Adversarial Risk Analysis for Counterterrorism Modeling

Jesus Rios
Manchester Business School, United Kingdom

(David Rios Insua and David Banks)

Outline

- Game Theory vs Subjective Expected Utility Theory
  Adversarial Risk Analysis

- Supporting the Defender against the Attacker
  The assessment of Attacker's decision

- Solving
  - Defend-Attack sequential game
  - Defend-Attack simultaneous game
  - Defend-Attack-Defend sequential game

- Discussion

Critiques to the Game Theory approach

- Strict and unrealistic assumptions
  - Full and common knowledge assumption
  - Common prior assumption for games with incomplete information
- Symmetric predictive and descriptive approach
  - What if multiple equilibria
  - Passive understanding
- Equilibria does not provide partisan advise

One-sided prescriptive support

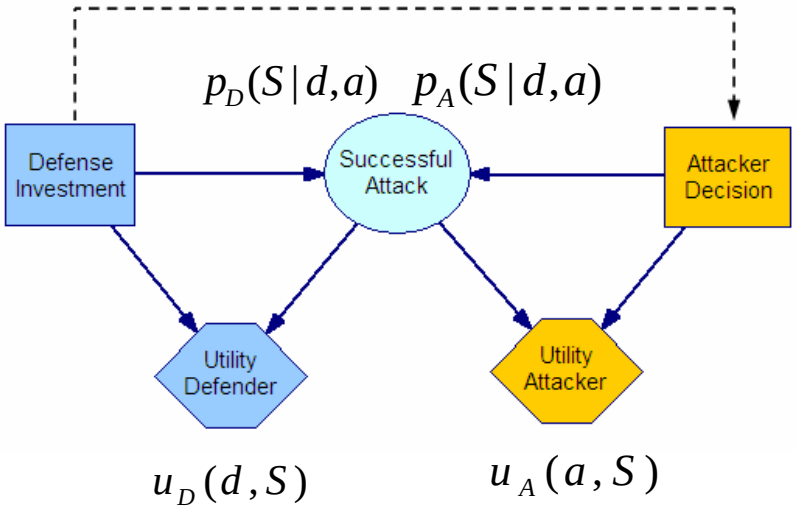
- Asymmetric prescriptive/descriptive approach (Raiffa)
  - Prescriptive advice to one party conditional on a (probalistic) description of how others will behave
- A bayesian approach (Kadane, Larkey…)
  - Use a SEU model for supporting the Defender
  - Treat the Attacker's decision as uncertain
  - Help the Defender to assess probabilities of Attacker's decisions
- Adversarial Risk Analysis
  - Weaken common (prior) knowledge assumption
  - Develop methods for the analysis of the adversaries' thinking to anticipate their actions.
    - We assume that the Attacker is a *expected utility maximizer*
    - But other models may be possible

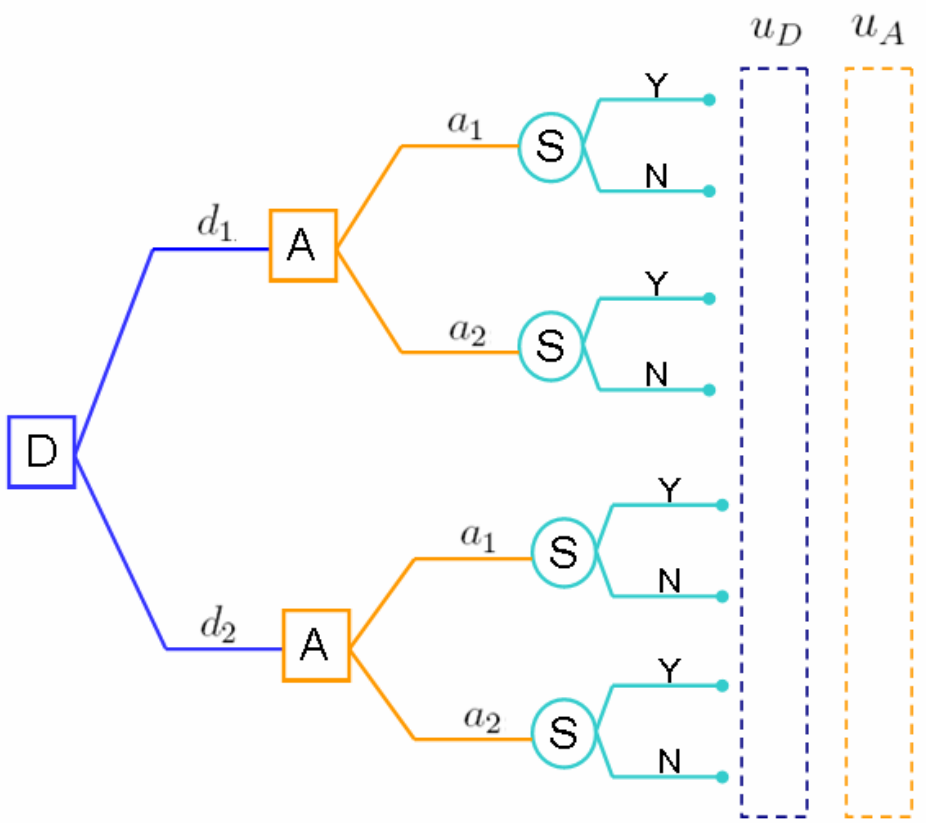
Assessing adversary's intelligent decisions

- Distinction between uncertainty stemming from
  - Nature
  - Intelligent adversaries' actions
- How to assess probabilities of Attacker's actions??
- Assuming the Attacker is a SEU maximizer
  - Based on an analysis of his decision problem
    - Assess Attacker' probabilities and utilities
    - Find his action of maximum expected utility
- Uncertainty about Attacker' decision should reflect
  - Defender's uncertainty about Attacker's probabilities and utilities
- Sources of information
  - Available past statistical data of Attacker's decision behavior
  - Expert knowledge
  - Non-informative (or reference) distributions

Defend-Attack sequential model

- Two intelligent players
  - Defender and Attacker
- Sequential moves
  - First Defender, afterwards Attacker knowing Defender's decision





