



Comp Sci 288: AI for Social Impact

Lecture 2: Case Studies from past 15 years of work

MILIND TAMBE

Announcements

- Over to TFs
- Questions?



HARVARD UNIVERSITY



AI For Social Impact Seminar Series

Every Monday, Jan 25 - Apr 19, 11 AM Eastern (UTC-5) • On Zoom

This seminar series is bringing together a diverse set of perspectives that spans across algorithmic fairness, human-centered computing, and sustained deployment.

SIGN UP:

url.seas.harvard.edu/aisi

Jan 25: Danielle Belgrave, Ph.D.
Microsoft Research Cambridge (UK)

Feb 1: Omer Reingold, Ph.D.
Stanford University

Feb 8: Heather Lynch, Ph.D.
Stony Brook University

Feb 22: Munmun De Choudhury, Ph.D.
Georgia Institute of Technology

Mar 8: Courtney Cogburn, Ph.D.
Columbia University

Mar 15: Lauren Wilcox, Ph.D.
Georgia Institute of Technology,
Google

Mar 22: Tiffany Veinot, MLS, Ph.D.
University of Michigan

Mar 29: Kush Varshney, Ph.D.
Thomas J. Watson Research Center,
IBM

Apr 5: Christopher Le Dantec, Ph.D.
Georgia Institute of Technology

Apr 12: Nyalleng Moorosi, Ph.D.
Google AI

Apr 19: Michael J. Mina, MD, Ph.D.
Harvard T.H. Chan School of Public
Health, Harvard Medical School

Goals of next two lectures

- Summarizing 15 years of research on AI for social impact
- How to “AI for Social Impact” via case studies:
 - Types of problems solved, AI tools available to solve them
 - Challenges in applying AI tools
 - Challenges in measuring impact
 - Ethical challenges
 - Pitfalls
- Chronological ordering
- Partnership lessons
- General lessons learned

Goals of next two lectures

- Summarizing 15 years of research on AI for social impact
- How to “AI for Social Impact” via case studies:
 - Types of problems solved, AI tools available to solve them
 - Challenges in applying AI tools
 - Challenges in measuring impact
 - Ethical challenges
 - Pitfalls
- You may have different technical backgrounds
- **Understand the problem and model used**, not essential to understand details of the algorithms used – you can pick that up for your project if necessary

