

# Adversarial Risk Analysis

Concepts, Applications and Challenges

David Ríos Insua

Royal Academy of Sciences

IE Univ, June '12

# Outline

- From risk analysis to adversarial risk analysis
- Motivation
- Sequential games
- Simultaneous games
- Auctions
- **Security**
- Intelligent interfaces
- Challenges

# Risk analysis

A systematic analytical process for assessing, managing and communicating the risk performed to understand the nature of unwanted, negative consequences to human life, health, property or the environment (so as to reduce and eliminate it)

- 1. Risk assessment.** Information on the extent and characteristics of the risk attributed to a hazard.
- 2. Risk management.** The activities undertaken to control the hazard
- 3. Risk communication.** Exchange of info and opinion concerning risk and risk-related factors among risk assessors, risk managers and other interested parties.

1 bis. Concern assessment

# Which is the best security resource allocation in a city?

City as a map with cells

Each cell has a value

For each cell, a predictive model of delictive acts

Allocate security resources (constraints)

For each cell predict the impact of resource allocation

Optimal resource allocation

NB: The bad guys also operate intelligent and organisedly!!!

SECONOMICS (Metro Barcelona, UK Grid, Anadolu Airport)

# Which is the best HW/SW maintenance for the university ERP?

Model HW/SW system (interacting HW and SW blocks)

Forecast block reliability

Forecast system reliability

Design maintenance policies

Forecast impact on reliability (and costs)

Optimal maintenance policy

NB: Again, what happens with the bad guys attacking our system?

RIESGOS (MICINN), RIESGOS-CM (CM)

# The risk management process

## 1. Determination of objectives

Preserve the operating effectiveness of the organisation

## 2. Identification of risks

## 3. Evaluation of risks

## 4. Considering alternatives and selecting the risk treatment device

## 5. Implementing the decision

## 6. Evaluation and review

# A framework for risk analysis: starting assumptions

- Only interested in costs...
- An existing alternative
- Just my organisation is relevant
- Aim. Maximise expected utility

# Risk analysis framework

- Forecast **costs** under normal circumstances
- Identify hazard events, estimate probabilities and impacts on costs (additional induced costs)
- Forecast costs (a “**mixture**” model). Compute changes in expected utility. If too big,...
- Identify interventions, estimate impact on probabilities and/or costs.
- Compute expected utilities. Choose best intervention (if gain is sufficient)



# Basic setting

- Design given (no interventions, status quo)
- (Random) costs are identified
- Expected utility computed



$$\Psi = \int u(c)\pi(c)dc$$

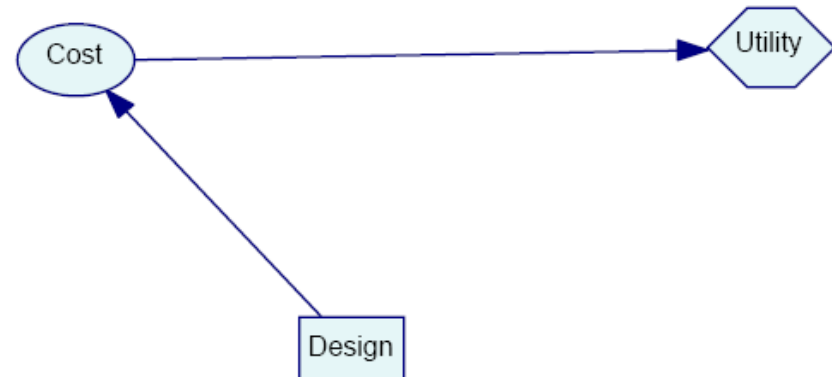
# Basic setting

- Design given



$$\Psi = \int u(c)\pi(c)dc$$

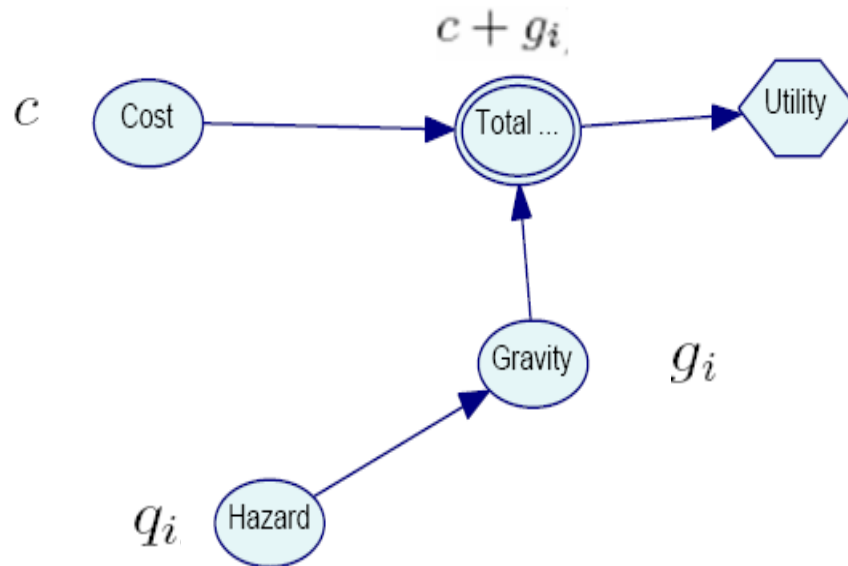
- Including design choice



$$\max_a \Psi(a) = \int u(c)\pi(c|a)dc$$

# Risk assessment

- Likelihood and impact of identified hazards. They



happen with a certain probability and entail an additional cost

- Compute expected utility after risk assessed:

$$\Psi_r = \int \int \int \sum q_i u(c + g_i) \pi(q) \pi(g) dq dg \pi(c) dc$$

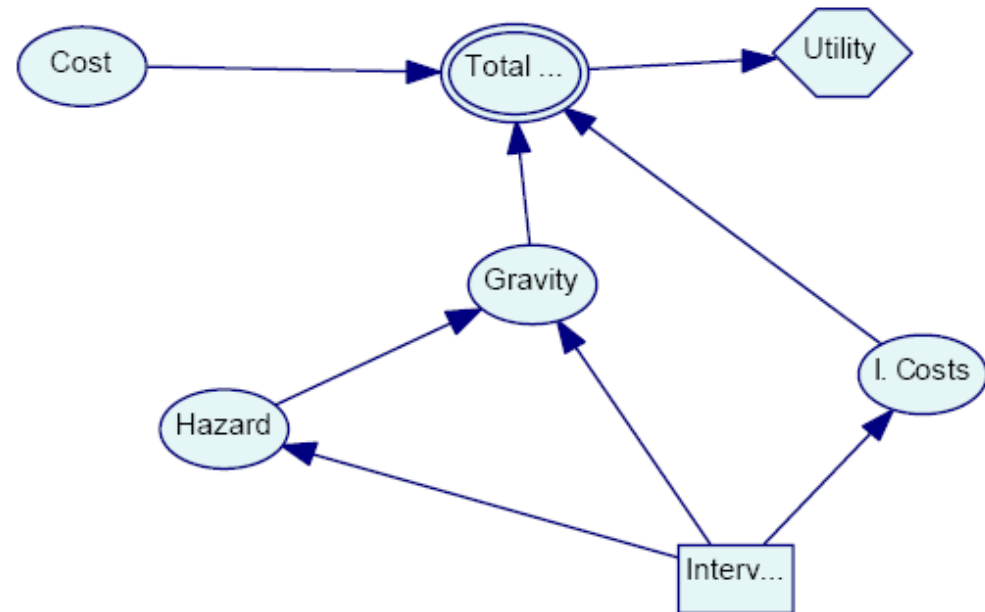
- Impact of risks:  $\Psi - \Psi_r$

If impact is too high, we need to manage risks

# Risk management

- Intervention to be chosen:

Interventions tend to reduce the likelihood of hazard appearance and its gravity... but they also entail a cost

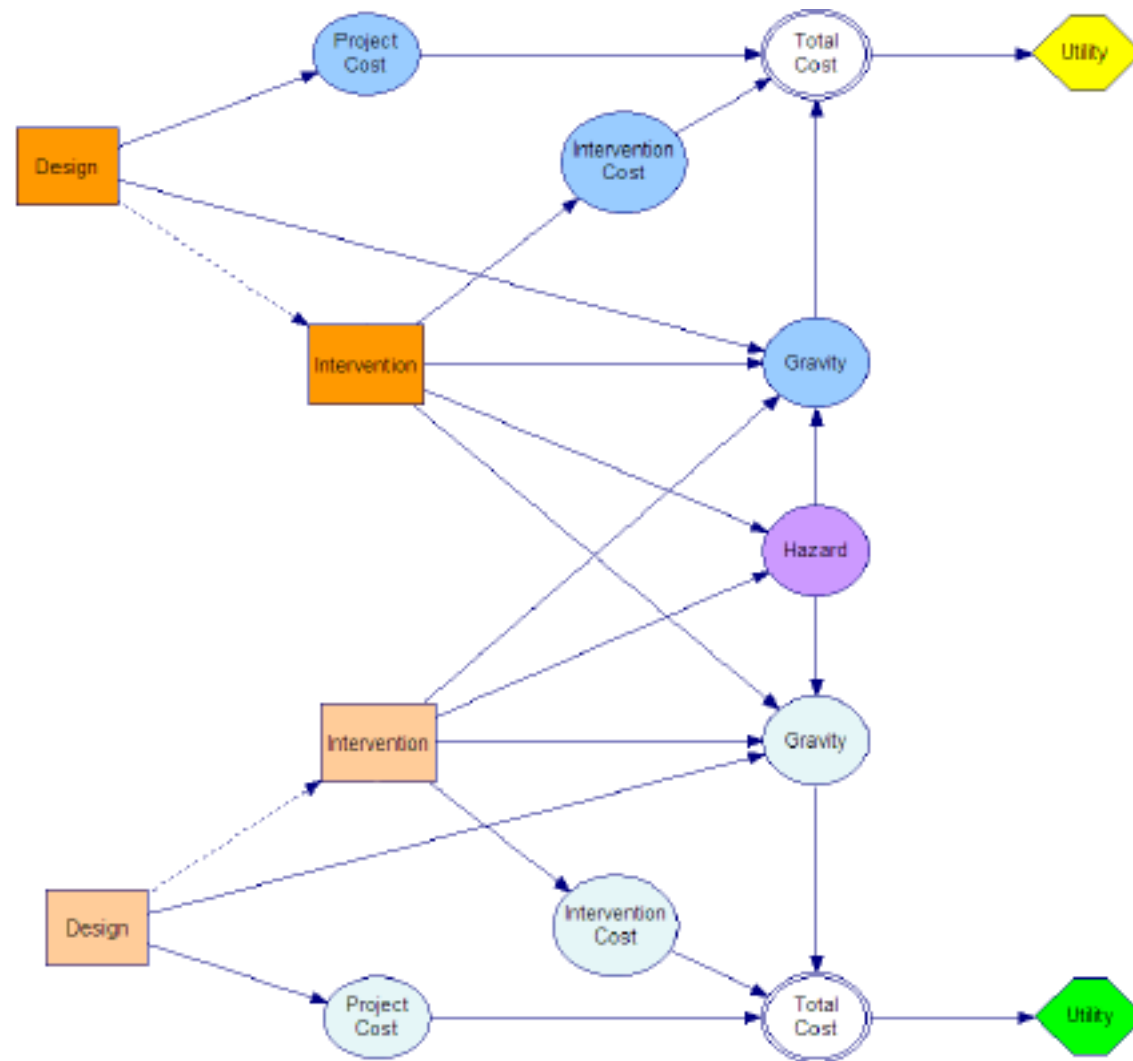


$$\Psi_d = \max_d \Psi_r(d) = \max_d \int \int \int \int \sum q_i u(c + g_i + c_d) \pi(q|d) \pi(g|d) dq dg \pi(c) \pi(c_d) dc_d dc$$

- Gain through managed risk:  $\Psi_d - \Psi_r$

Choose the intervention which provides the biggest gain, if it is sufficiently big...

# Adversarial risk analysis



# Outline

- From risk analysis to adversarial risk analysis
- **Motivation**
- Sequential games
- Simultaneous games
- Auctions
- Security
- Intelligent interfaces
- Challenges

# Adversarial Risk Analysis

- Traditional RA extended to include adversaries ready to increase our risks
- S-11, M-11 lead to large security investments globally, some of them criticised
- Many modelling efforts to efficiently allocate such resources
- Parnell et al (2008) NAS review
  - Standard reliability/risk approaches not take into account intentionality
  - Game theoretic approaches. Common knowledge assumption...
  - Decision analytic approaches. Forecasting the adversary action...
- Merrick, Parnell (2011) review approaches commenting favourably on Adversarial Risk Analysis

# Adversarial Risk Analysis

- A framework to manage risks from actions of intelligent adversaries (DRI, Rios, Banks, JASA 2009)
- One-sided prescriptive support
  - Use a SEU model
  - Treat the adversary's decision as uncertainties
- Method to predict adversary's actions
  - We assume the adversary is a *expected utility maximizer*
    - Model his decision problem
    - Assess his probabilities and utilities
    - Find his action of maximum expected utility
  - But other *descriptive* models are possible
- Uncertainty in the Attacker's decision stems from
  - *our* uncertainty about his probabilities and utilities
  - but this leads to a hierarchy of nested decision problems

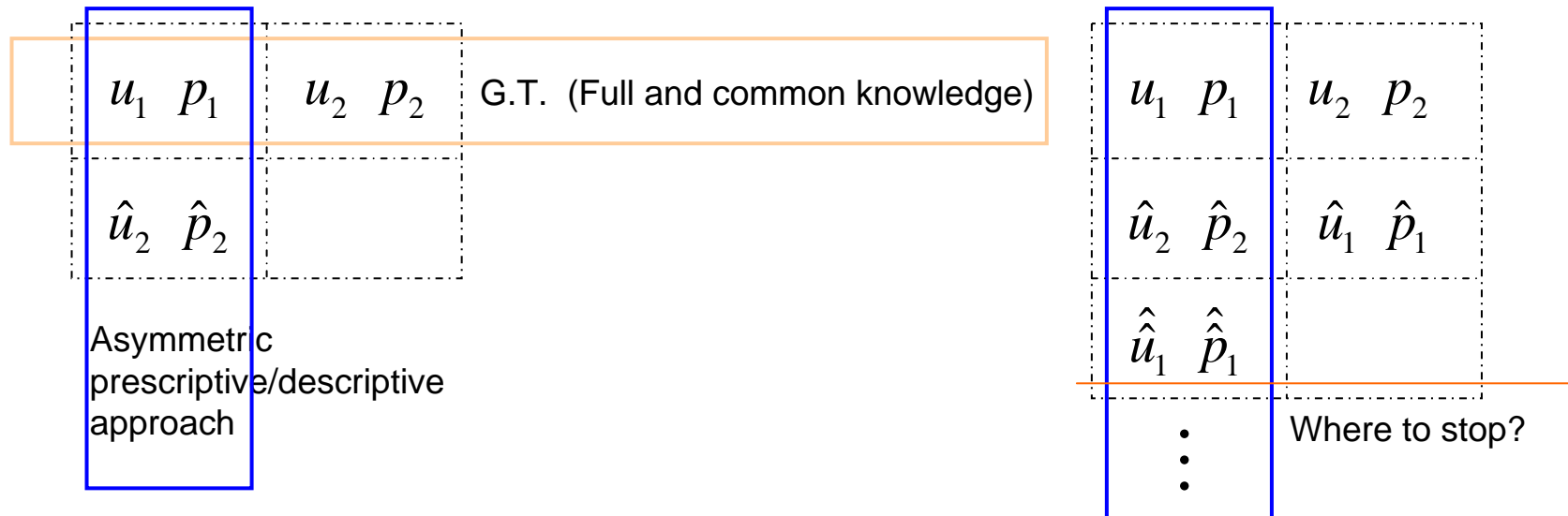
(noninformative, heuristic, mirroring argument) vs (common knowledge)