Standard Game Theory Analysis

Expected utilities at node S

$$\psi_D(d,a) = p_D(S=0|d,a)\ u_D(d,S=0)\ +\ p_D(S=1|d,a)\ u_D(d,S=1)$$

$$\psi_A(d,a) = p_A(S=0\mid d,a)\ u_A(a,S=0)\ +\ p_A(S=1\mid d,a)\ u_A(a,S=1)$$

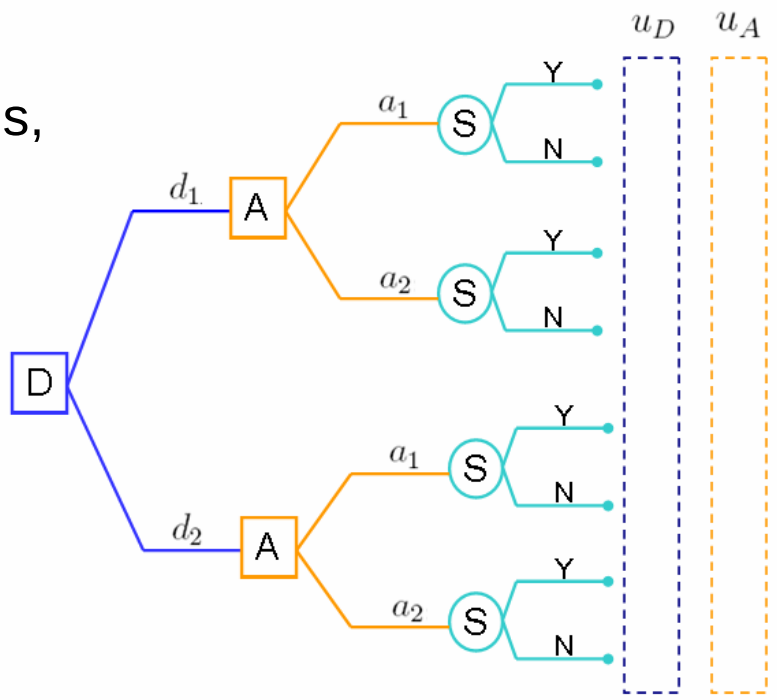
Best Attacker's decision at node A

$$a^*(d) = \text{argmax}_{a\in\mathcal{A}}\psi_A(d,a)$$

Assuming Defender knows Attacker's analysis, Defender's best decision at node D

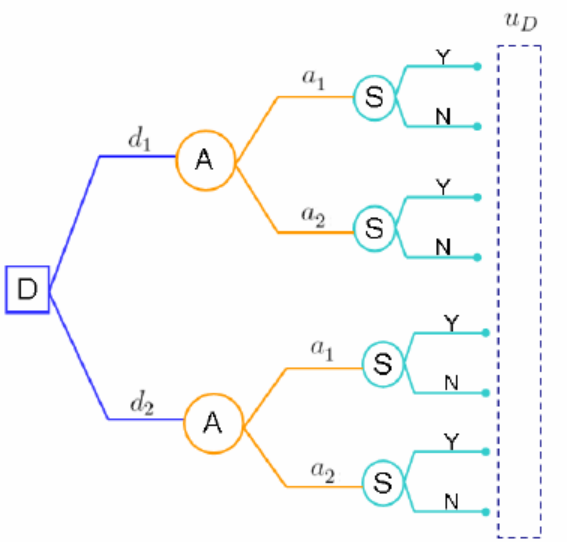
$$d^* = \text{argmax}_{d\in\mathcal{D}}\psi_D(d,a^*(d))$$

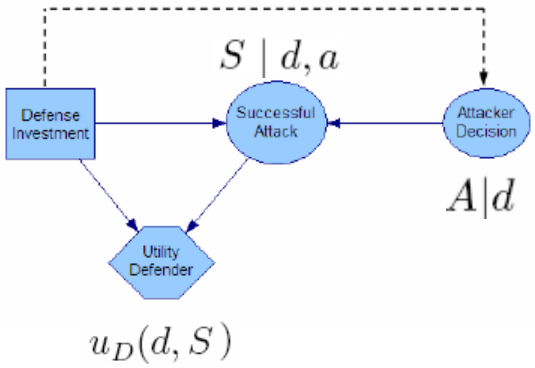
Nash Solution: $(d^*, a^*(d^*))$



Supporting the Defender

Defender's problem	Defender's solution of maximum SEU





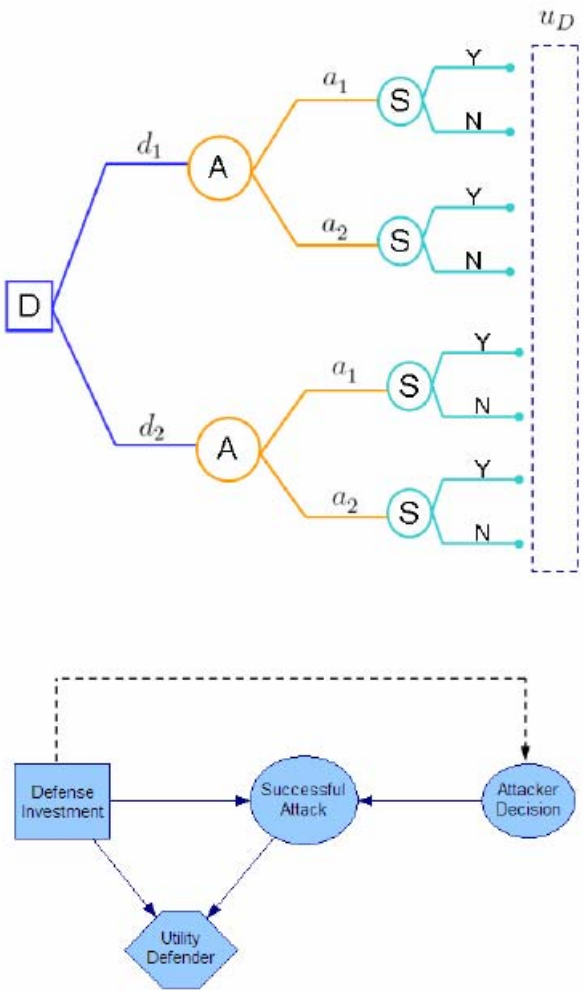
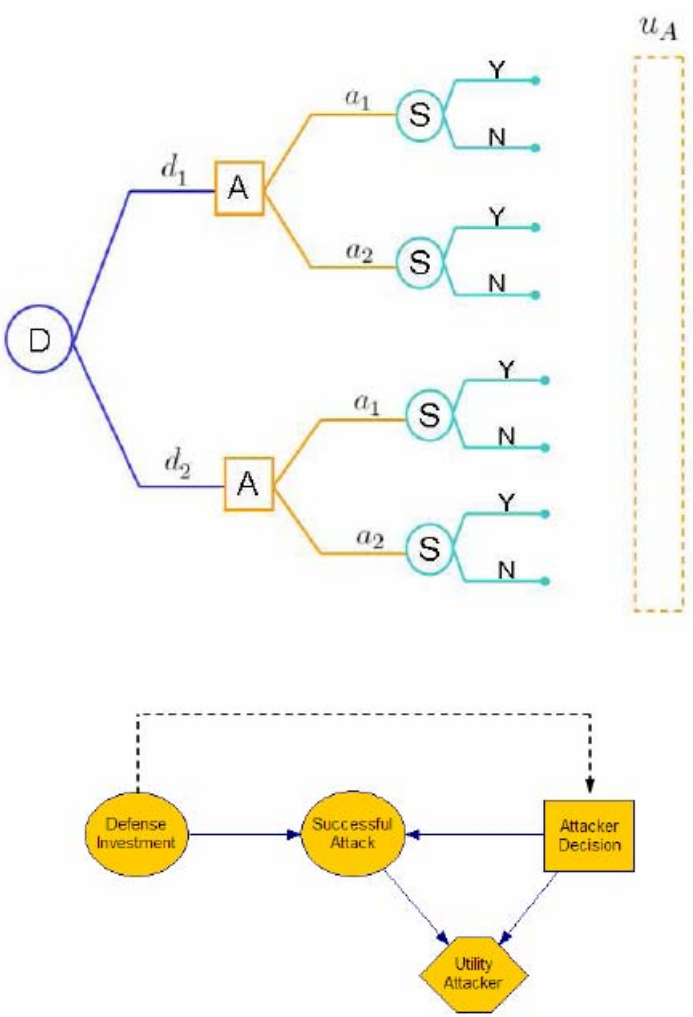
$$\psi_D(d,a) = p_D(S=0|d,a)\ u_D(d,S=0)\ + \\ p_D(S=1|d,a)\ u_D(d,S=1)$$