AI & Multiagent Systems Research for Social Impact



Public Health



Conservation

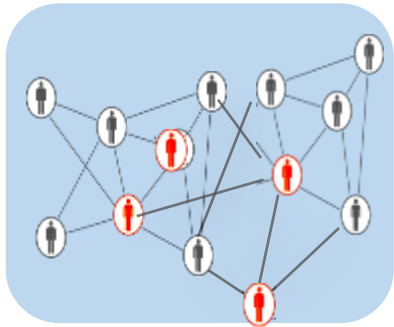


**Public Safety
and Security**

Key Research Challenge

Optimize Our Limited Intervention Resources

Optimizing Limited Intervention Resources



**Social
Networks &
Bandits**

Public Health



**Green
security
games**



Conservation



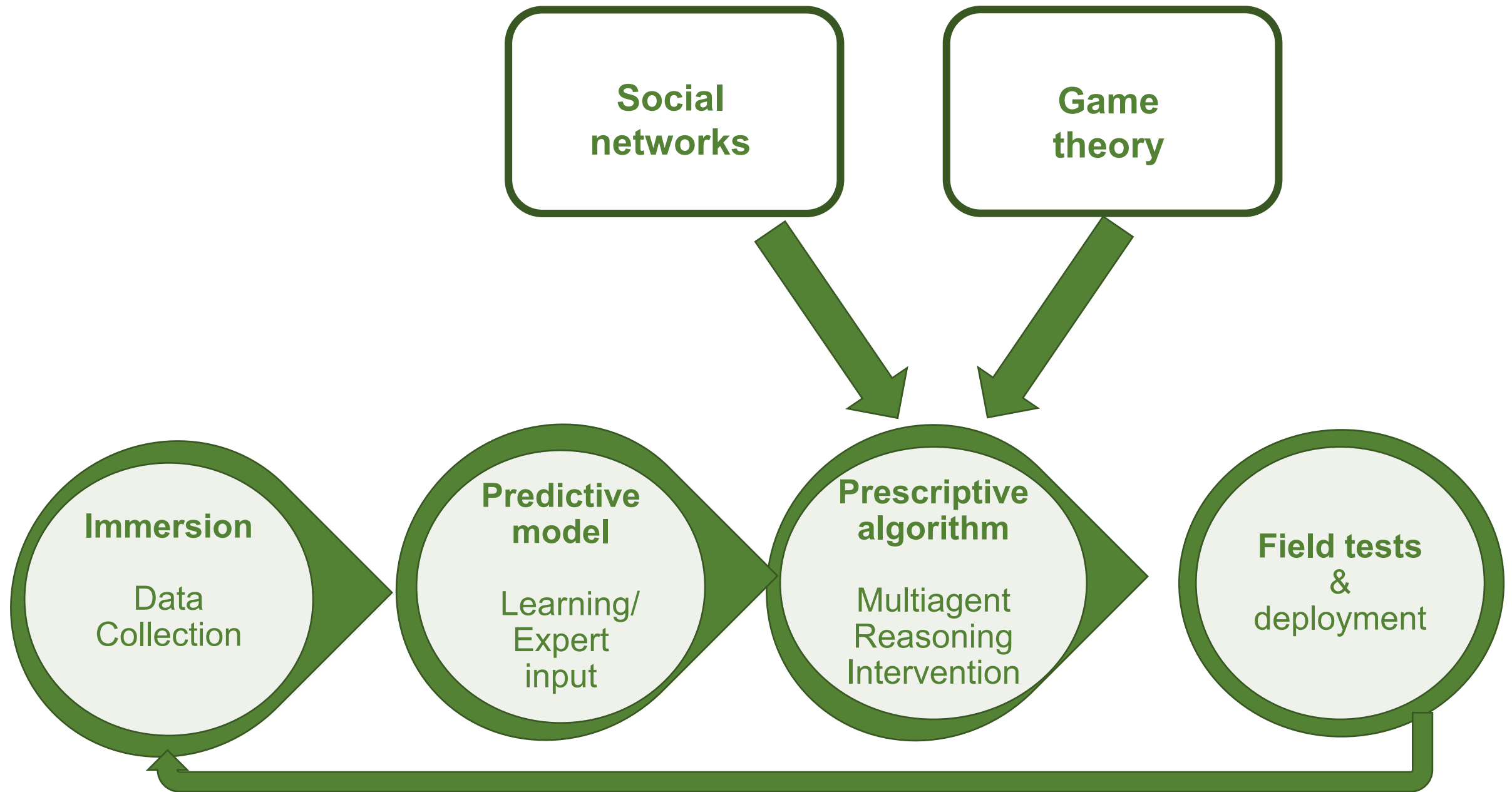
**Public Safety
& Security**



**Stackelberg
security
games**

Three Common Themes

Multiagent systems, Data-to-deployment pipeline, Interdisciplinary partnerships

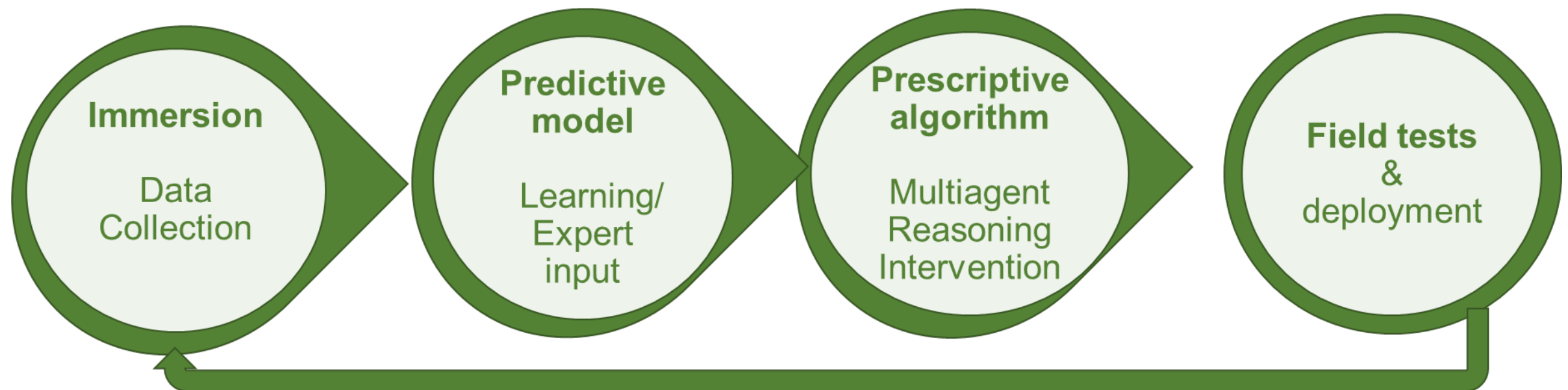


Three Common Themes

Multiagent systems, Data-to-deployment pipeline, Interdisciplinary partnerships

Field test & deployment: Social impact is a key objective

Lack of data is a norm: Must be part of project strategy



Three Common Themes

Multiagent systems, Data-to-deployment pipeline, Interdisciplinary partnerships



Empower non-profits to use AI tools; avoid being gatekeepers to AI4SI technology



ASHOKA TRUST FOR RESEARCH IN
ECOLOGY & THE ENVIRONMENT



Outline



Public Safety and Security:
Stackelberg Security Games

Conservation/Wildlife Protection:
Green Security Games

11 July 2006: Mumbai

TRAIN OF TERROR

Mumbai continues to be the prime target for terrorist groups. It has borne the brunt of seven attacks in the past 13 years.



Explosive used
High-quality explosive.
Most likely RDX
(Cyclotrimethylenetrinitramine)



Quantity of explosive
At least 5 kg per blast;
possibly packed into
bags or tiffin boxes



Why attack the first class compartments?
It is easier to enter a first class compartment at peak hour than a second class with a bag filled with up to 5 kg of explosives



Where were bombs placed?
In the luggage racks where commuters keep their bags and tiffin boxes



How many bombers were there?
At least 20, 2 for each blast and a logistics base of 6 people

WARNING

JAN 6, 2006:
seized from three youths in Mumbai

JAN 30, 2006:
powder and 2 rifles from 2 people at

MAY 9, 2006:
2,000 live cartridges, AK-56s seized in

MAY 12, 2006:
live cartridges and grenades seized

MAY 14, 2006:
three AK-47s and cartridges seized

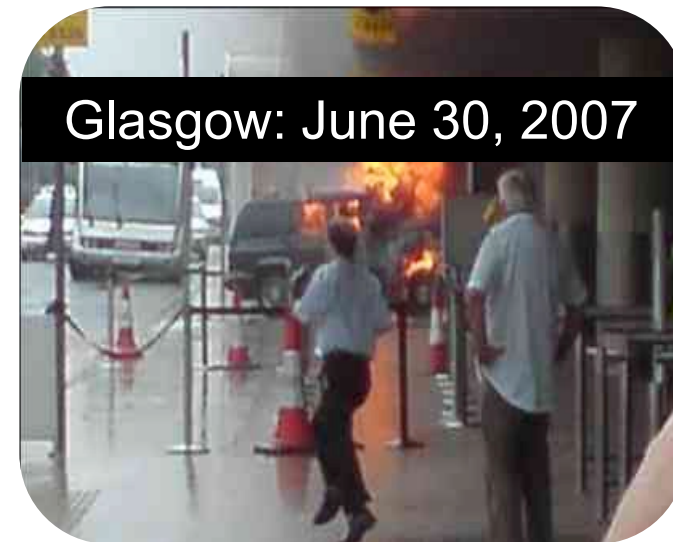


ARMOR Airport Security: LAX(2007)