# Adversarial Risk Analysis

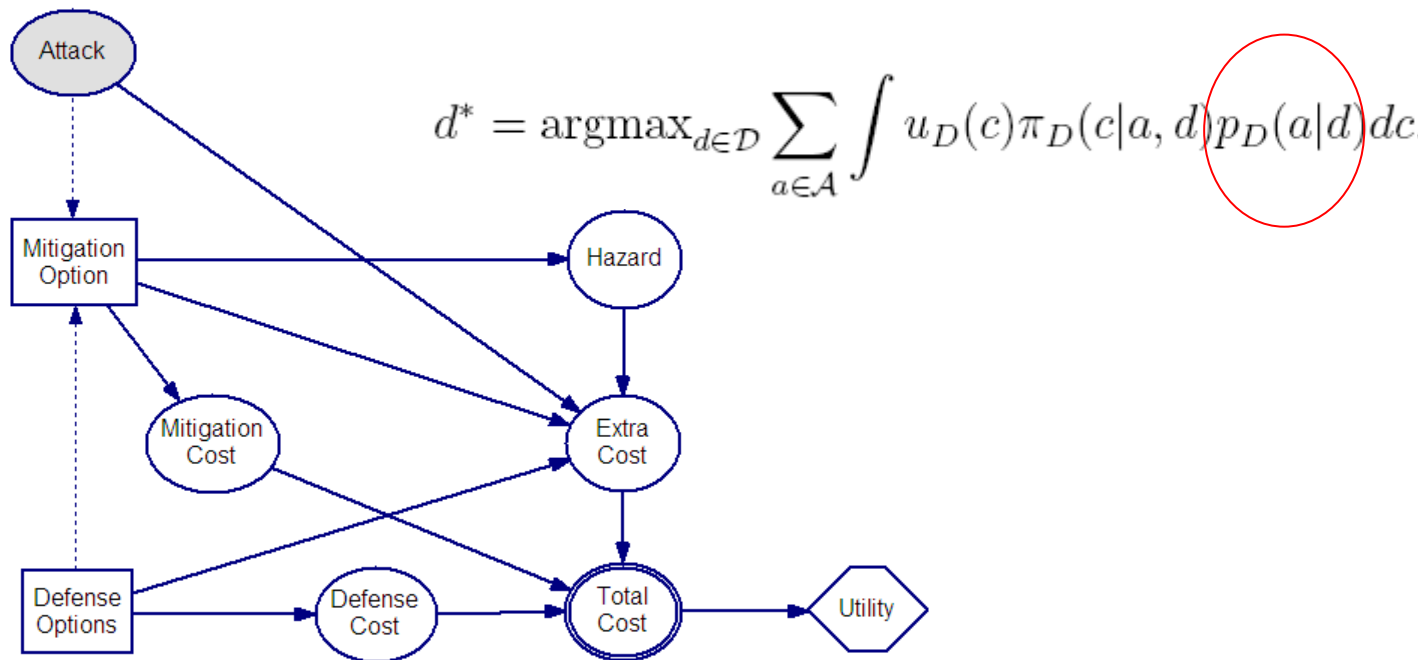
- ARA applications to counterterrorism models (Rios, DRI, 2009, 2012 Risk Analysis)
  - Sequential Defend-Attack
  - Simultaneous Defend-Attack
  - Sequential Defend-Attack-Defend
  - Sequential Defend-Attack with private information
- Somali pirates case (Sevillano, Rios, DRI, 2012 Decision Analysis)
- Routing games (anti IED war) (Wang, Banks, 2011)
- Borel games (Banks, Petralia, Wang, 2011)
- Auctions (DRI, Rios, Banks, 2009; Rothkopf, 2007)
- Kadane, Larkey (1982), Raiffa (1982), Lippman, McCardle (2012)
- Stahl and Wilson (1994, 1995) D. Wolpert (2012)
- Rotschild, MacLay, Guikema (2012)

# Adversarial risk analysis



# Asymmetric prescriptive/descriptive approach

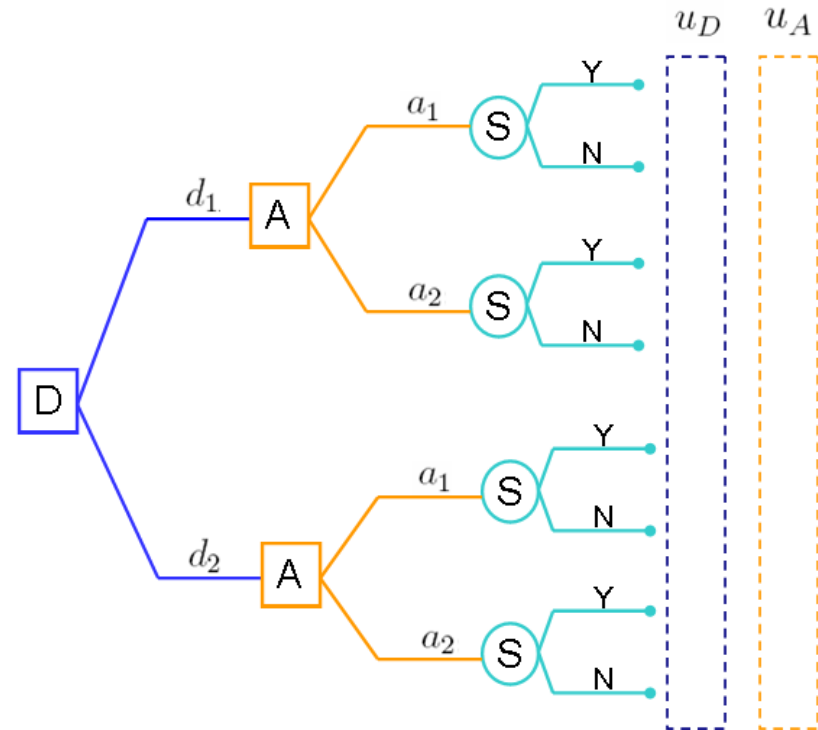
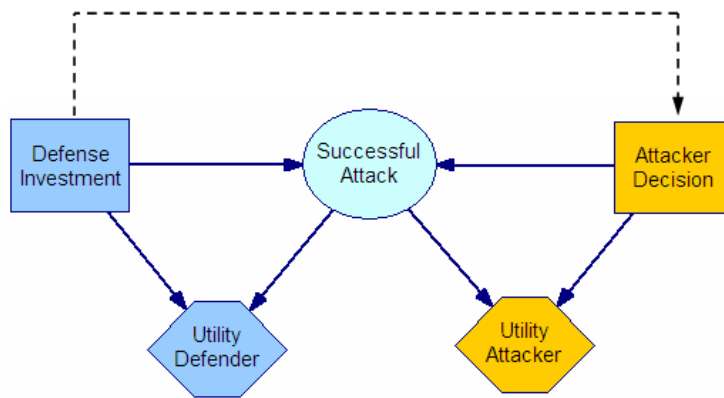
- Bayesian approach (Raiffa, Kadane, Larkey...)
  - Prescriptive advice to one party conditional on a (probabilistic) description of how others will behave
  - Treat the other participant's decisions as uncertain



# Outline

- From risk analysis to adversarial risk analysis
- Motivation
- **Sequential games**
- Simultaneous games
- Auctions
- Security
- Intelligent interfaces
- Challenges

# Sequential moves: First Defender, afterwards Attacker

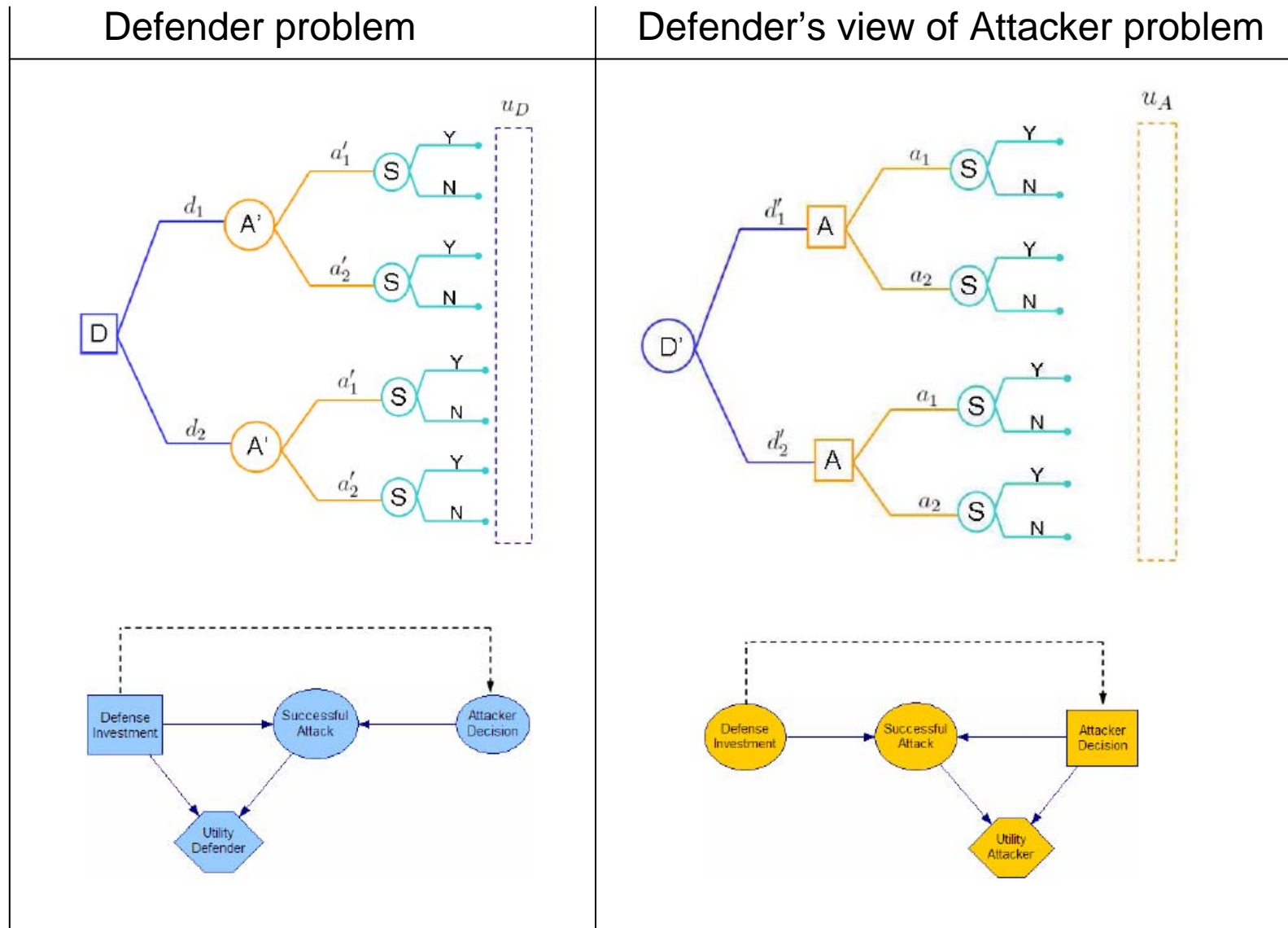


$$a^*(d) = \arg \max_{a \in X_A} u_A(d, a)$$

$$d^* = \arg \max_{d \in X_D} u_A(d, a^*(d))$$

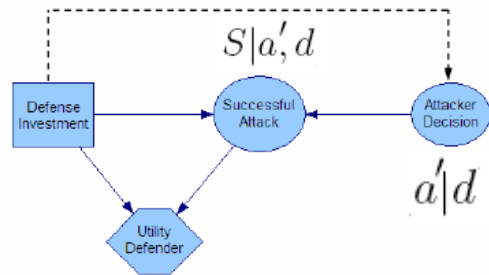
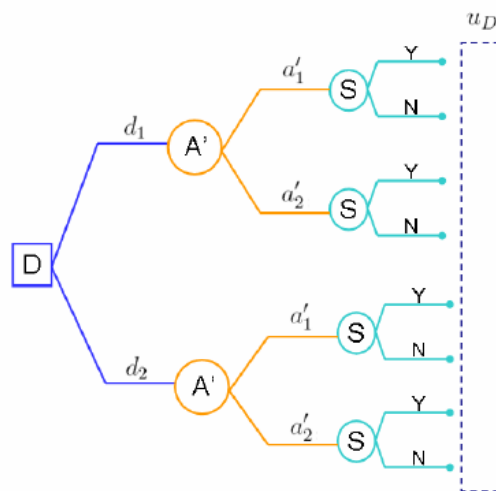
Nash Solution:  $(d^*, a^*(d^*))$

# The sequential game: Supporting the Defender



# Supporting the Defender

## Defender problem



## Defender's solution

$$\psi_D(d, a') = u_D(d, S = Y) p_D(S = Y | X_D = d, X'_A = a') + u_D(d, S = N) p_D(S = N | X_D = d, X'_A = a')$$

$$\psi_D(d) = \psi_D(d, a'_1) p_D(a'_1 | d) + \psi_D(d, a'_2) p_D(a'_2 | d)$$

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

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

# Supporting the Defender: The assessment problem

Defender's view of Attacker problem	Elicitation of $p_D(a' d)$
	<p>A is a EU maximizer</p> <p>D's beliefs about <math>(\hat{u}_A, \hat{p}_A) \sim F</math></p> $\hat{\psi}_A(d', a) = \hat{u}_A(a, S = Y) \hat{p}_A(S = Y   X'_D = d', X_A = a) + \hat{u}_A(a, S = N) \hat{p}_A(S = N   X'_D = d', X_A = a)$ $\hat{\psi}_A \sim \hat{\Psi}_A$ $p_D(a' d) = Pr \left[ a' = \arg \max_{x \in X'_A} \hat{\Psi}_A(d, x) \right]$ <p><u>MC simulation</u></p> $\hat{p}_D(a d) \approx n^{-1} \sum_i \#\{a = \operatorname{argmax}_{x \in \mathcal{A}} \hat{\psi}_A^i(x, d)\}$ <p>where <math>\hat{\psi}_A^i \sim \hat{\Psi}_A, i = 1, \dots, n</math></p>