$$\psi_D(d) = \psi_D(d,a_1)\ p_D(A=a_1|d) + \\ \psi_D(d,a_2)\ p_D(A=a_2|d)$$

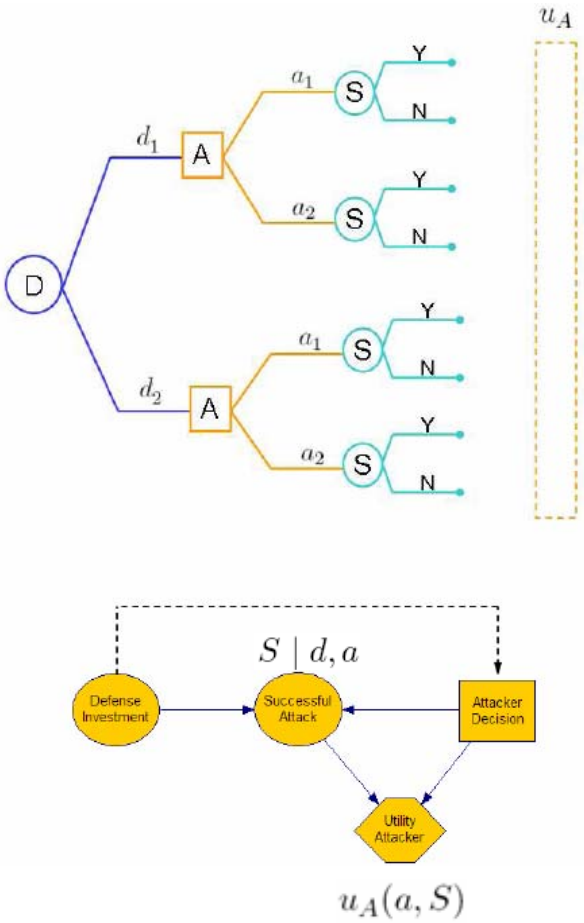
$$d^* = \arg\max_{d \in X_D} \psi_D(d)$$

Modeling input: $p_D(S|a,d)$ $p_D(A\,|d)$??

Supporting the Defender assessing Attacker's decision

Defender problem	Defender's view of Attacker problem
	

Solving the assessment problem

Defender's view of Attacker problem	Elicitation of $p_D(A \mid d)$
	A is a EU maximizer D's beliefs about $(u_A, p_A) \sim (P_A, U_A) = F$ $\Psi_A(d,a) = P_A(S=0 \mid d,a)\, U_A(a,S=0) + P_A(S=1 \mid d,a)\, U_A(a,S=1)$ $p_D(A=a\|d) = \mathbb{P}_F[a = \text{argmax}_{x\in\mathcal{A}} \Psi_A(d,x)]$ <u>MC simulation</u> $\{(p_A^i, u_A^i)\}_{i=1}^n \sim F \longrightarrow \psi_A^i \sim \Psi_A$ $a_i^*(d) = \text{argmax}_{x\in\mathcal{A}} \psi_A^i(x,d)$ $p_D(A=a\|d) \approx \#\{a = a_i^*(d)\}/n$

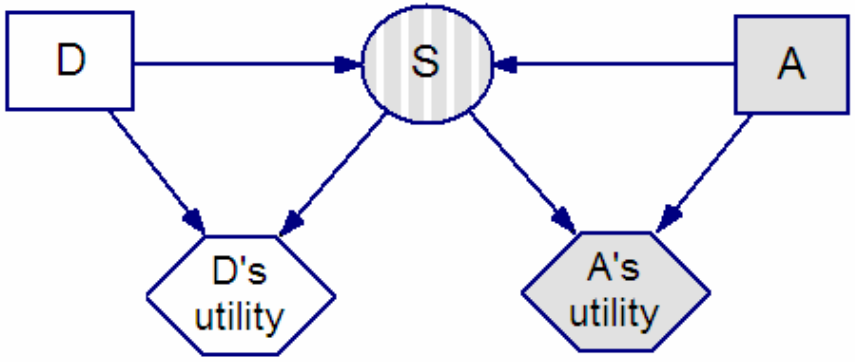
Bayesian solution for the Defend- Attack sequential model

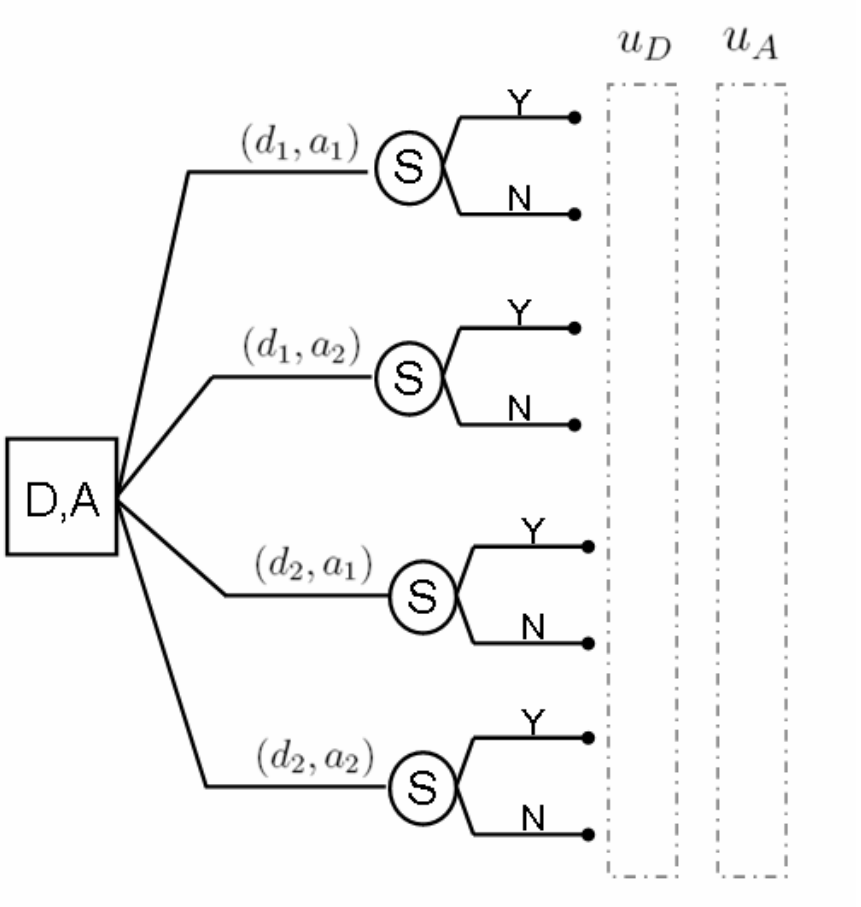
1. Assess (p_D, u_D) from the Defender
2. Assess $F = (P_A, U_A)$, describing the Defender's uncertainty about (p_A, u_A)
3. For each d, simulate to assess $p_D(A|d)$ as follows:
   (a) Generate $(p_A^i, u_A^i) \sim F$, $i = 1, \ldots, n$
   (b) Solve $a_i^*(d) = \text{argmax}_{a \in \mathcal{A}} \ \psi_A^i(d, a)$
   (c) Approximate $\hat{p}_D(A = a|d) = \#\{a = a_i^*(d)\}/n$
4. Solve the Defender's problem

$$d^* = \text{argmax}_{d \in \mathcal{D}} \ \psi_D(d, a_1)\, \hat{p}_D(A = a_1|d) \ + \ \psi_D(d, a_2)\, \hat{p}_D(A = a_2|d)$$

