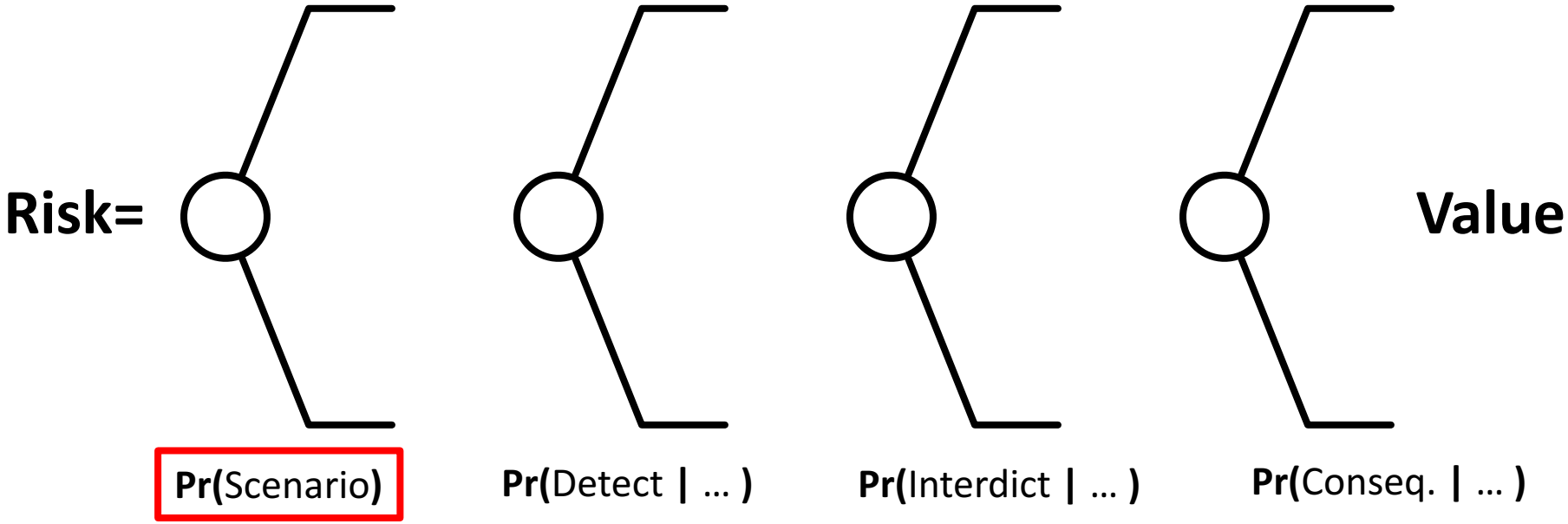
Strategic Stability and Adversary Modeling



- Based on ***uncertainty***
- Historically ***qualitative*** *in most cases*
- Significantly under-theorized for:
  - More than 2 agents
  - Escalation
  - Less than existential threats
- Based on commonly held assumptions of adversary desires and decision processes

- Based in ***physics***
- Historically ***quantitative***
- Decades of research
- In some cases, significant data exists

Conclusions



- Many approaches to adversary modeling can be helpful and yield powerful insights to decision makers.
- The goal is NOT TO PREDICT adversary actions, but to understand when adversaries might be incentivized to act.
- There is a significant need for continued research and application of quantitative adversary models.