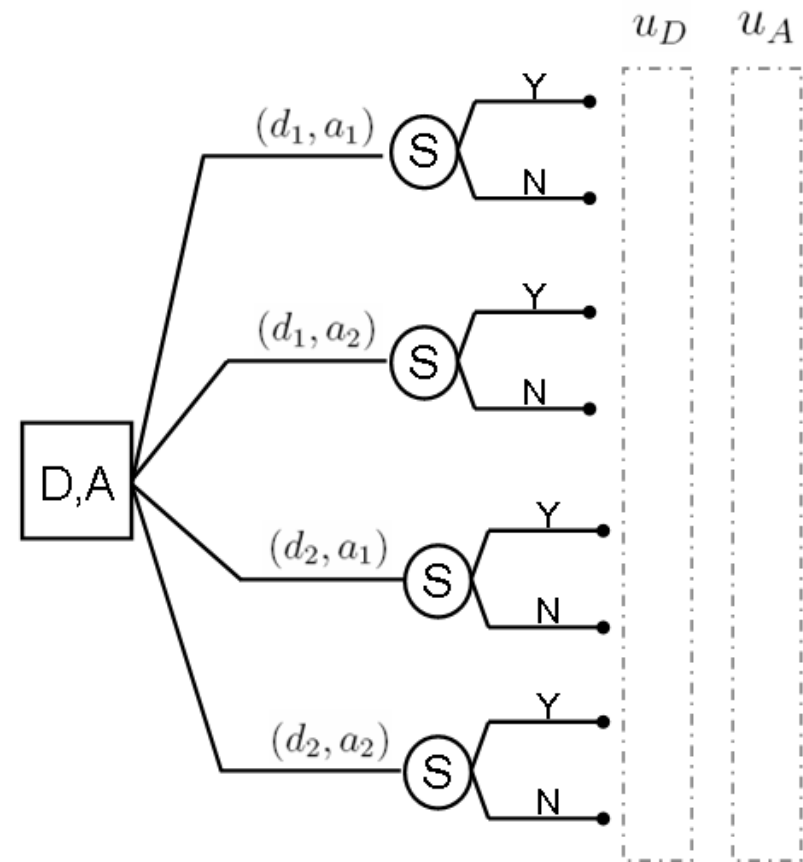
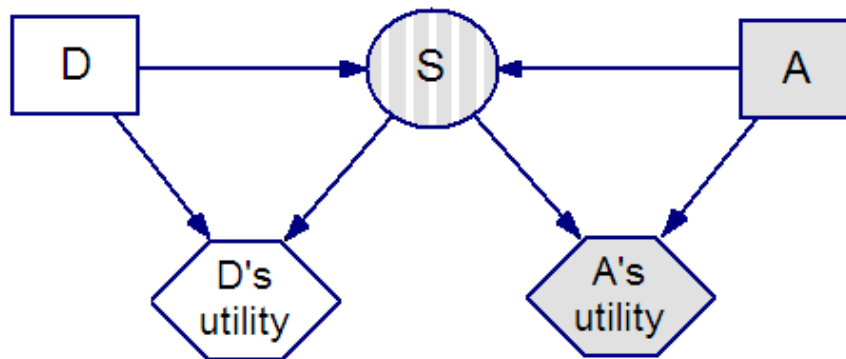


# Outline

- From risk analysis to adversarial risk analysis
- Motivation
- Sequential games
- **Simultaneous games**
- Auctions
- Security
- Intelligent interfaces
- Challenges

# Defend-Attack simultaneous model

- Decisions are made 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

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

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

- 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  $\psi_D(d(\tau_D), a, \tau_D) =$   
$$= \int \left[ \sum_{s \in \mathcal{S}} u_D(d(\tau_D), s, \tau_D) p_D(S = s \mid d(\tau_D), a(\tau_A), \tau_D) \right] \pi(\tau_A \mid \tau_D) d\tau_A$$
- Similarly for Attacker, given his type  $\psi_A(d, a(\tau_A), \tau_A)$

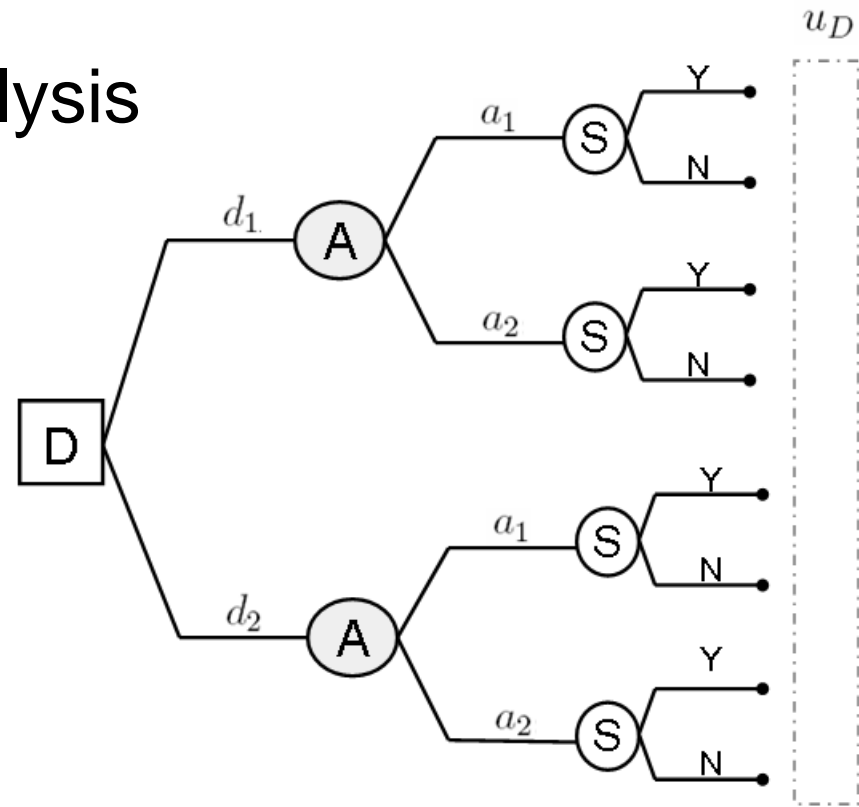
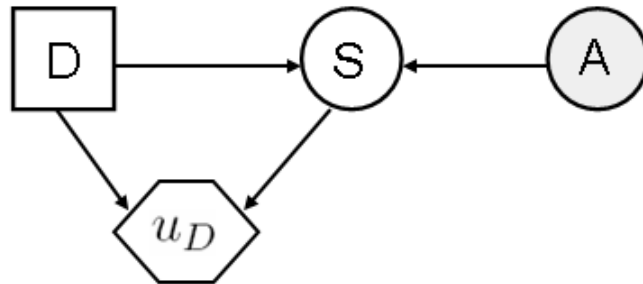
## – Bayes-Nash Equilibrium $(d^*, a^*)$ satisfying

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

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

# Supporting the Defender

- Defender's decision analysis

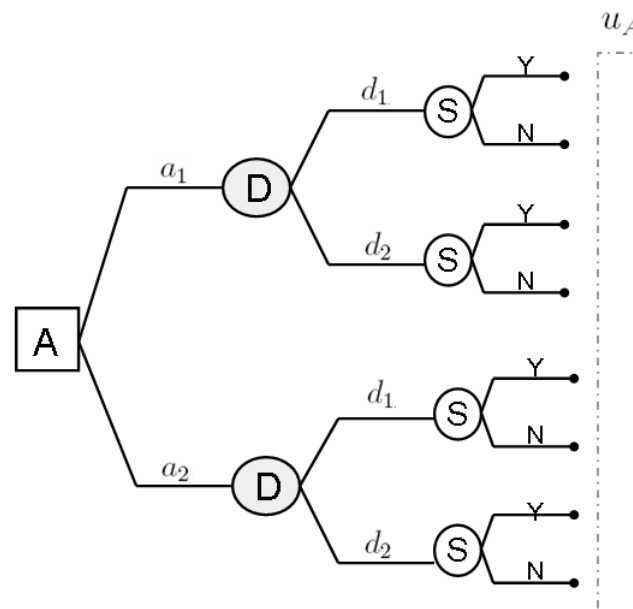
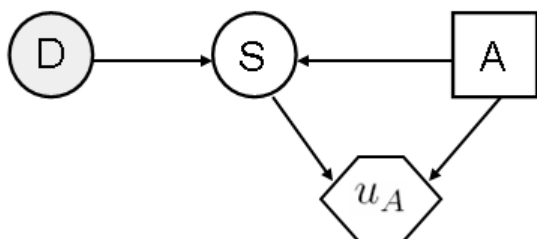


$$d^* = \operatorname{argmax}_{d \in \mathcal{D}} \sum_{a \in \mathcal{A}} \left[ \sum_{s \in \{0,1\}} u_D(d, s) p_D(S = s \mid d, a) \right] \pi_D(A = a)$$

How to elicit it ??

## Assessing $\pi_D(A = a)$

- Attacker's decision analysis as seen by the Defender



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

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

## Assessing $\pi_D(A = a)$

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

- $\Pi_A(D = d)$ 
  - 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

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

# The assessment problem

- To predict Attacker's decision  
The Defender needs to solve Attacker's decision problem  
She needs to assess  $(u_A, p_A, \pi_A)$
- Her beliefs about  $(u_A, p_A, \pi_A)$  are modeled through a probability distribution  $(U_A, P_A, \Pi_A)$
- 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

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

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

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

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

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

# Outline

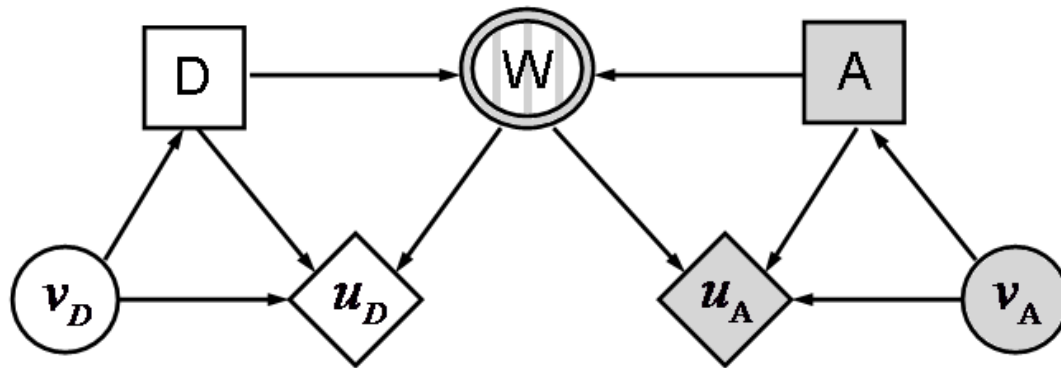
- From risk analysis to adversarial risk analysis
- Motivation
- Sequential games
- Simultaneous games
- **Auctions**
- Security
- Intelligent interfaces
- Challenges

# Bidding in a two-person sealed-bid Auction

- Two sealed bids, the highest one wins
  - Simultaneous decision problem
- The standard Game Theory Analysis
  - D knows  $v_D$  but A does not:  $p_A(v_D)$
  - A knows  $v_A$  but D does not:  $p_D(v_A)$
  - Common knowledge assumption

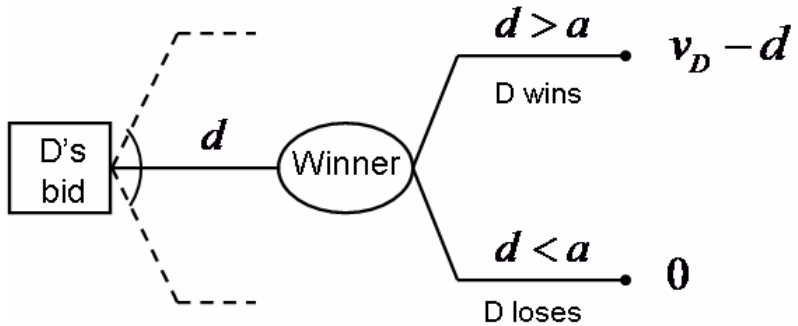
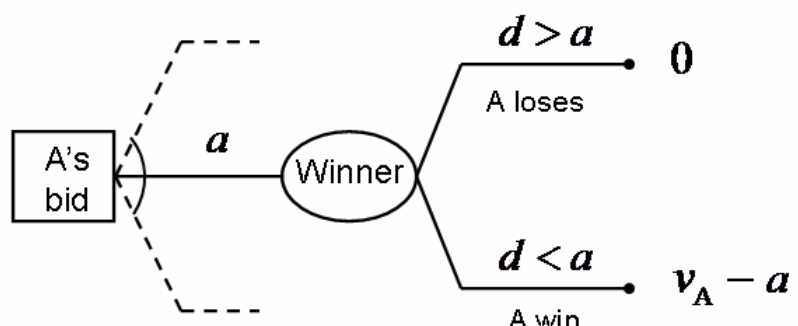
$$p_A(v_D) = p(v_D)$$

$$p_D(v_A) = p(v_A)$$



- Bayesian Nash Eq. (Harsanyi)
- Is it rational that players' beliefs about the opponent's object value will be disclosed??