Defend-Attack simultaneous model

- Decision are taken without knowing each other's decisions





Game Theory Analysis

- Common knowledge
  - Each knows expected utility of every pair (d,a) for both of them
  - Nash equilibrium: (d*, a*) satisfying

$$\psi_D(d^*, a^*) \geq \psi_D(d, a^*) \;\; \forall\, d \in \mathcal{D}$$

$$\psi_A(d^*, a^*) \geq \psi_A(d^*, a) \;\; \forall\, a \in \mathcal{A}$$

- When some information is not common knowledge
  - Private information
    - Type of Defender and Attacker

$$\tau_D \in T_D \longrightarrow u_D(d, s, \tau_D) \quad p_D(S \mid d, a, \tau_D)$$

$$\tau_A \in T_A \longrightarrow u_A(d, s, \tau_D) \quad p_A(S \mid d, a, \tau_D)$$

  - Common prior over private information $\pi(\tau_D, \tau_A)$
  - Model the game as one of incomplete information

Bayes Nash Equilibrium

– Strategy functions
  - Defender $d : \tau_D \rightarrow d(\tau_D) \in \mathcal{D}$
  - Attacker $a : \tau_A \rightarrow a(\tau_A) \in \mathcal{A}$

– Expected utility of (d,a)
  - for Defender, given her type $\psi_D(d(\tau_D), a, \tau_D) =$

$$= \int \left[\sum_{s \in S} u_D(d(\tau_D), s, \tau_D)\, p_D(S = s \mid d(\tau_D), a(\tau_A), \tau_D) \right] \pi(\tau_A \mid \tau_D)\, \mathrm{d}\tau_A$$

  - Similarly for Attacker, given his type $\psi_A(d, a(\tau_A), \tau_A)$

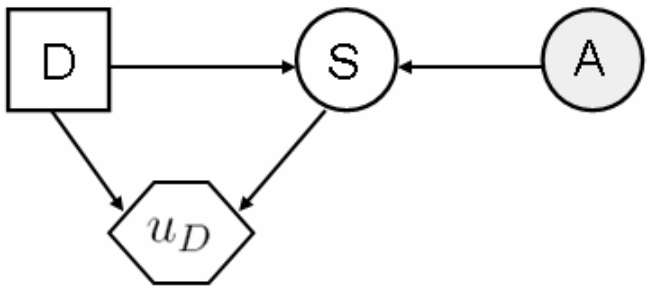
– Bayes-Nash Equlibrium (d*, a*) satisfying

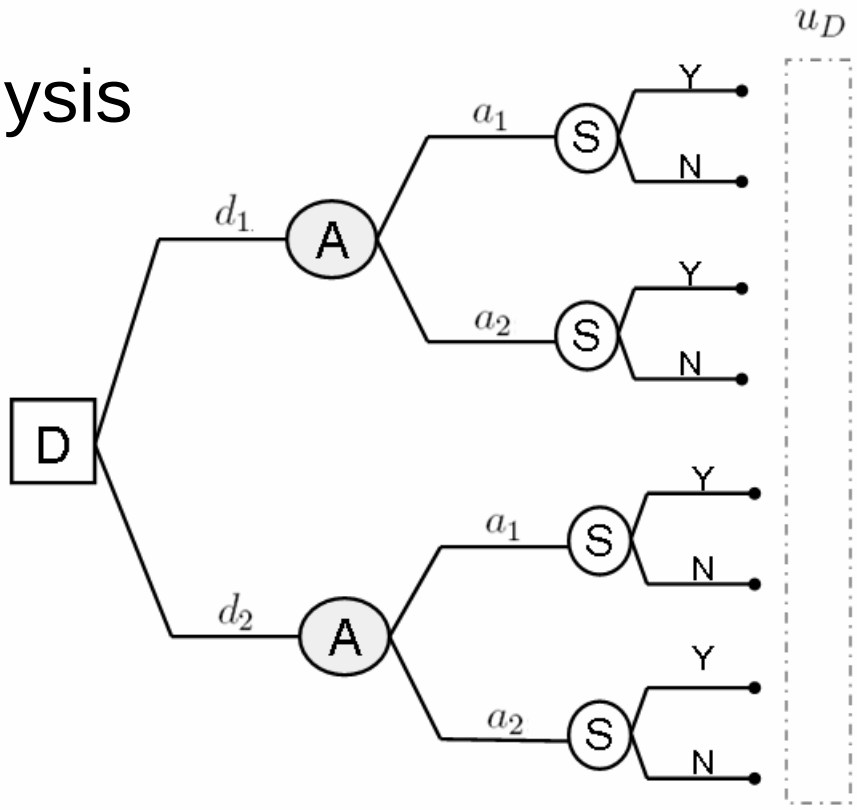
$$\psi_D(d^*(\tau_D), a^*, \tau_D) \geq \psi_D(d(\tau_D), a^*, \tau_D) \quad \forall\, d : \tau_D \rightarrow d(\tau_D)$$

$$\psi_A(d^*, a^*(\tau_A), \tau_A) \geq \psi_A(d^*, a(\tau_A), \tau_A) \quad \forall\, a : \tau_A \rightarrow a(\tau_A)$$

Supporting the Defender

- Defender's decision analysis



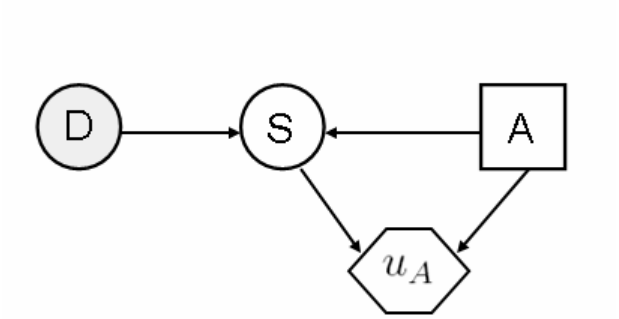


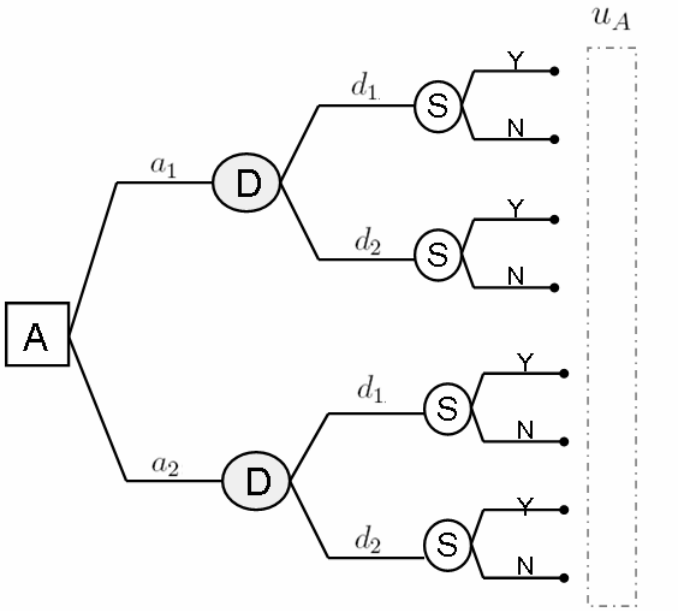
$$d^* = \text{argmax}_{d \in \mathcal{D}} \sum_{a \in \mathcal{A}} \left[\sum_{s \in \{0,1\}} u_D(d,s)\, p_D(S = s \mid d, a) \right] \pi_D(A = a)$$

How to elicit it ??

Assessing $\pi_D(A = a)$

- Attacker's decision analysis as seen by the Defender





$$a^* = \text{argmax}_{a \in \mathcal{A}} \sum_{d \in \mathcal{D}} \left[\sum_{s \in \{0,1\}} u_A(a, s)\, p_A(S = s \mid d, a) \right] \pi_A(D = d)$$

$$(u_A, p_A, \pi_A) \sim (U_A, P_A, \Pi_A)$$

Assessing $\pi_D(A = a)$

$$A \mid D \sim \text{argmax}_{a \in \mathcal{A}} \sum_{d \in \mathcal{D}} \left[\sum_{s \in \{0,1\}} U_A(a, s)\ P_A(S = s \mid d, a) \right] \Pi_A(D = d)$$

- $\Pi_A(D = d)$
  - Attacker's uncertainty about Defender's decision $\pi_A(D = d)$
  - Defender's uncertainty about the model used by the Attacker to predict what defense the Defender will choose $\pi_A \sim \Pi_A$

- The elicitation of $\Pi_A(D = d)$ may require further analysis at the next level of recursive thinking

$$D \mid A^1 \sim \text{argmax}_{d \in \mathcal{D}} \sum_{a \in \mathcal{A}} \left[\sum_{s \in \{0,1\}} U_D(d, s)\ P_D(S = s \mid d, a) \right] \Pi_D(A^1 = a)$$

The assessment problem

- To predict Attacker's decision
  The Defender needs to solve Attacker's decision problem
  She needs to assess (u_A, p_A, π_A)

- Her beliefs about (u_A, p_A, π_A) are modeled through a probability distribution (U_A, P_A, Π_A)

- The assessment of $\Pi_A(D = d)$ requires deeper analysis
  - D's analysis of A's analysis of D's problem

- It leads to an infinite regress thinking-about-what-the-other-is-thinking-about…

Hierarchy of nested models

Repeat

Find $\Pi_{D^{i-1}}(A^i)$ by solving

$$A^i \mid D^i \sim \text{argmax}_{a \in \mathcal{A}} \sum_{d \in \mathcal{D}} \left[\sum_{s \in \{0,1\}} U_A^i(a,s)\, P_A^i(S = s \mid d, a) \right] \Pi_{A^i}(D^i = d)$$

where $(U_A^i, P_A^i) \sim F^i$

Find $\Pi_{A^i}(D^i)$ by solving

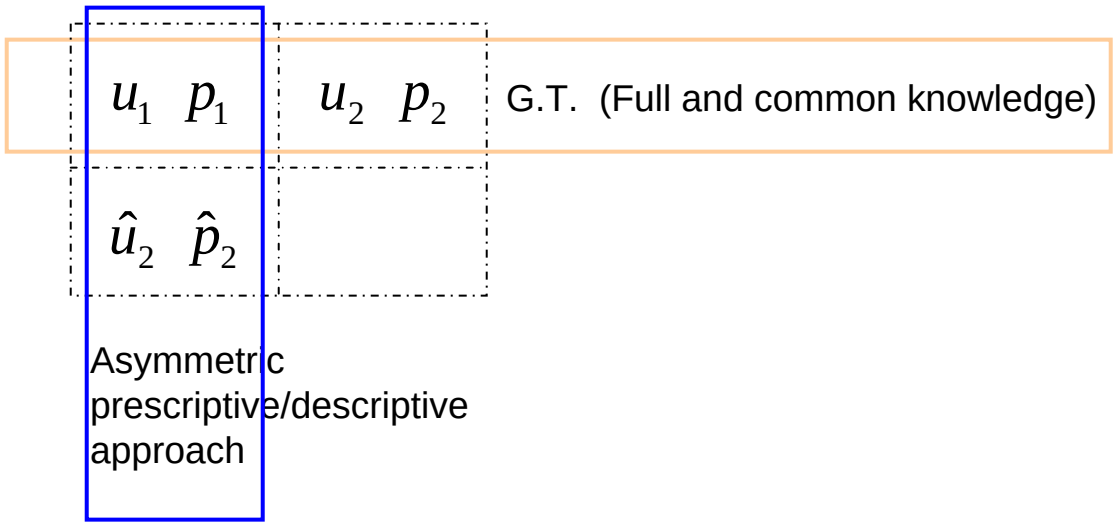
$$D^i \mid A^{i+1} \sim \text{argmax}_{d \in \mathcal{D}} \sum_{a \in \mathcal{A}} \left[\sum_{s \in \{0,1\}} U_D^i(d,s)\, P_D^i(S = s \mid d, a) \right] \Pi_{D^i}(A^{i+1} = a)$$

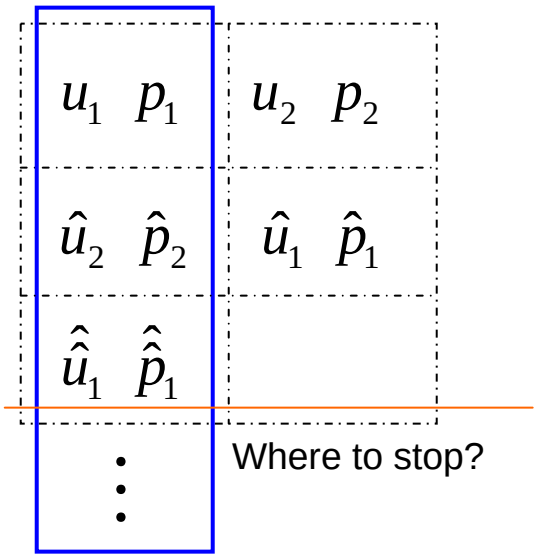
where $(U_D^i, P_D^i) \sim G^i$

$i = i + 1$

Stop when the Defender has no more information about utilities and probabilities at some level of the recursive analysis

How to avoid infinite regress?





A numerical example

- Defender chooses d_1 or d_2
- Simultaneously Attacker must choose a_1 or a_2
- Defender assessments:

$u_D(d, s)$

	$s = 1$	$s = 0$
d_1	50	80
d_2	0	100

$p_D(S = 1 \mid d, a)$

	a_1	a_2
d_1	0.1	0
d_2	0.9	0

  - Two different types of Attacker
    - Type I prob 0.8
    - Type II prob 0.2

$(U_{A_I}, P_{A_I}) \sim F_I$:

$U_{A_I}(a, s)$

	$s = 1$	$s = 0$
a_1	$Tri(20, 100, 100)$	$Tri(0, 20, 100)$
a_2	100	$Tri(0, 40, 100)$

$P_{A_I}(S = 1 \mid d, a)$

	a_1	a_2
d_1	$\mathcal{U}[0, 1]$	0
d_2	$Tri(0.5, 1, 1)$	0

$(U_{A_{II}}, P_{A_{II}}) \sim F_{II}$:

$U_{A_{II}}(a, s)$

	$s = 1$	$s = 0$
a_1	$\mathcal{U}[0, 100]$	$Tri(0, 20, 100)$
a_2	100	$Tri(40, 80, 90)$

$P_{A_{II}}(S = 1 \mid d, a)$

	a_1	a_2
d_1	$Tri(0, 0, 1)$	0
d_2	$Tri(0, 1, 1)$	0

- Defender thinks that a Type I Attacker is smart enough to analyze her problem
  - A Type I Attacker's beliefs about her utilities and probabilities are

$(U_{D_I}, P_{D_I}) \sim G_I$:

$U_{D_I}(d, s)$

	$s = 1$	$s = 0$
d_1	$Tri(0, 0, 40)$	$\mathcal{U}[50, 100]$
d_2	$Tri(0, 0, 40)$	$\mathcal{U}[50, 100]$