# Supporting D

D's problem	D's analysis of A's problem
	
$\max_d u_D(v_D - d) \mathbb{P}_D(d > \underline{a}   d)$	$\max_a \hat{u}_A(\hat{v}_A - a) \underbrace{\hat{\mathbb{P}}_A(a > \underline{d}   a)}_{\int_{-\infty}^a \hat{\pi}_A(d) dd}$
<p>??</p> <p><math>(\hat{u}_A, \hat{v}_A, \hat{\mathbb{P}}_A)</math></p>	<p><math>d \sim \hat{\pi}_A</math></p> <p>A's prob. of winning given his bid a</p>

# The assessment problem (III)

- Assessment of  $d \sim \hat{\pi}_A$
- D's analysis of A's analysis of D's problem
  - It leads to a infinite analysis of previous analysis...
- Avoiding infinite regress
  - Available past statistical data (Capen et al, Keefer et al)
  - Expert knowledge
  - Non-informative distribution
  - Heuristic based elicitation (\*)
- Heuristic elicitation  $\hat{\pi}_A(d)$ 
  - Identification of relevant variables in which A can base his assessment of D's bid  $d \sim \hat{\pi}_A$

# Relevant variables

- Auctioned object (true) value for
  - D:  $v_D$
  - A: ?
- D's analysis of A's problem (D's guessed values)

- A's value:  $v_A \square V_A$
- A's guess of ...
  - D's value:  $\hat{v}_D$
  - D's guess of A's value:  $\hat{v}_A$

Used by A to guess D's bid

$$d \sim \hat{\pi}_A$$

as a function of  $\hat{v}_D$  and  $\hat{v}_A$

- Variables that D needs to assess

$v_D$	$v_A$
$\hat{v}_D$	$\hat{v}_A$

# The assessment solution: An heuristic elicitation approach

- D's analysis of A's problem
- Helping D in the assessment of  $\hat{\pi}_A(d)$ 
  - D's analysis of A's analysis about D's bid

$$\hat{\pi}_A(d) = N(\min(\alpha \hat{v}_D, \beta \hat{v}_A), \sigma) \quad \alpha, \beta \in (0, 1) \quad \text{truncated}$$

$$(\alpha, \hat{v}_D, \beta, \hat{v}_A, \sigma) \sim \Pi_A$$

- Assess  $F = (V_A, \Pi_A)$  from D
  - D's uncertainties in her analysis of A's problem

# Outline

- From risk analysis to adversarial risk analysis
- Motivation
- Sequential games
- Simultaneous games
- Auctions
- **Security**
- Intelligent interfaces
- Challenges



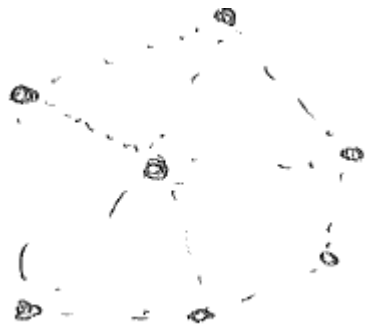
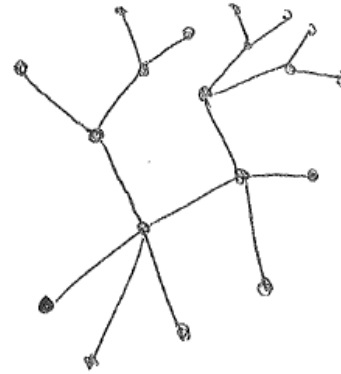
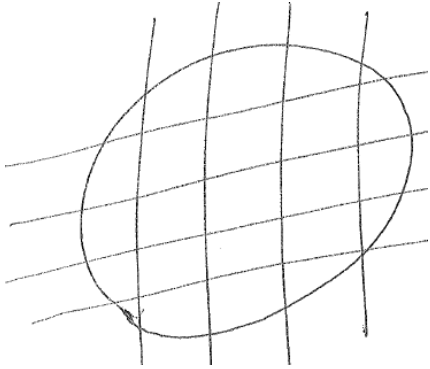
# Security

- One of 'The World's (23) Biggest Problems' (Lomborg, 2008)
  - Arms proliferation
  - Conflicts
  - Corruption
  - Terrorism
  - Drugs
  - Money laundering

# Security

- One of FP7 priorities
- SECONOMICS (2012-2015)
  - Anadolu Airport
  - Barcelona underground
  - National Grid, UK

# Security



# Security from a modelling perspective

- Criminology
- Becker (1968) Economic theory of delict
- Clarke and Cornish (1986) Situational crime prevention. The reasoning criminal
  - Rational Choice in criminology
  - Routine activities theory
  - Delictive pattern theory
  - Problem-oriented policing
- Displacement theory
- Policing performance measures

# Security from a modelling perspective

- COMPSTAT (1994)
- Crime Mapping
- Patrol Car Allocation Models (Tongo, 2010)
- Police Patrol Area Covering Models (Curtin et al, 2007)
- Police Patrol Routes Models (Chawathe, 2007)
- ARMOR at LAX (CREATE, 2007, 2009, 2011)
  
- The Numbers behind NUMB3RS (Devlin, Lorden, 2007)

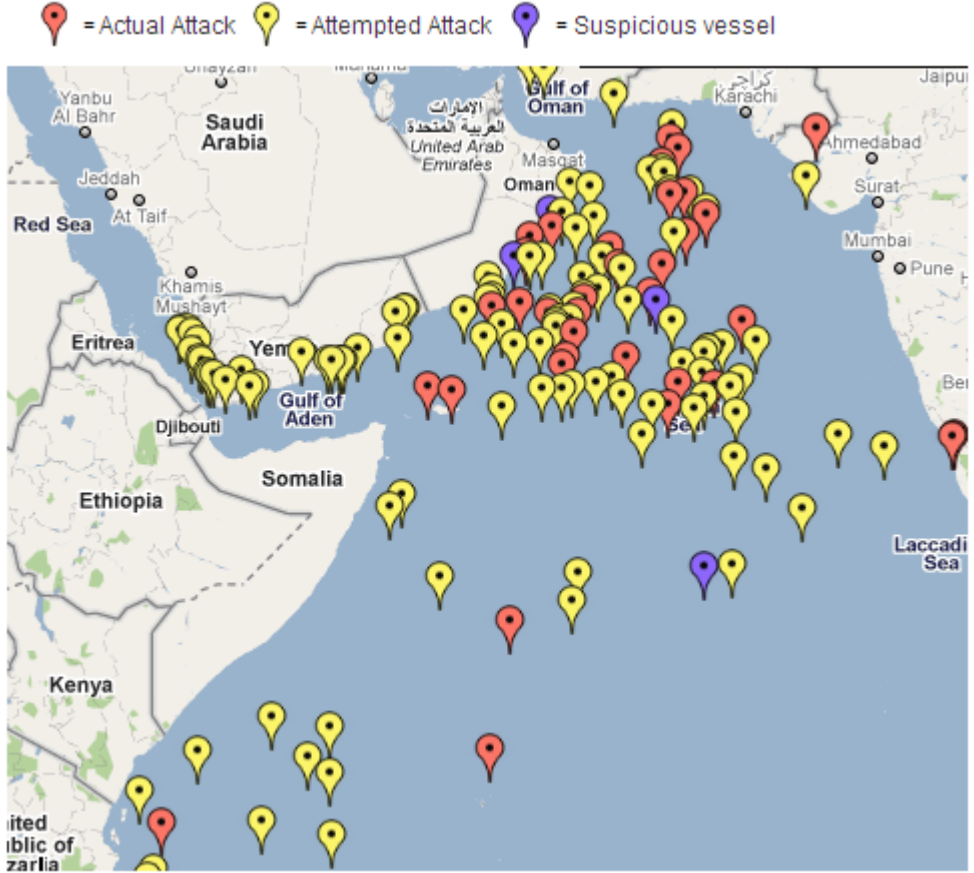
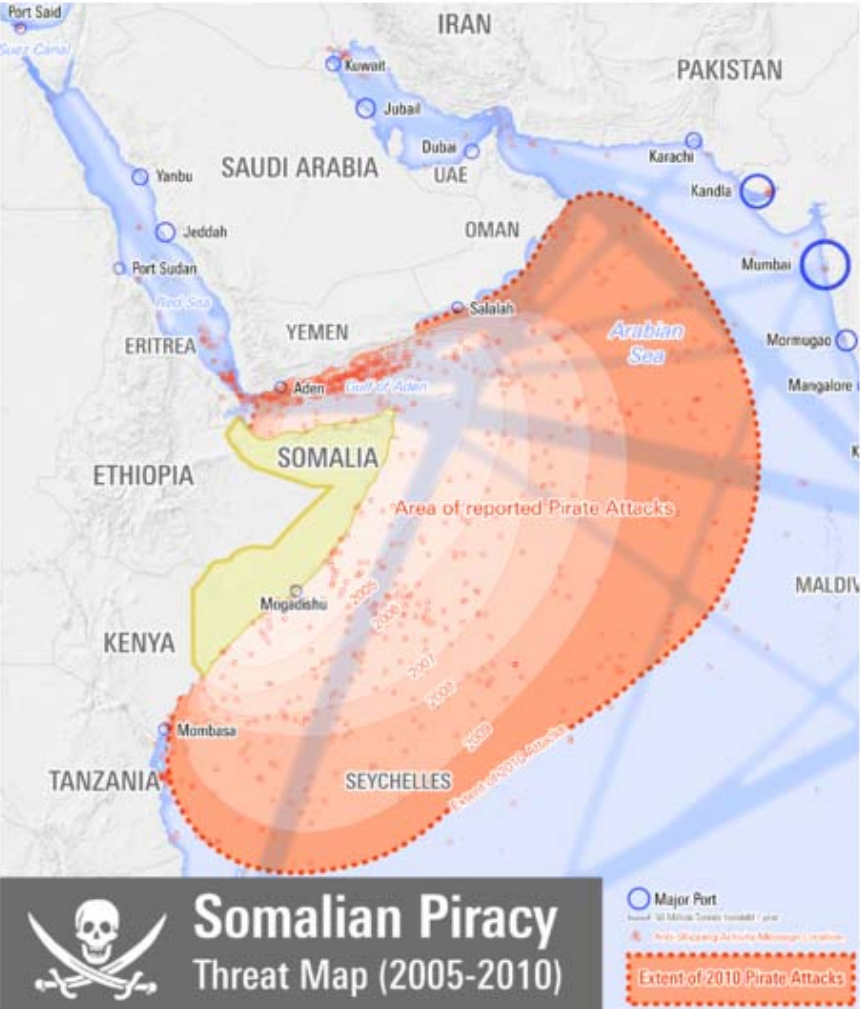
# Piracy in Somalia

- 2010. 1181 Hostages.

2011 (May)

- Worldwide
  - Total Attacks: 211
  - Total Hijackings: 24
- Somalia
  - Total Incidents: 139
  - Total Hijackings: 21
  - Total Hostages: 362
  - Total Killed: 7
  
  - Vessels held by Somali pirates: 26
  - Hostages: 522

# Piracy in Somalia



Piracy and armed robbery incidents reported to the IMB Piracy Reporting Centre  
47 2011

# Piracy in Somalia



Best route between Europe and Asia  
More than 20,000 ships/year passing through  
the Suez Canal

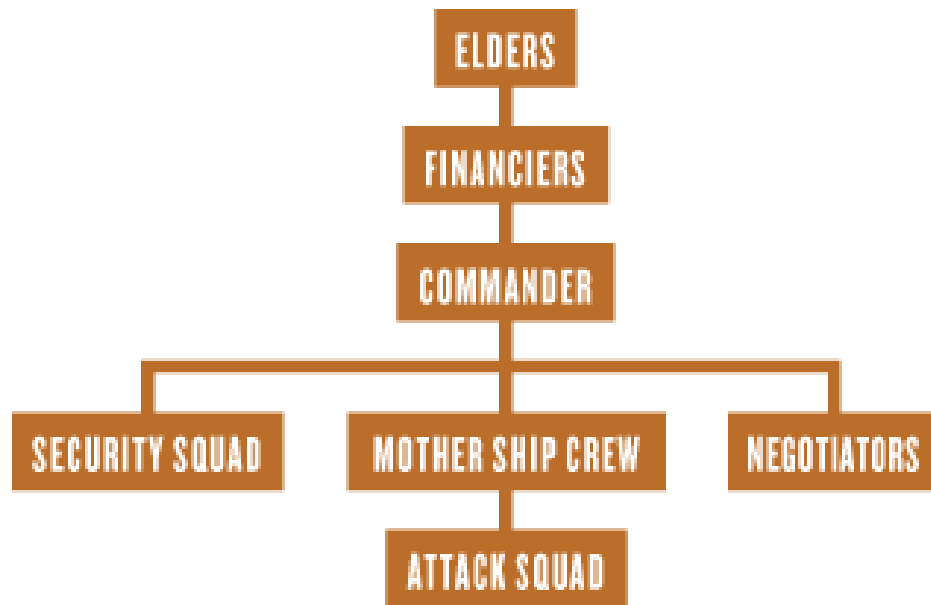
Fishery

48

Volvo Ocean Race 2011



# Piracy in Somalia



*Somalis collect up to \$100M/year from ransoms... Europeans and Asians poach around \$300M a year in fish from Somali waters*

- Cutthroat capitalism. An economic analysis of the Somali pirates business model, Carney (2009) WIRED
- Behind the business plan of Pirates Inc, Siegel (2009) NPR
- Wikipedia page on Piracy in Somalia

# Piracy in Somalia



A major security issue worldwide

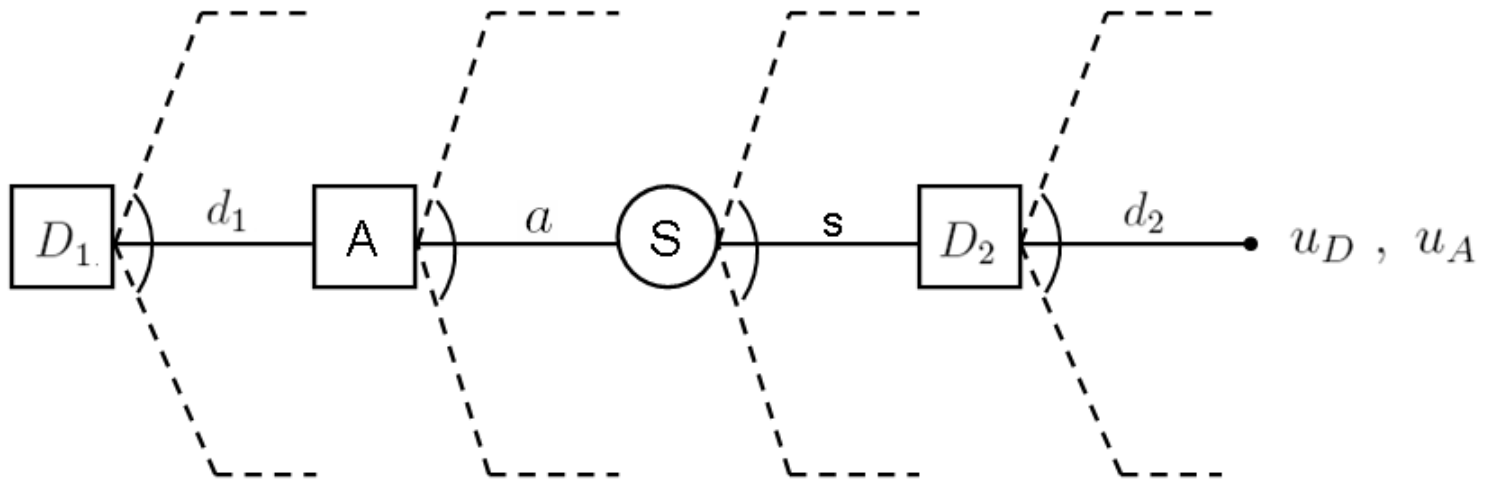
# The Somali Pirates case

- An Illustrative application of ARA
- Support the owner of a Spanish fishing ship managing risks from piracy
- Modeled as a Defend-Attack-Defend decision problem
- Develop predictive models of Pirates' behaviour
  - By thinking about their decision problem

# The Defend–Attack–Defend model

- Two intelligent players
  - Defender and Attacker
- Sequential moves
  - First, Defender moves
  - Afterwards, Attacker knowing Defender's move
  - Afterwards, Defender again responding to attack