$P_{D_I}(S = 1 \mid d, a)$

	a_1	a_2
d_1	$Tri(0, 0, 0.5)$	0
d_2	$\mathcal{U}[0, 1]$	0

$$\Pi_{A_I}(D_I = d_1) \sim \mathcal{B}e(\alpha, 10 - \alpha), \text{ where } \alpha = \pi_{A_I}(D_I = d_1) \times 10$$

- However, the Defender does not know how a Type II Attacker would analyze her problem, but believes that

$$\Pi_{A_{II}}(D_{II} = d_1) \sim \mathcal{B}e(75, 25)$$

- Defender: what does Type I Attacker think to be her beliefs about what he will do

$$\Pi_{D_I}(A_I^1 = a_1) \sim \mathcal{U}[0, 1]$$

- Solving Defender's decision problem
  - Computing her defense of max. expected utility
- She first needs to compute
  - Her predictive distribution about what an Attacker will do

$$\pi_D(A = a_1) = 0.8 \times \pi_D(A_I = a_1) + 0.2 \times \pi_D(A_{II} = a_1)$$

$\pi_D(A_I = a_1)$ ⟶

1. For $k = 1, \ldots, n$, repeat
   - Draw $\pi^k_{D_I} \sim \Pi_{D_I}$, that is $\pi^k_{D_I}(A^1_I = a_1) \sim \mathcal{U}[0,1]$.
   - Draw $(u^k_{D_I}, p^k_{D_I}) \sim (U_{D_I}, P_{D_I}) = G_I$
   - Compute

$$d^k_I = \text{argmax}_{d \in \mathcal{D}} \sum_{a \in \mathcal{A}} \left[\sum_{s \in \{0,1\}} u^k_{D_I}(d,s)\, p^k_{D_I}(S = s \mid d, a) \right] \pi^k_{D_I}(A^1_I = a)$$

2. Approximate $\pi_{A_I}(D_I = d_1)$ through $\hat{\pi}_{A_I}(D_I = d_1) = \#\{d^k_I = d_1\}/n$.

   Set $\hat{\Pi}_{A_I}(D_I = d_1) \sim \mathcal{B}e(\alpha, 10 - \alpha)$, with $\alpha = \hat{\pi}_{A_I}(D_I = d_1) \times 10$.

3. For $k = 1, \ldots, n$, repeat
   - Draw $\hat{\pi}^k_{A_I} \sim \hat{\Pi}_{A_I}$, that is $\hat{\pi}^k_{A_I}(D_I = d_1) \sim \hat{\Pi}_{A_I}(D_I = d_1)$
   - Draw $(u^k_{A_I}, p^k_{A_I}) \sim (U_{A_I}, P_{A_I}) = F_I$
   - Compute

$$a^k_I = \text{argmax}_{a \in \mathcal{A}} \sum_{d \in \mathcal{D}} \left[\sum_{s \in \{0,1\}} u^k_{A_I}(a,s)\, p^k_{A_I}(S = s \mid d, a) \right] \hat{\pi}^k_{A_I}(D_I = d)$$

4. Approximate $\pi_D(A_I = a_1)$ through $\hat{\pi}_D(A_I = a_1) = \#\{a^k_I = a_1\}/n$.

$\pi_D(A_{II} = a_1)$ →

1. For $k = 1, \ldots, n$, repeat
   - Draw $\pi^k_{A_{II}} \sim \Pi_{A_{II}}$, that is $\pi^k_{A_{II}}(D_{II} = d_1) \sim \mathcal{B}e(75, 25)$.
   - Draw $(u^k_{A_{II}}, p^k_{A_{II}}) \sim (U_{A_{II}}, P_{A_{II}}) = F_{II}$
   - Compute

$$a^k_{II} = \text{argmax}_{a \in \mathcal{A}} \sum_{d \in \mathcal{D}} \left[\sum_{s \in \{0,1\}} u^k_{A_{II}}(a, s)\, p^k_{A_{II}}(S = s \mid d, a) \right] \pi^k_{A_{II}}(D_{II} = d)$$

2. Approximate $\pi_D(A_{II} = a_1)$ through $\hat{\pi}_D(A_{II} = a_1) = \#\{a^k_{II} = a_1\}/n$.

– In a run with n=1000, we got

$$\hat{\pi}_D(A_I = a_1) = 0.97 \times 0.8$$

$$\hat{\pi}_D(A_{II} = a_1) = 0.82 \times 0.2$$

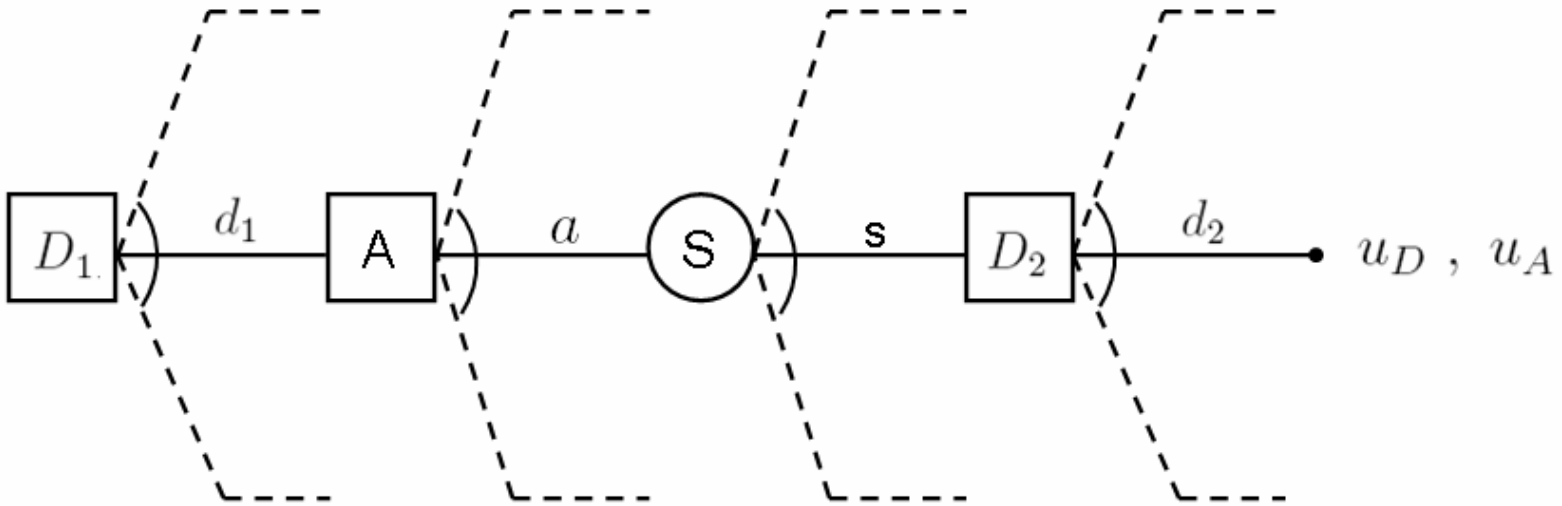
$$\hat{\pi}_D(A = a_1) = 0.94$$

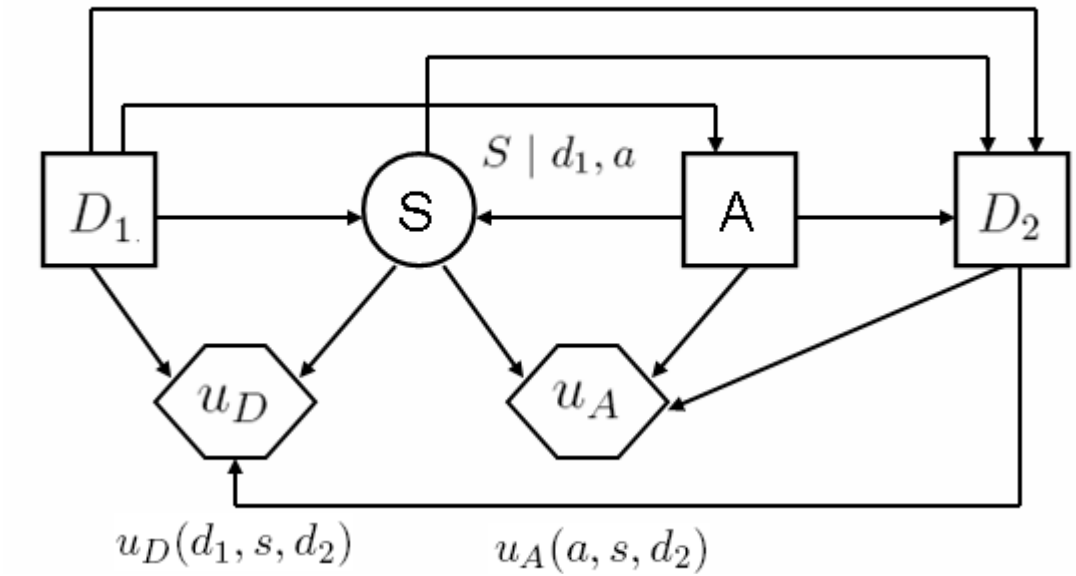
- And, now the Defender can solve her problem

$$d^* = \text{argmax}_{d \in \mathcal{D}} \sum_{a \in \mathcal{A}} \left[\sum_{s \in \{0,1\}} u_D(d, s)\, p_D(S = s \mid d, a) \right] \pi_D(A = a)$$

$d^* = d_1$ with (MC estimated) expected utility 77, against d_2 with 15

Defend – Attack – Defend model

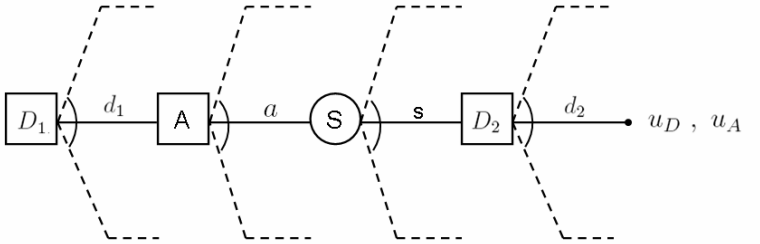




Standard Game Theory Analysis

- Under common knowledge of utilities and probs
- At node D_2

$$d_2^*(d_1, s) = \text{argmax}_{d_2 \in \mathcal{D}_2} u_D(d_1, s, d_2)$$



- Expected utilities at node S

$$\psi_D(d_1, a) = \int u_D(d_1, s, d_2^*(d_1, s))\, p_D(s \mid d_1, a)\, \mathrm{d}s$$

$$\psi_A(d_1, a) = \int u_A(a, s, d_2^*(d_1, s))\, p_A(s \mid d_1, a)\, \mathrm{d}s$$

- Best Attacker's decision at node A

$$a^*(d_1) = \text{argmax}_{a \in \mathcal{A}} \psi_A(d_1, a)$$

- Best Defender's decision at node D_1

$$d_1^* = \text{argmax}_{d_1 \in \mathcal{D}_1} \psi_D(d_1, a^*(d_1))$$

- Nash Solution

$$d_1^* \in \mathcal{D}_1 \qquad a^*(d_1^*) \in \mathcal{A} \qquad d_2^*(d_1^*, s) \in \mathcal{D}_2$$

Supporting the Defender

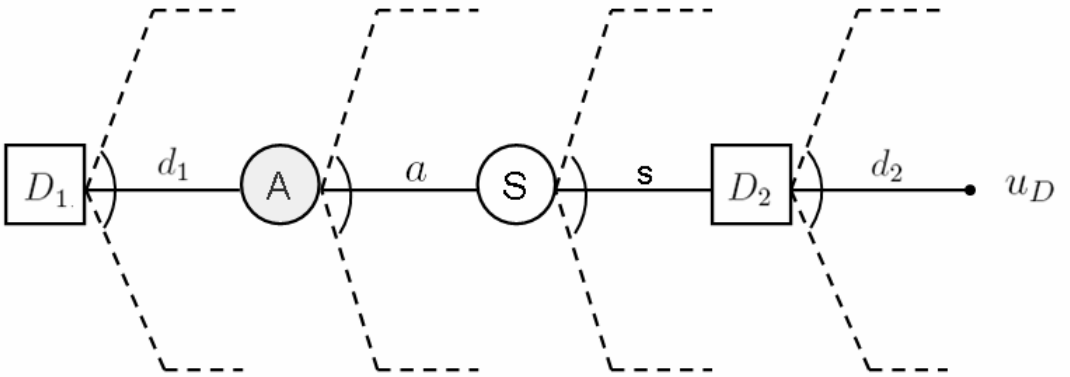
- At node A

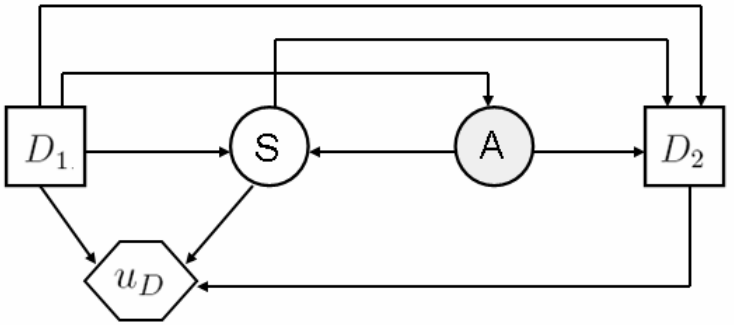
$$\psi_D(d_1) = \int \psi_A(d_1, a)\, p_D(a \mid d_1)\, \mathrm{d}a$$

- At node D_1

$$d_1^* = \mathrm{argmax}_{d_1 \in \mathcal{D}_1} \psi_D(d_1)$$

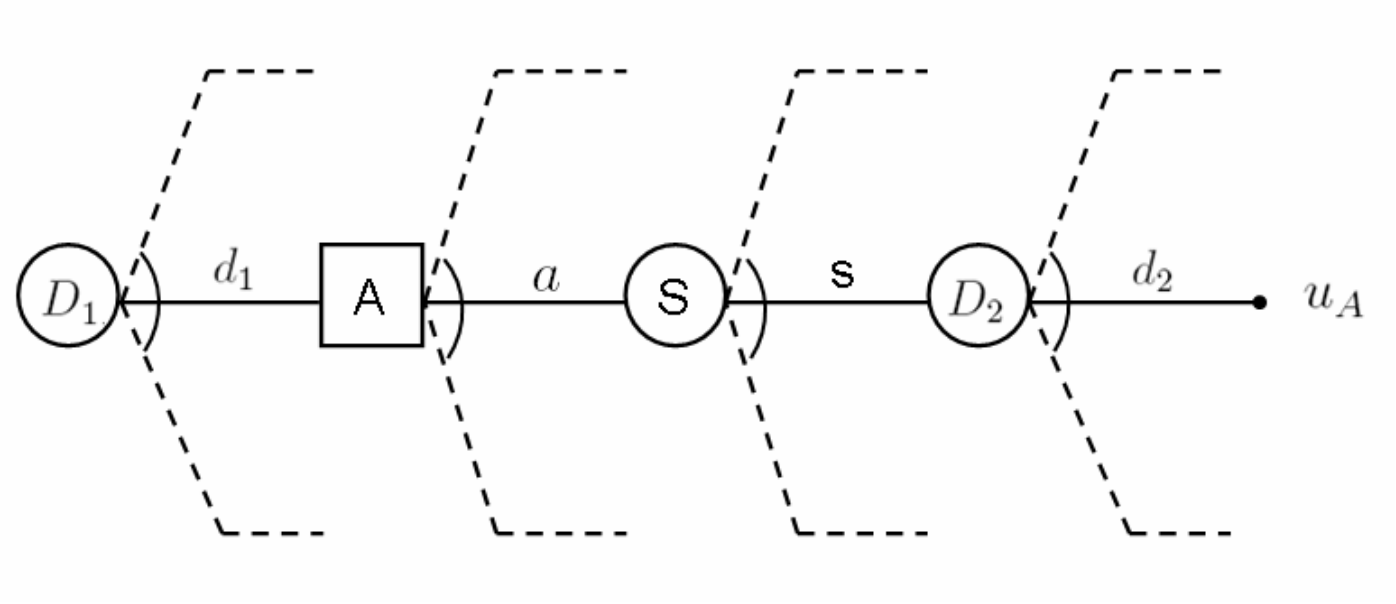
- $p_D(A \mid d_1)$??