# Defend-Attack-Defend model



# ARA:

## Supporting the Defender against the Attacker

- At node  $D_2$

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

- Expected utilities at node S

$$\psi_D(d_1, a) = \int u_D(d_1, s, d_2^*(d_1, s)) p_D(s | d_1, a) ds$$

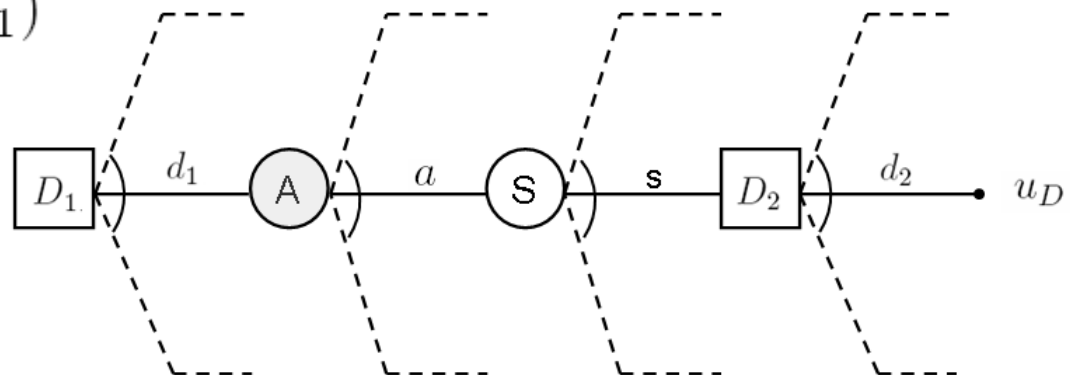
- At node A

$$\psi_D(d_1) = \int \psi_A(d_1, a) p_D(a | d_1) da$$

- Best Defender's decision at node  $D_1$

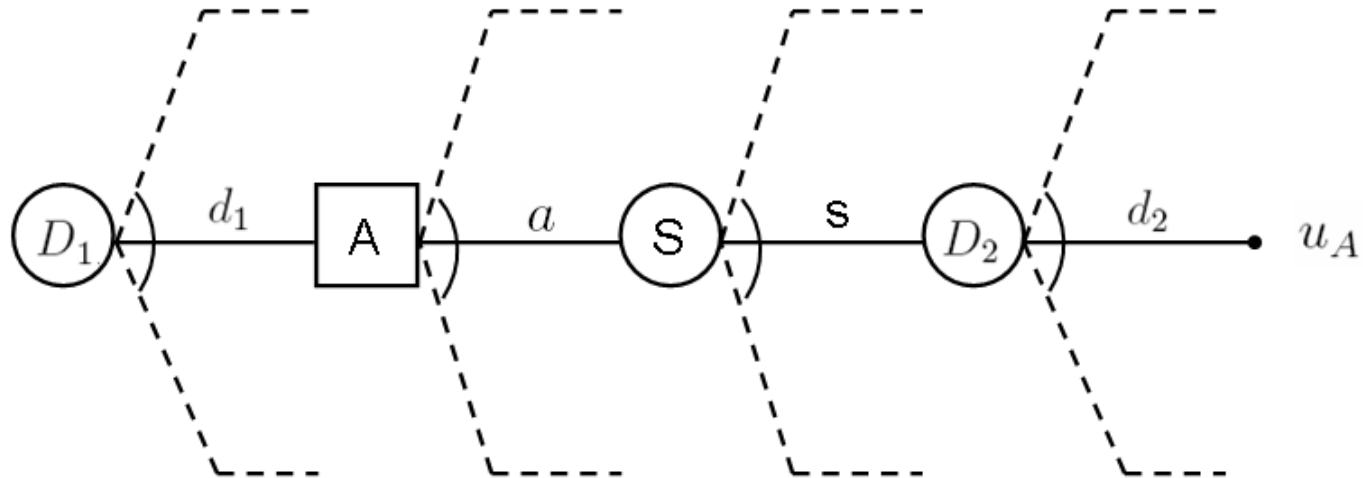
$$d_1^* = \operatorname{argmax}_{d_1 \in \mathcal{D}_1} \psi_D(d_1)$$

- $p_D(A | d_1)$  ??



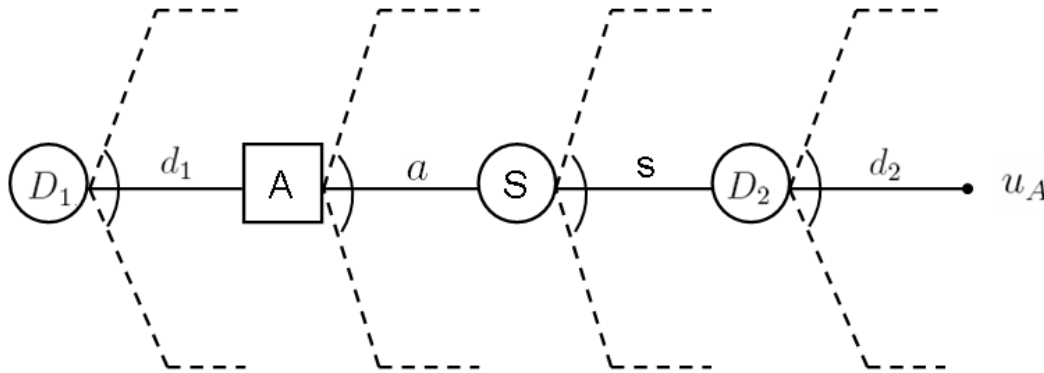
# Predicting $p_D(A | d_1)$

- Attacker's problem as seen by the Defender



## Assessing $p_D(A \mid d_1)$

Given  $F = (U_A(a, s, d_2), P_A(S \mid d_1, a), P_A(D_2 \mid d_1, s))$



- At chance node  $D_2$

$$(d_1, a, s) \rightarrow \Psi_A(d_1, a, s) = \int U_A(a, s, d_2) P_A(D_2 = d_2 \mid d_1, s) dd_2$$

- At chance node  $S$

$$(d_1, a) \rightarrow \Psi_A(d_1, a) = \int \Psi_A(d_1, a, s) P_A(S = s \mid d_1, a) ds$$

- At decision node  $A$

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

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



## Monte-Carlo approximation of $p_D(A | d_1)$

- Drawn  $\{(u_A^i(a, s, d_2), p_A^i(S | d_1, a), p_A^i(D_2 | d_1, s))\}_{i=1}^n \sim F$

- Generate  $\{a_i^*(d_1)\}_{i=1}^n \sim A^*(d_1)$  by

- At chance node  $D_2$

$$(d_1, a, s) \rightarrow \psi_A^i(d_1, a, s) = \int u_A^i(a, s, d_2) p_A^i(D_2 = d_2 | d_1, s) dd_2$$

- At chance node  $S$

$$(d_1, a) \rightarrow \psi_A^i(d_1, a) = \int \psi_A^i(d_1, a, s) p_A^i(S = s | d_1, a) ds$$

- At decision node  $A$

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

- Approximate

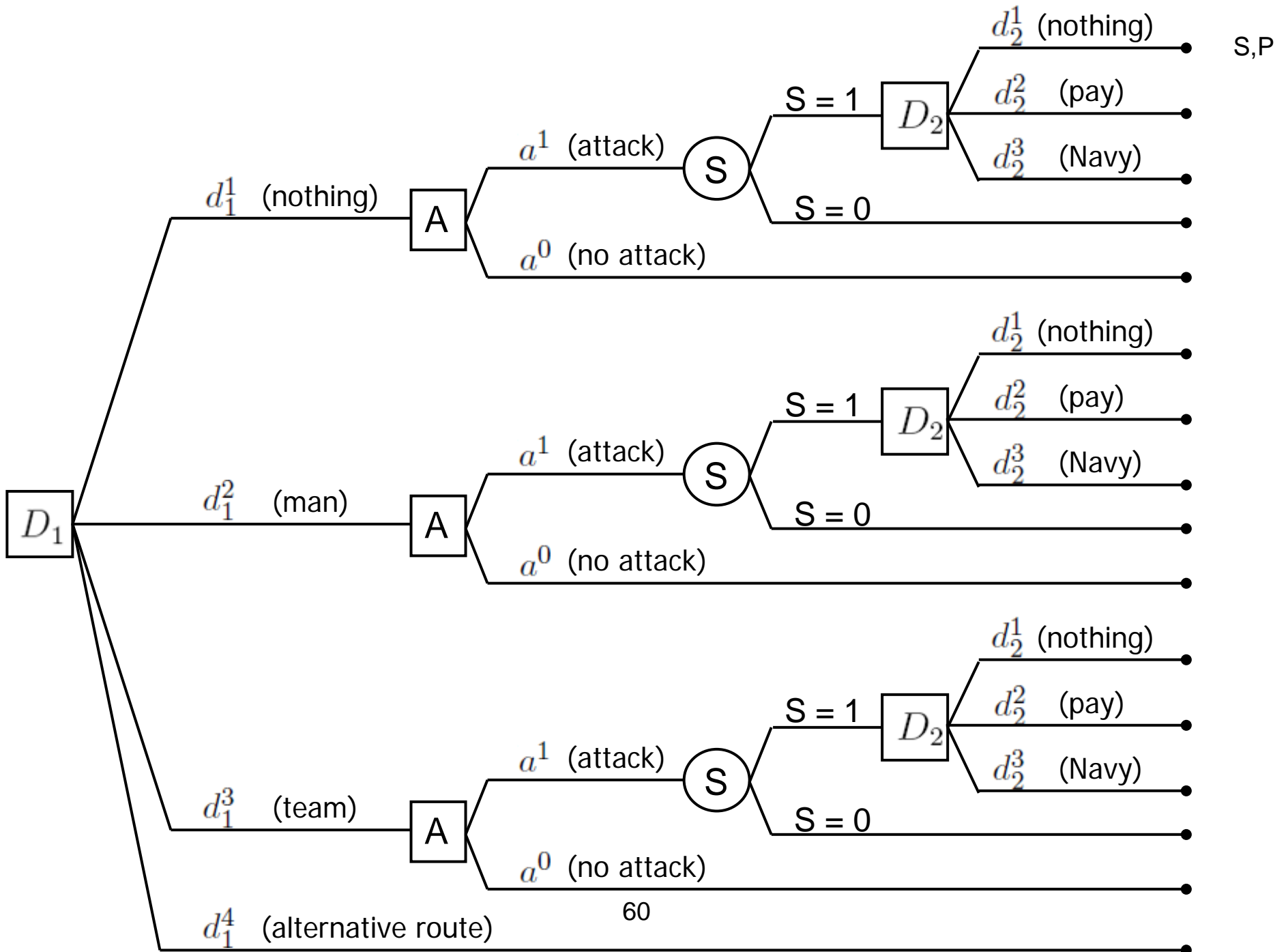
$$p_D(A = a | d_1) \approx \#\{1 \leq i \leq n : a_i^*(d_1) = a\} / n$$

## 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

# The Somali Pirates Case: Problem formulation

- Two players
  - Defender: Ship owner
  - Attacker: Pirates
- Defender first move
  - Do nothing
  - Private protection with an armed person
  - Private protection with a team of two armed persons
  - Go through the Cape of Good Hope avoiding the Somali coast
- Attacker's move
  - Attack or not to attack the Defender's ship
- Defender response to an eventual kidnapping
  - Do nothing
  - Pay the ransom
  - Ask the Navy for support to release the boat and crew



# Defender's own preferences and beliefs

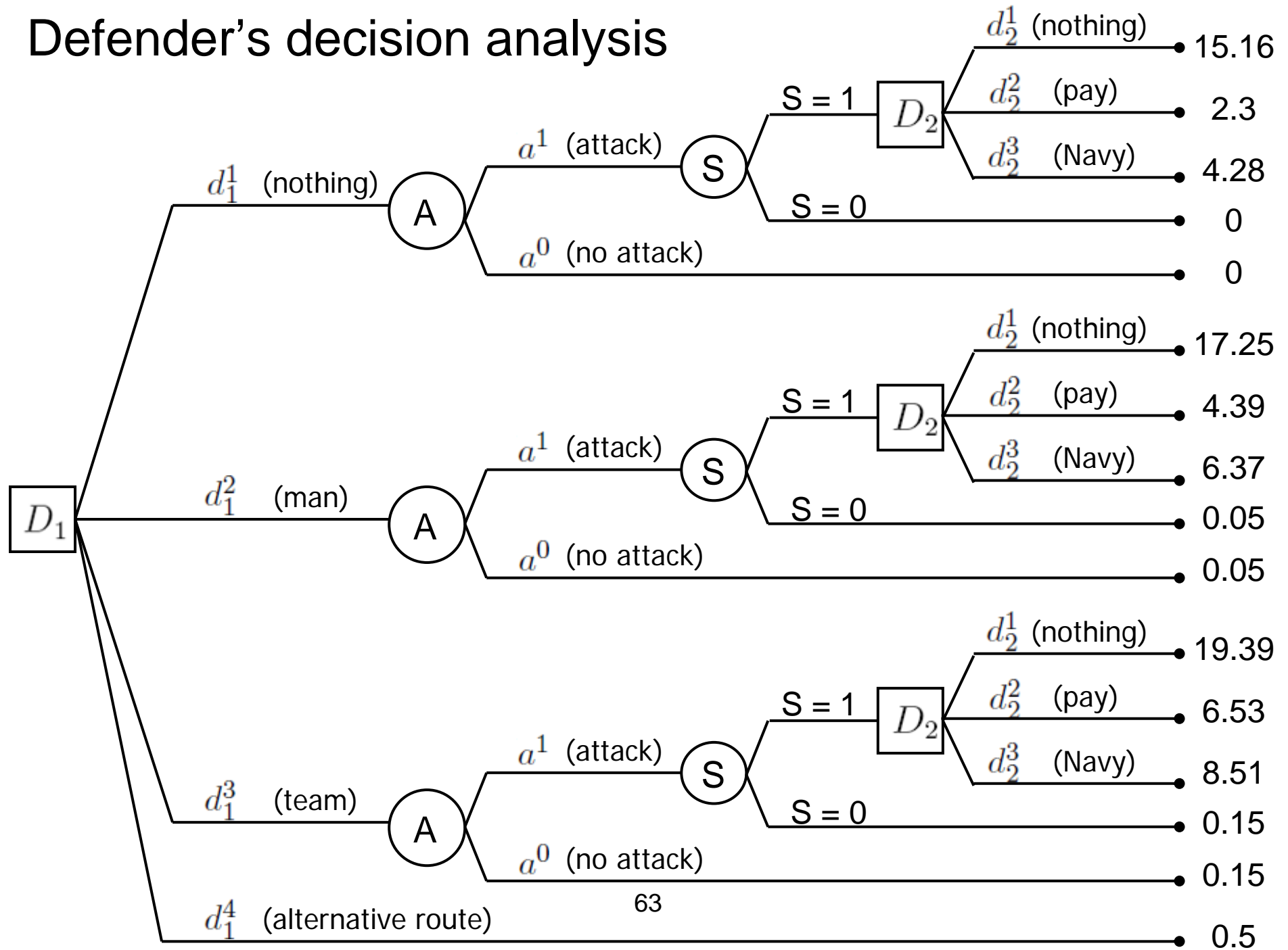
- Assessments from the Defender
  - Multi-attribute consequences
  - Preferences over consequences
  - Beliefs about  $S \mid d_1, a^1$
  - Beliefs about  $A \mid d_1$
- Defender's relevant consequences
  - Loss of the boat
  - Costs of protecting and responding to an eventual attack
  - Number of deaths on her crew
- Defender's monetary values of
  - a Spanish life: 2.04M Euros
  - the ship: 7M Euros