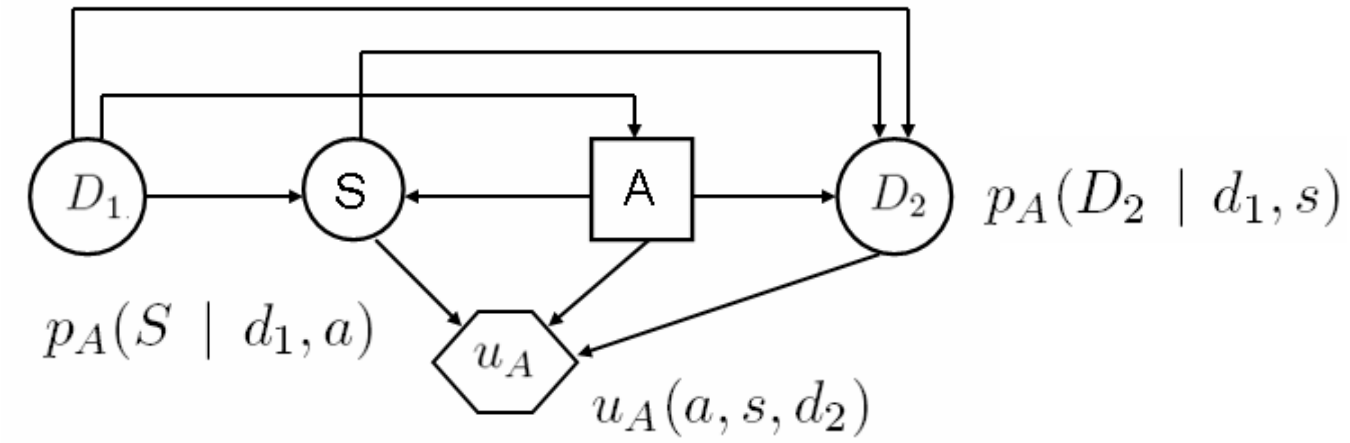




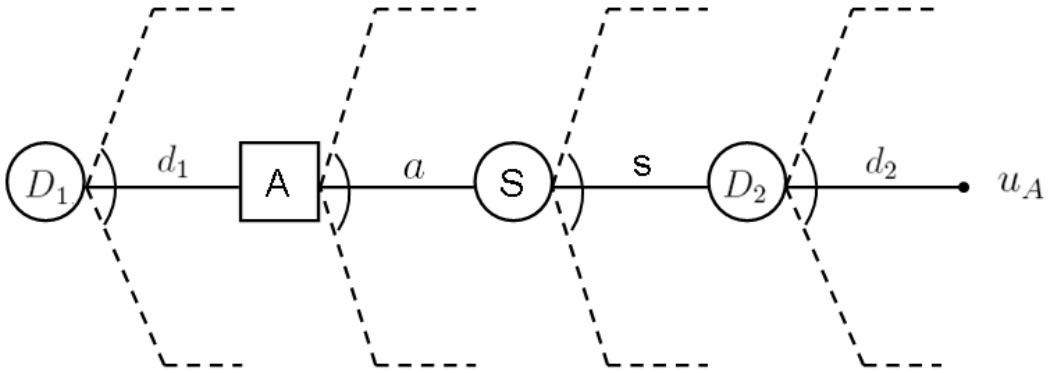
Assessing $p_D(A \mid d_1)$

- Attacker's problem as seen by the Defender





Assessing $p_D(A \mid d_1)$



- At chance node D_2, compute

$$(d_1, a, s) \rightarrow \Psi_A(d_1, a, s) = \int U_A(a, s, d_2)\ P_A(D_2 = d_2 \mid d_1, s)\ \mathrm{d}d_2$$

- At chance node S

$$(d_1, a) \rightarrow \Psi_A(d_1, a) = \int \Psi_A(d_1, a, s)\ P_A(S = s \mid d_1, a)\ \mathrm{d}s$$

- At decision node A

$$d_1 \rightarrow A^*(d_1) = \mathrm{argmax}_{a \in \mathcal{A}} \Psi_A(d_1, a)$$

- $p_D(A = a \mid d_1) = \Pr(A^*(d_1) = a)$

Monte-Carlo approximation of $p_D(A \mid d_1)$

- Drawn $\{\left(u_A^i(a,s,d_2), p_A^i(S \mid d_1, a), p_A^i(D_2 \mid d_1, s)\right)\}_{i=1}^n \sim F$
- Generate $\{a_i^*(d_1)\}_{i=1}^n$ by
  - At chance node D_2

    $(d_1, a, s) \rightarrow \psi_A^i(d_1, a, s) = \int u_A^i(a, s, d_2)\, p_A^i(D_2 = d_2 \mid d_1, s)\, \mathrm{d}d_2$
  - At chance node S

    $(d_1, a) \rightarrow \psi_A^i(d_1, a) = \int \psi_A^i(d_1, a, s)\, p_A^i(S = s \mid d_1, a)\, \mathrm{d}s$
  - At decision node A

    $d_1 \rightarrow a_i^*(d_1) = \mathrm{argmax}_{a \in \mathcal{A}} \psi_A^i(d_1, a)$
- Approximate

$$p_D(A = a \mid d_1) \approx \#\{a_i^*(d) = a\}/n$$

The assessment of $p_A(D_2 \mid d_1, s)$

- The Defender may want to exploit information about how the Attacker analyzes her problem
- Hierarchy of recursive analysis
- Stop when there is no more information to elicit
  - Unconditional probability assessment, or
  - Non-informative (or reference) distribution

Discussion

- DA vs GT
  - A Bayesian prescriptive approach to support Defender against Attacker
  - Weaken common (prior) knowledge assumption
  - Analysis and assessment of Attacker' thinking to anticipate their actions assuming Attacker is a expected utility maximizer
  - Computation of her defense of maximum expected utility
- Several simple but illustrative models
  - sequential D-A, simultaneous D-A and D-A-D decision problems
  - What if
    - more complex dynamic interactions?
    - against more than one Attacker?
    - an uncertain number of Attackers?
- The assessment problem under infinite regress
- Implementation issues
  - Elicitation of a valuable judgmental input from Defender
  - Computational issues