# Defender's own preferences and beliefs

- Consequences of the tree paths for the Defender

$D_1$	$S$	$D_2$	Boat loss	Action costs	Lives lost	Aggregate cost
$d_1^1$ (nothing)	$S = 1$	$d_2^1$ (nothing)	1	0 + 0	0 + 4	15.16
$d_1^1$ (nothing)	$S = 1$	$d_2^2$ (pay)	0	0 + 2.3M	0 + 0	2.3
$d_1^1$ (nothing)	$S = 1$	$d_2^3$ (army)	0	0 + 0.2M	0 + 2	4.28
$d_1^1$ (nothing)	$S = 0$		0	0	0	0
$d_1^2$ (man)	$S = 1$	$d_2^1$ (nothing)	1	0.05M + 0	1 + 4	17.25
$d_1^2$ (man)	$S = 1$	$d_2^2$ (pay)	0	0.05M + 2.3M	1 + 0	4.39
$d_1^2$ (man)	$S = 1$	$d_2^3$ (army)	0	0.05M + 0.2M	1 + 2	6.37
$d_1^2$ (man)	$S = 0$		0	0.05M	0	0.05
$d_1^3$ (team)	$S = 1$	$d_2^1$ (nothing)	1	0.15M + 0	2 + 4	19.39
$d_1^3$ (team)	$S = 1$	$d_2^2$ (pay)	0	0.15M + 2.3M	2 + 0	6.53
$d_1^3$ (team)	$S = 1$	$d_2^3$ (army)	0	0.15M + 0.2M	2 + 2	8.51
$d_1^3$ (team)	$S = 0$		0	0.15M	0	0.15
$d_1^4$ (alternative route)			0	0.5 M	0	0.5

# Defender's decision analysis



# Defender's own preferences and beliefs

- The Defender is constant risk adverse to monetary costs
  - Defender's utility function strategy equivalent to

$$u_D(c_D) = -\exp(c \times c_D), \text{ with } c > 0$$

- We perform sensitivity analysis on “c”
- Defender's beliefs about  $S|a^1, d_1$

$$p_D(S = 1|a^1, d_1^1) = 0.40$$

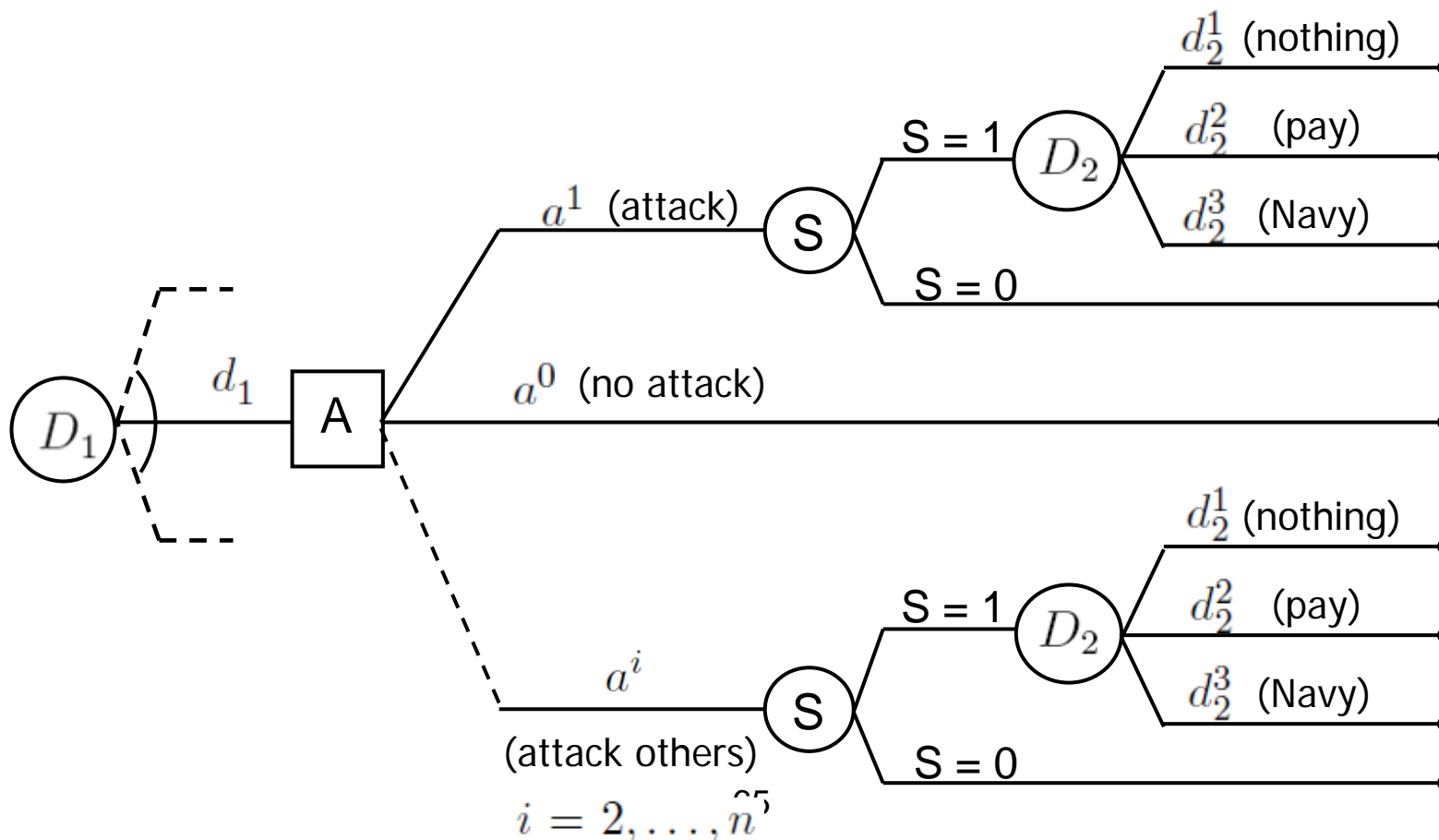
$$p_D(S = 1|a^1, d_1^2) = 0.10$$

$$p_D(S = 1|a^1, d_1^3) = 0.05$$



# Predicting Attacker's behavior

- The objective is to assess  $p_D(A = a^1 | d_1)$
- Attacker's decision problem as seen by the Defender



## Defender's beliefs over the Attacker's beliefs and preferences

- Assess from the Defender the Pirates' preferences  $U_A(a, s, d_2)$
- Perceived relevant consequences for the Pirates
  - Whether they keep the boat
  - Money earned.
  - Number of Pirates' lives lost.

			$c_A(a, s, d_2)$			
$A$	$S$	$D_2$	Boat kept	Profit	Lives lost	Aggregate profit
$a^0$ (no attack)			0	0	0	0
$a^i$ (attack)	$S = 1$	$d_2^1$ (nothing)	1	-0.03M	0	0.97
$a^i$ (attack)	$S = 1$	$d_2^2$ (pay rescue)	0	2.27M	0	2.27
$a^i$ (attack)	$S = 1$	$d_2^3$ (Navy sent)	0	-0.03M	5	-1.28
$a^i$ (attack)	$S = 0$		0	-0.03M	2	-0.53

$i = 1, \dots, n$  (no difference in consequences of attacking the Defender's and other boats)

- The Defender thinks the Pirates are (constant) risk prone for profits
  - Pirates' utility function strategically equivalent to  $U_A(c_A) = \exp(c \times c_A)$ , with  $c \sim \mathcal{U}(0, 20)$
- Defender assessment of Pirates' beliefs on
  - $S \mid a, d_1$ 

$$P_A(S = 1 \mid a^1, d_1^1) \sim \text{Be}(40, 60)$$

$$P_A(S = 1 \mid a^1, d_1^2) \sim \text{Be}(10, 90)$$

$$P_A(S = 1 \mid a^1, d_1^3) \sim \text{Be}(50, 950)$$

$$P_A(S = 1 \mid a^i) \sim \text{Be}(1, 1) \quad \text{for boat } i = 2, \dots, n$$
  - $D_2 \mid d_1, a^1, S=1$ 

$$P_A(D_2 \mid d_1^1, a^1, S = 1) \sim \text{Dir}(1, 1, 1)$$

$$P_A(D_2 \mid d_1^2, a^1, S = 1) \sim \text{Dir}(0.1, 4, 6)$$

$$P_A(D_2 \mid d_1^3, a^1, S = 1) \sim \text{Dir}(0.1, 1, 10)$$
  - $D_2 \mid a^i, S=1$ 

$$P_A(D_2 \mid a^i, S = 1) \sim \text{Dir}(1, 1, \hat{1}) \quad \text{for } i = 2, \dots, n$$

# Predicting Pirates' uncertain behavior

- We use MC simulation to approximate  $p_D(A = a^1 | d_1)$  by

$$\frac{\#\{1 \leq k \leq N : \psi_A^k(d_1, a^1) > \max\{u_A^k(a^0), \psi_A^k(a^2), \dots, \Psi_A^k(a^n)\}\}}{N}$$

- For illustrative purposes, assume that  $n = 4$ 
  - There will be 3 boats (of similar characteristics) at the time the Defender's boat sails through the Gulf of Aden
- *Based on 1000 MC iterations, we have*
  - $\hat{p}_D(A = a^1 | d_1^1) = 0.1931$
  - $\hat{p}_D(A = a^1 | d_1^2) = 0.0181$
  - $\hat{p}_D(A = a^1 | d_1^3) = 0.0002$

# Max EU defense strategy

- *For different risk aversion coefficients “c”*

- $c = 0.1$  and  $c = 0.4$

- $d_1^* = d_1^2$  (protect with an armed man) and  
if kidnapped ( $S = 1$ ), pay the ransom ( $d_2^* = d_2^2$ )

- $c = 2$

- $d_1^* = d_1^4$  (Going through GH Cape)

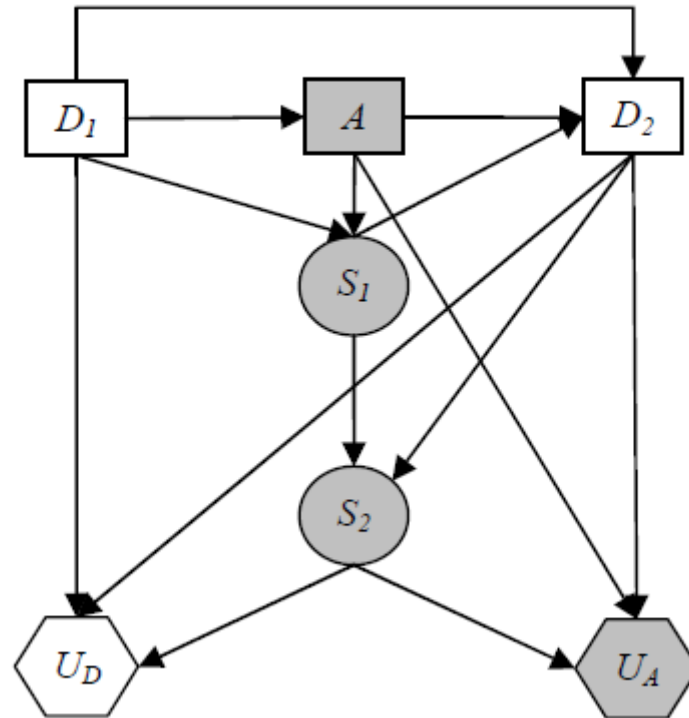
# ARA for Urban Security. Basics

- City divided into cells  $(i,j)$
- Each cell has a value  $v_{ij}$
- Actors
  1. Defender, D, Police. Aims at maintaining value
  2. Attacker, A, Mob. Aims at gaining value
- D allocates resources to prevent  $\sum_{ij} d_{ij}^1 \leq D_1$
- A performs attacks  $\sum_{ij} a_{ij} \leq A$
- D allocates resources to recover  $\sum_{ij} d_{ij}^2 \leq D_2$   
Plus other constraints

# ARA for Urban Security. Basics

At each cell, a  
coupled  
influence diagram

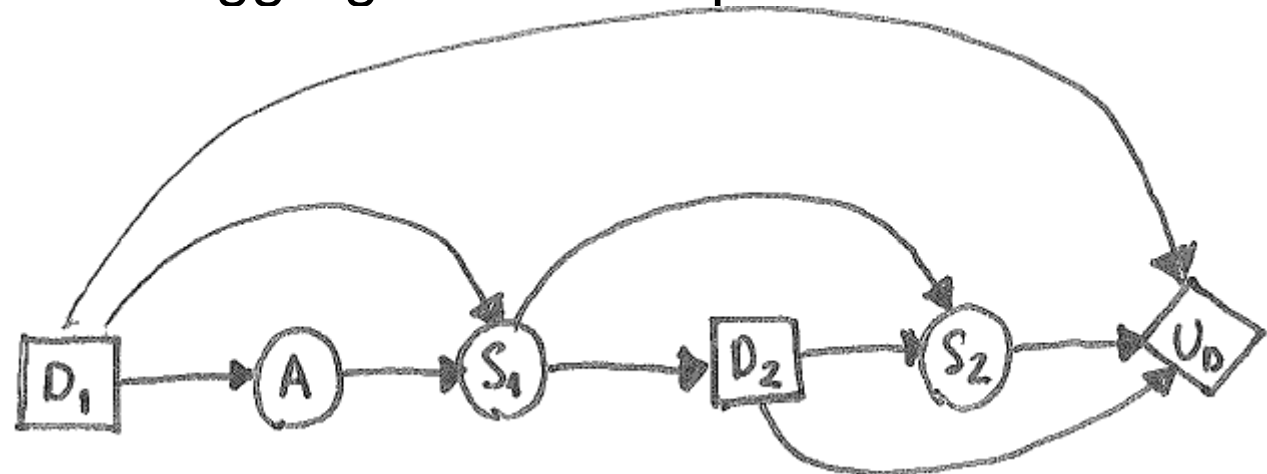
Cell decision making  
coordinated by  
constraints on  
resources



# ARA for Urban Security. Police dynamics

At each cell:

- Makes resource allocation  $d_{ij}^1$
- Faces a level of delinquency  $a_{ij}$  with impact  $s_{ij}^1$
- Recovers as much as she can with resources  $d_{ij}^2$  with a level of success  $s_{ij}^2$
- Gets a consequence
- Aggregates utilities/Aggregates consequences





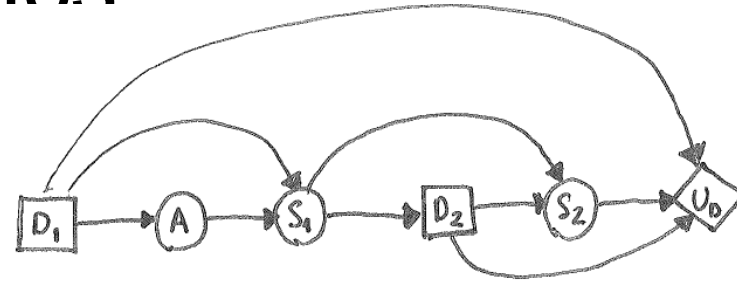
# ARA for Urban Security. Police dynamics

The assessments required from the defender are

- $p_D(a|d_1)$  \*\*\*\*\*
- $p_D(s_1|a, d_1)$
- $p_D(s_2|s_1, d_2)$
- $u_D(d_1, s_2, d_2, v)$

# ARA for Urban Security. Police dynamics

The Police solves sequentially



- At node  $U_D$ ,  $u_D(d_1, s_2, d_2, v)$ .
- At node  $S_2$ , compute  $\psi_D(d_1, s_1, d_2, v) = \int u_D(d_1, s_2, d_2, v) p_D(s_2 | s_1, d_2) ds_2$ .
- At node  $D_2$ , compute  $\psi_D(d_1, s_1, v) = \max_{\sum d_2^{ij} \leq D_2} \psi_D(d_1, s_1, d_2, v)$  and store optimal allocation.
- At node  $S_1$ , compute  $\psi_D(d_1, v, a) = \int \psi_D(d_1, s_1, v) p_D(s_1 | a, d_1) ds_1$ .
- At node  $A$ , compute  $\psi_D(d_1, v) = \int \psi_D(d_1, v, a) p_D(a | d_1) da$
- At node  $D$ , compute  $\psi_D(v) = \max_{\sum d_1^{ij} \leq D_1} \psi_D(d_1, v)$  and store optimal allocation.

$$\max_{\sum d_1^{ij} \leq D_1} \max_{\sum d_2^{ij} \leq D_2} \int \int \int u_D(d_1, s_2, d_2, v) p_D(s_2 | s_1, d_2) p_D(s_1 | a, d_1) p_D(a | d_1) ds_2 ds_1 da$$

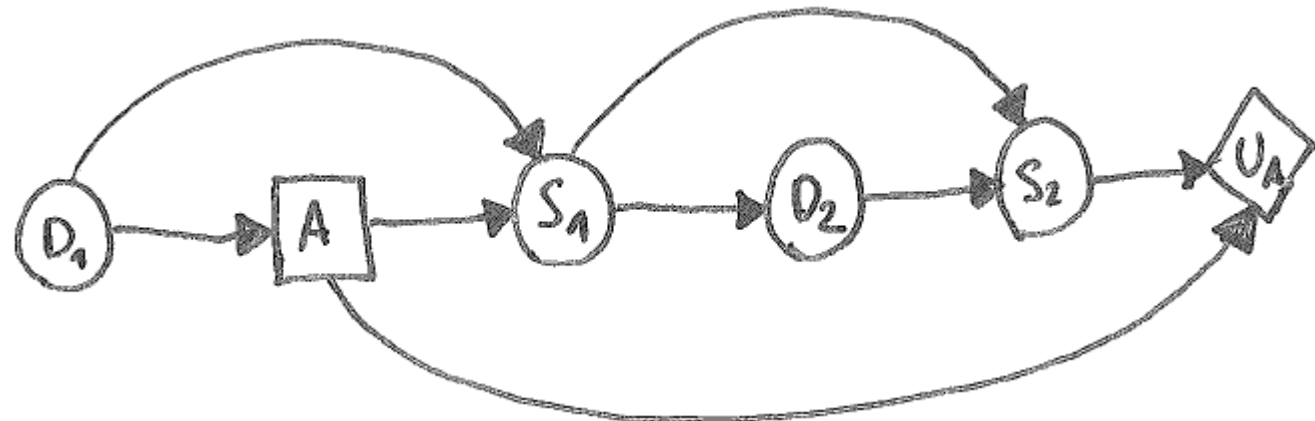
Augmented probability simulation (Bielza, Muller, DRI, 1999 Mansci)

$$p_D(a | d_1)$$

# ARA for Urban Security. Mob dynamics

At each cell:

- Observes resource allocation  $d_{ij}^1$
- Undertakes attack  $a_{ij}$ , with impact  $s_{ij}^1$
- Observes recovery with resources  $d_{ij}^2$  with a level of success  $s_{ij}^2$
- Gets a consequence
- Aggregates utilities/Aggregates consequences



# ARA for Urban Security. Mob Dynamics

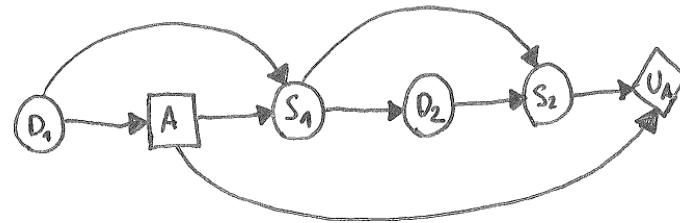
- The assessments for the Mob are
  - $p_A(d_2|s_1)$
  - $p_A(s_1|a, d_1)$
  - $p_A(s_2|s_1, d_2)$
  - $u_A(a, s_2, v)$

- We model our uncertainty about them through

- •  $P_A(d_2|s_1)$
- $P_A(s_1|a, d_1)$
- $P_A(s_2|s_1, d_2)$
- $U_A(a, s_2, v)$

# ARA for Urban Security. Mob dynamics

- We propagate such uncertainty through the scheme



- At node  $U_A$ ,  $U_A(a, s_2, v)$ .
- At node  $S_2$ , compute  $\Psi_D(a, s_1, d_2, v) = \int U_A(a, s_2, v) P_A(s_2 | s_1, d_2) ds_2$ .
- At node  $D_2$ , compute  $\Psi_D(a, s_1, v) = \int \Psi_D(a, s_1, d_2, v) P_A(d_2 | s_1) dd_2$
- At node  $S_1$ , compute  $\Psi_D(d_1, v, a) = \int \Psi_D(a, s_1, v) P_A(s_1 | a, d_1) ds_1$ .
- At node  $A$ , compute  $\Psi_D(d_1, v) = \max_{\sum a^{ij} \leq A} \Psi_D(d_1, v, a)$  and stores optimal random allocation  $A^*(d_1, v)$ .

$$\int_{\{x \leq a\}} p_D(A = x | d_1, v) dx = Pr(A^*(d_1, v) \leq a)$$

# ARA for Urban Security. Mob dynamics

- We can estimate it by Monte Carlo
- Sample from

$$F = \{U_A(a, s_2, v), P_A(s_1 | a, d_1), P_A(d_2 | s_1), P_A(s_2 | s_1, d_2)\}$$

- Solve for maximum expected utility attack (EU computed in one step+ augmented prob. Simulation)

$$\hat{P}_D(A \leq a | d_1) = \frac{\#\{A_k^*(v, d_1) \leq a\}}{n}$$

# ARA for Urban Security. Mob dynamics

Inicializar parámetros

Generar la estructura del ataque  $\{d_1, a, s_1, d_2\}$  y  $P_A^i(d_2 | d_1, a, s_1)$

1. Para el Atacante, desde  $i = 1, 2, \dots, N$  repetir

En el nodo  $S_2$  y  $\forall d_1, a, s_1, d_2$  factibles

Generar  $P_A^i(s_2 | s_1, d_2)$

$$\text{Obtener } \Psi_A^i(a, s_1, d_2, v) = \sum_{s_2} U_A^i(a, s_2, v) \prod_j P_A^i(s_j^2 | s_j^1, d_j^2)$$

En el nodo  $D_2$  y  $\forall d_1, a, s_1$  factibles

$$\text{Obtener } \Psi_A^i(d_1, a, s_1, v) = \sum_{d_2} \Psi_A^i(a, s_1, d_2, v) P_A^i(d_2 | d_1, a, s_1)$$

En el nodo  $S_1$  y  $\forall d_1, a$  factibles

Generar  $P_A^i(s_1 | d_1, a)$

$$\text{Obtener } \Psi_A^i(d_1, a, v) = \sum_{s_1} \Psi_A^i(d_1, a, s_1, v) \prod_j P_A^i(s_j^1 | d_j^1, a_j)$$

En el nodo  $A$  y  $\forall d_1$  factible

$$\text{Obtener } (d_1, v) \rightarrow A_i^*(d_1, v) = \arg \max_{a \in A} \Psi_A^i(d_1, a, v)$$

2. Aproximar  $P_D(a | d_1)$  mediante

$$\hat{P}_D(a | d_1) = \frac{\#\{A_i^*(d_1, v) = a\}}{n}$$

3. Para el Defensor, hacer

En el nodo  $S_2$ ,  $\forall d_1, a, s_1, d_2$

$$\text{Obtener } \Psi_D(d_1, s_1, d_2, v) = \sum_{s_2} u_D(s_2, v) \prod_j p_D(s_j^2 | s_j^1, d_j^2)$$

En el nodo  $D_2$ ,  $\forall d_1, a, s_1$

$$\text{Obtener } \Psi_D(d_1, s_1, v) = \arg \max_{d_2} \Psi_D(d_1, s_1, d_2, v) \text{ y guardar } d_2^*(d_1, a, s_1)$$

En el nodo  $S_1$ ,  $\forall d_1, a$

$$\text{Obtener } \Psi_D(d_1, a, v) = \sum_{s_1} \Psi_D(d_1, s_1, v) \prod_j p_D(s_j^1 | d_j^1, a_j)$$

En el nodo  $A$ , obtener  $\forall d_1$

$$\text{Obtener } \Psi_A(d_1, v) = \sum_a \Psi_D(d_1, a, v) p_D(a | d_1)$$

En el nodo  $D_1$

$$\text{Obtener } \Psi_D(v) = \arg \max_a \Psi_A(d_1, v) \text{ y guardar } d_1^*$$

# Example

1	2	3

1	2	3
1,0	0,75	2,0

1	2	3
1	1	2

1	2	3
4	0	0

1	2	3
2	0	0

1	2	3
1	1	0

1	2	3
$d_1^1$	$d_2^1$	$d_3^1$

1	2	3
$d_1^2$	$d_2^2$	$d_3^2$

$$\sum_j d_j^2 \leq 3$$

$$d_j^2 \geq 0, d_j^2 \text{ entero } ,y,$$

$$d_1^2 \leq d_1^1 + d_2^1$$

$$d_2^2 \leq d_1^1 + d_2^1 + d_3^1$$

$$d_3^2 \leq d_2^1 + d_3^1$$



# Example

		A		
		0	1	2
D	0	0,0	0,85	0,95
	1	0,0	0,6	0,75
	2	0,0	0,3	0,5
	3	0,0	0,05	0,1
	4	0,0	0,0	0,05

Figura 9:  
 $p_D(s_1 = 1 | a, d_1)$

		0 a 4	0	1	2	3	4	$d_2$
		0	1					$s_1$
$s_2$	0	1,0	0,0	0,05	0,2	0,4	0,6	
	1	0,0	1,0	0,95	0,8	0,6	0,4	

Figura 10:  $p_D(s_2 | s_1, d_2)$

$$u_D = -\exp\left[c \sum_j v_j (1 - \rho_j)\right], c > 0$$

# Example

$P_A(s_1 | d_1, a) \sim \beta e(\alpha, \beta)$ , de modo que  $E(P_A(s_1 | d_1, a)) = p_D(s_1 | d_1, a)$  y  $\sigma(P_A(s_1 | a, d_1)) = 0,05$ .

$$P_A(s_2 | s_1, d_2)$$

- si  $s_j^1 = 0$ , entonces  $d_j^2 \leq d_j^1$  (a menos que las unidades  $d_j^1$  sean requeridas en otros lugares), con probabilidad 1.
- si  $s_j^1 = 1$ , entonces hacemos  $d_j^2 \geq a_j$  (moviendo los recursos a esa celda). Asumimos respuestas proporcionadas en el sentido de que

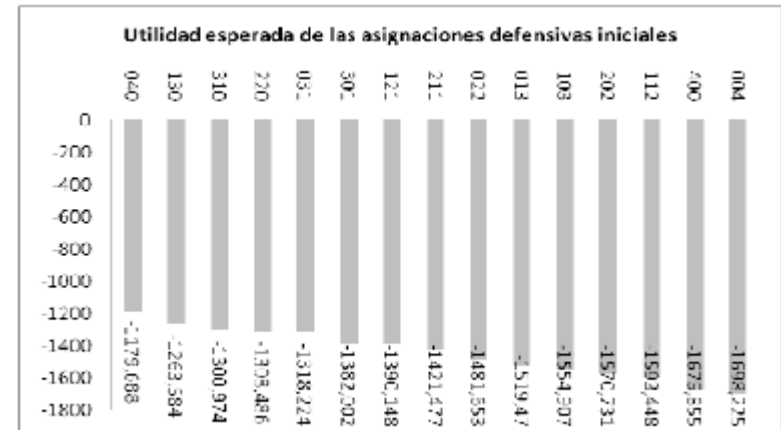
- $d_j^2 = a_j$  con probabilidad 0,5
- $d_j^2 = a_j + 1$  con probabilidad 0,25
- $d_j^2 = a_j + 2$  con probabilidad 0,125
- ...

$$P_A(d_2 | d_1, a, s_1) = \prod_j P_A(d_j^2 | d_j^1, a_j, s_j^1)$$

$$u_A = \exp \left[ c \sum_j (\rho v_j - a_j k) \right].$$

# Example

d1	a	000	001	002	010	011	020	100	101	110	200
004		0,46	0	0	0,54	0	0	0	0	0	0
013		0,03	0	0	0	0	0	0,71	0	0,26	0
022		0	0,03	0,01	0	0,01	0	0,48	0,47	0	0
031		0	0,25	0,01	0	0,03	0	0,01	0,7	0	0
040		0	0,22	0,02	0	0	0	0	0,76	0	0
103		0,07	0	0	0,28	0	0	0,24	0	0,41	0
112		0,13	0,2	0,03	0,01	0,05	0	0,17	0,3	0,11	0
121		0	0,38	0,05	0	0,09	0	0	0,48	0	0
130		0	0,53	0	0	0,01	0	0	0,46	0	0
202		0,07	0,16	0,08	0,26	0,37	0	0	0,06	0	0
211		0	0,45	0,09	0	0,3	0	0	0,16	0	0
220		0	0,65	0,03	0	0,14	0	0	0,18	0	0
301		0	0,43	0	0	0,52	0	0	0,05	0	0
310		0	0,71	0	0	0,24	0	0	0,05	0	0
400		0,35	0	0	0,65	0	0	0	0	0	0



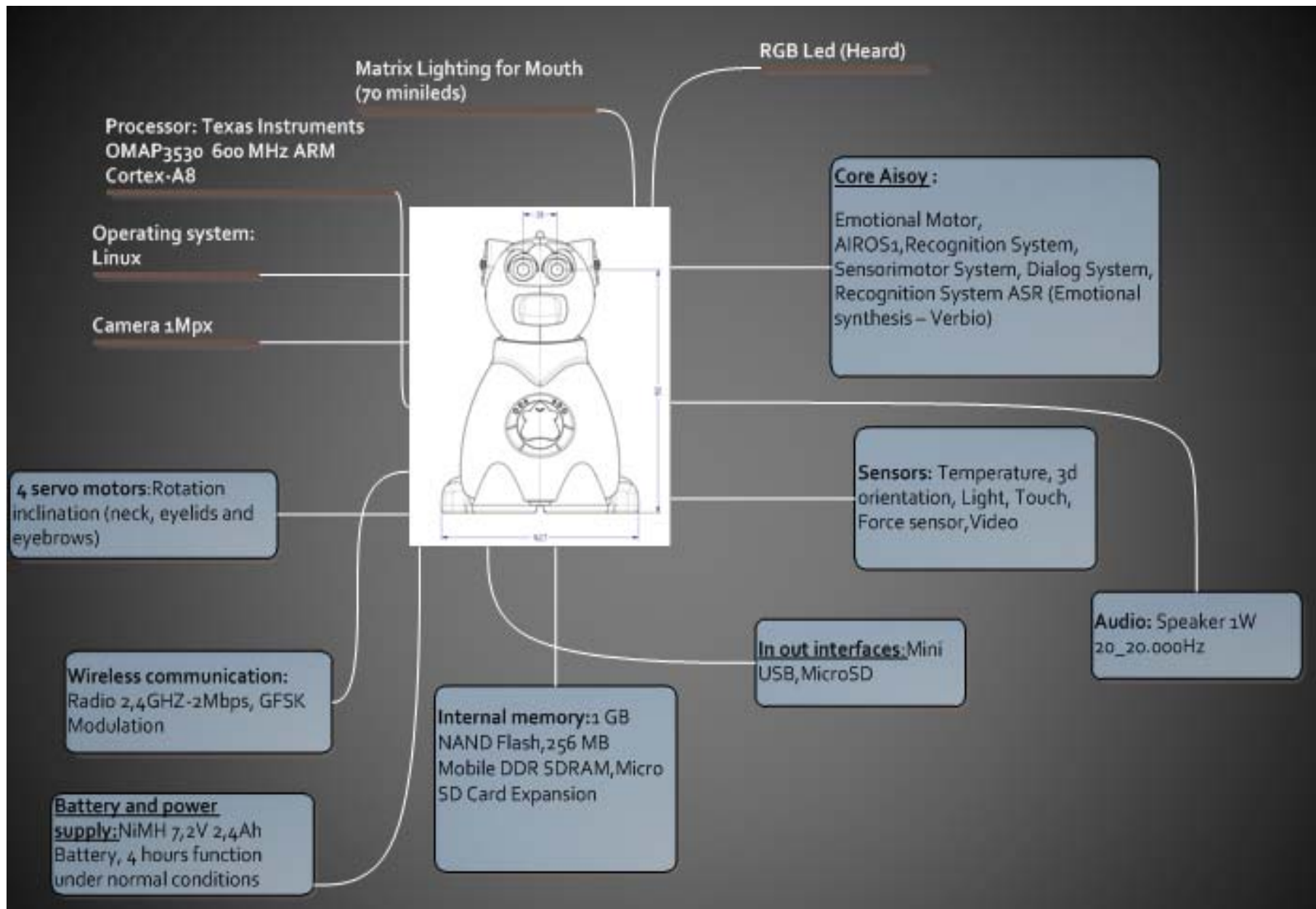
a	s1	d2	s2	Nodo	Utilidad esperada
000	000	040		nodo D2: max,ueD,D2	-1808,042
001	001	031		nodo D2: max,ueD,D2	-1241,771
001	000	040		nodo D2: max,ueD,D2	-1808,042
002	001	022		nodo D2: max,ueD,D2	-1331,182
002	000	040		nodo D2: max,ueD,D2	-1808,042
010	000	040		nodo D2: max,ueD,D2	-1808,042
010	010	040		nodo D2: max,ueD,D2	-1707,304
011	001	031		nodo D2: max,ueD,D2	-1241,771
011	011	031		nodo D2: max,ueD,D2	-1137,990
011	000	040		nodo D2: max,ueD,D2	-1808,042
011	010	040		nodo D2: max,ueD,D2	-1707,304
020	000	040		nodo D2: max,ueD,D2	-1808,042
020	010	040		nodo D2: max,ueD,D2	-1707,304
100	000	040		nodo D2: max,ueD,D2	-1808,042
100	100	130		nodo D2: max,ueD,D2	-1496,687
101	001	031		nodo D2: max,ueD,D2	-1241,771
101	000	040		nodo D2: max,ueD,D2	-1808,042
101	101	121		nodo D2: max,ueD,D2	-1027,931
101	100	130		nodo D2: max,ueD,D2	-1496,687
110	000	040		nodo D2: max,ueD,D2	-1808,042

# Outline

- From risk analysis to adversarial risk analysis
- Motivation
- Sequential games
- Simultaneous games
- Auctions
- Security
- **Intelligent interfaces**
- Challenges

# Problem

- An agent makes decisions in a finite set
- Has sensors providing information around it
- It relates with a user which makes decisions in an environment
- They're both within an environment which evolves (under the control of the user)





AlSoy Robotics  
 • 'AlSoy Y', con acento catalán

### Y, además, habla...

Se acaba de poner a la venta este robot capaz de hablar con nosotros, aprender cosas nuevas gracias a la interacción y reconocer hasta 14 sentimientos: alegría, ira, miedo, tristeza, sorpresa... «Somos ingenieros, informáticos, matemáticos... Hemos trabajado en consultoría, pero nos movía hacer algo que no habíamos visto en el mercado», dice Diego García, en la imagen, a la derecha, junto con David Siles.

tradicional. La novedad es que AlSoy Y se ha puesto a la venta este mismo agosto. «Le echo una suzeta-dilla», agrega Diego, «el primer robot de una nueva especie.»

El objetivo, claro, es crear dispositivos cada vez más complejos con aplicaciones directa en nuestra vida y, ahí sí, las aplicaciones tienen muy buenos ejemplos en los que estamos para aprender: las universidades. Rubi, por ejemplo, es un robot desarrollado en el Machine Perception Laboratory (algo así como el Laboratorio de Percepción de las Máquinas) de la Universidad de California, en San Diego. Como el AlSoy Y, también "aprende" de nosotros: memoriza nuevo vocabulario, reconoce nuestros rostros, nuestros estados de ánimo, interpreta el tono de voz, los gestos de la cara y hasta da clases a los más pequeños.

Sabe 'leer' si un niño presta o no atención, si le interesa la lección o piensa en las monedas. Y, según lo que ve, actúa de un modo u otro. Sus ojos son una pequeña cámara, y su cerebro, un refresco elaborado a partir de una base

de datos con más de 70.000 rostros en diferentes posiciones y estados de ánimo. Rubi tiene aún muchas limitaciones, pero ya se ha utilizado con pequeños autistas o menores de diez años, y sus creadores trabajan para mejorarlo.

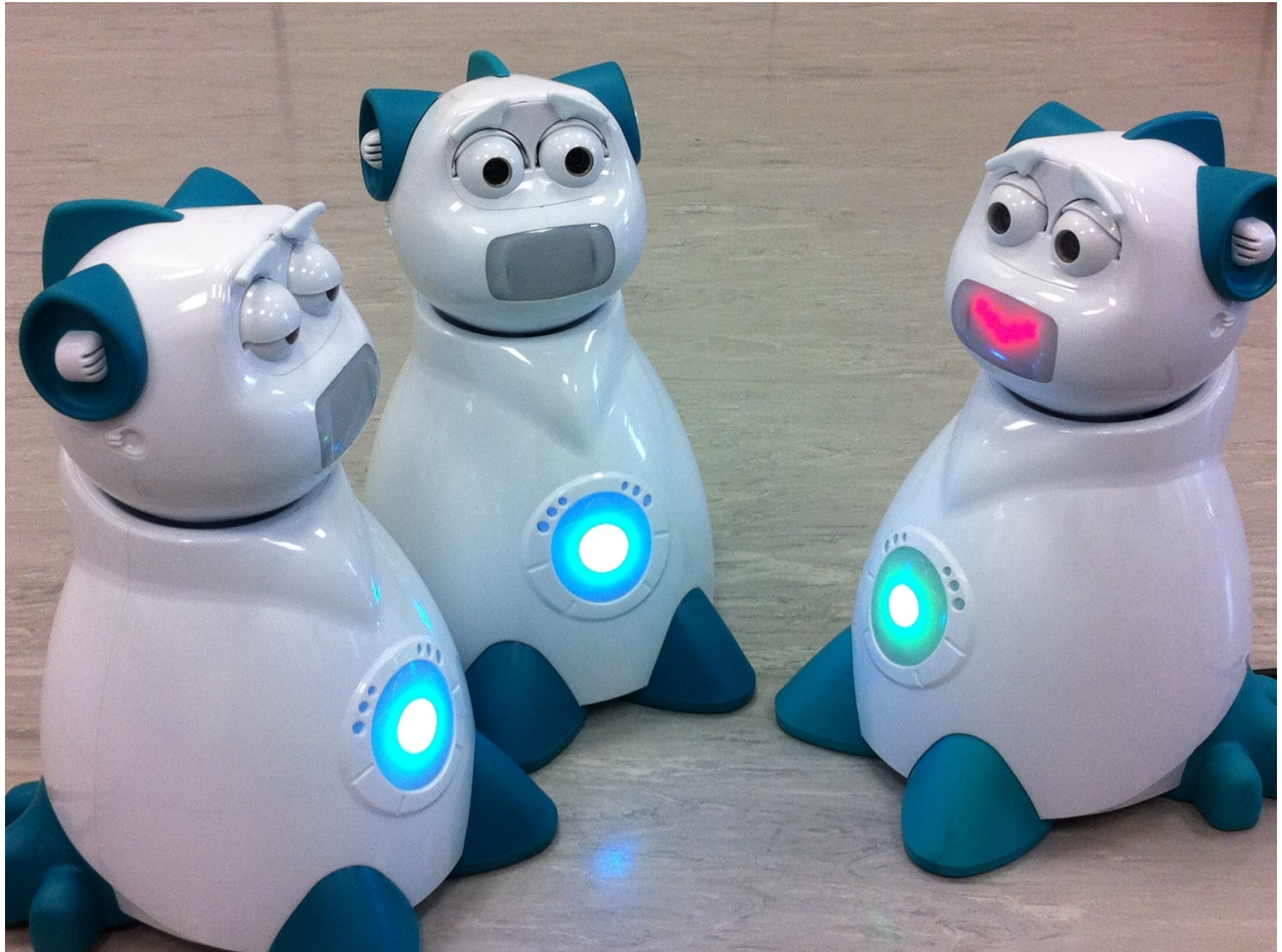
Todos los asuntos de la robótica contribuyen así a la ingeniería de un mundo no tan lejano y para el que la tecnología es cada vez menos un gran impedimento. «Para muchas de estas cosas, la tecnología está ya lista», dice el californiano Bob Allen, miembro del Humanoid Robotics

Club, investigador en la Universidad Carnegie Mellon (Pittsburgh, EE.UU.) y cofundador de la empresa OLogix, que cree que en 50 años podrán existir ya humanoides biónicos, que caminarán y nos ayudarán a cocinar, planchar, fregar, cortar el césped... «Contra a ellos, podemos dedicar nuestro tiempo libre a leer, estar con la familia o disfrutar de la vida», explica este experto, que trabaja en el desarrollo de pequeños 'seres' capaces de apagar incendios incendios. «La tecnología, insisto, ya está lista. Hay que trabajar y desarrollarla. En las próximas décadas podría cambiar mucho nuestra vida.» ■

#### PARA SABER MÁS

► [www.aysoyrobotics.com](http://www.aysoyrobotics.com) Página web con información sobre programación y configuración de robots como AlSoy Y, se puede adquirir kits de iniciación.

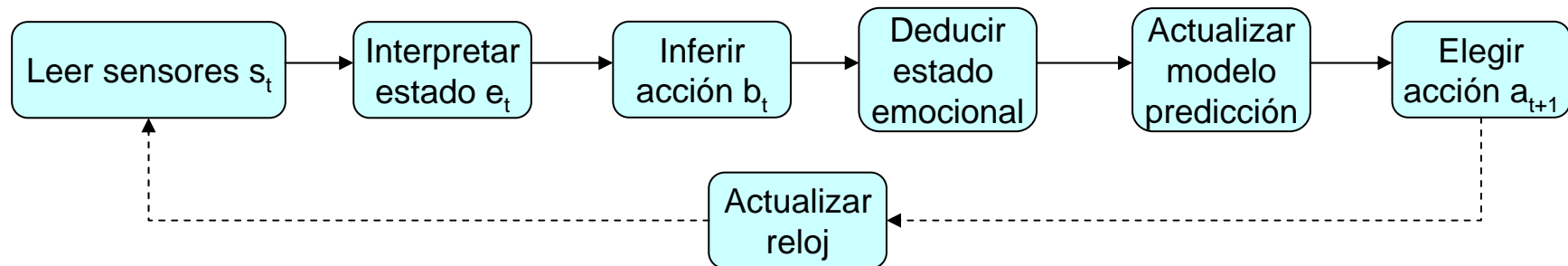
**'Rubi' sabe si a un niño le interesa la lección y actúa según lo que ve. Podría ayudar a niños autistas**





# Basic framework

$$\max_{a_t \in \mathcal{A}} \psi(a_t) = \sum_{b_t, e_t} u(a_t, b_t, e_t) \times p(b_t, e_t \mid a_t, (a_{t-1}, b_{t-1}, e_{t-1}), (a_{t-2}, b_{t-2}, e_{t-2}))$$



- # Basic framework

Several bots:

- Support each of the bots, treat the other bots as users (selfish 1). ARA
- Allow them to communicate, compute nash equilibria (selfish 2)
- If they communicate, from selfish to cooperative. ARA
- Emotions impacting degree of cooperativeness

# Outline

- From risk analysis to adversarial risk analysis
- Motivation
- Sequential games
- Simultaneous games
- Auctions
- Security
- Intelligent interfaces
- **Challenges**

# 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
  - What if the other not EU maximiser? Prospect theory, concept uncertainty
- **Several simple but illustrative models**
  - sequential D-A, simultaneous D-A, D-A-D, sequential DA with private information decision problems
  - What if
    - more complex dynamic interactions? (coupled IDs with shared nodes)
    - against more than one Attacker?
    - an uncertain number of Attackers?
    - several defenders? (risk sharing negotiations)
- **Implementation issues**
  - Elicitation of a valuable judgmental input from Defender
  - Computational issues (optimization + simulation)
    - Augmented simulation
  - Parallel
  - Portfolio theory
  - Both problem sin one shot
  - Templates
  - K.level. The value of information
  - Computational environment
- **Other applications**
  - Auctions
  - Cybersecurity

# Discussion

- Multiple Defenders to be coordinated (risk sharing).
  - Private security
  - Multiple Attackers possibly coordinated
  - Various types of resources
  - Various types of delinquency
  - Multivalued cells. The perception of security (concern analysis)
  - Multiperiod planning
  - Time and space effects (Displacement of delicts)
  - Insurance
  - General coupled influence diagrams
- 
- Networks with value only at nodes
  - Networks with value at nodes and arcs

# Discussion

- Educational environments
- Emotions and cooperativeness
- Multiperiod planning
- Mobility

# Thanks!!!

[david.rios@urjc.es](mailto:david.rios@urjc.es)

[www.analisisderiesgos.com](http://www.analisisderiesgos.com)

[www.aisoy.com](http://www.aisoy.com)