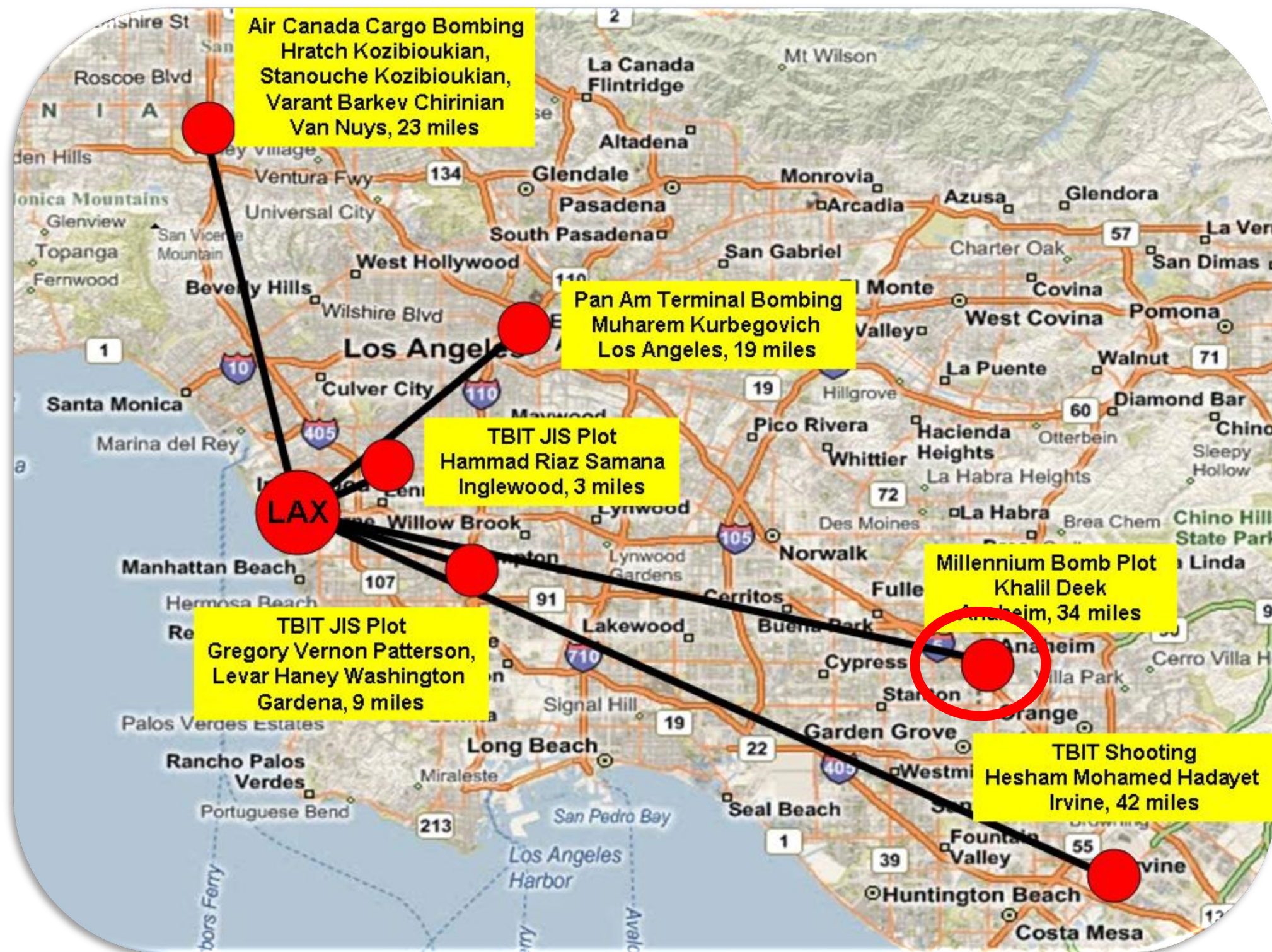
Erroll Southers



LAX Airport, Los Angeles



Background on LAX Airport Threats: Surveillance Opportunity



LAX Airport Case: Optimize Limited Security Resources

Eight Inbound Roads, Eight Terminals: Limited Staff, Canines

Can we propose game theory for security resource optimization?



Game Theory for Security Resource Optimization

New Model: Stackelberg Security Games, key aspects for tractability

Set of targets, payoffs based on targets covered or not
Stackelberg Leader-Follower formulation



Defender

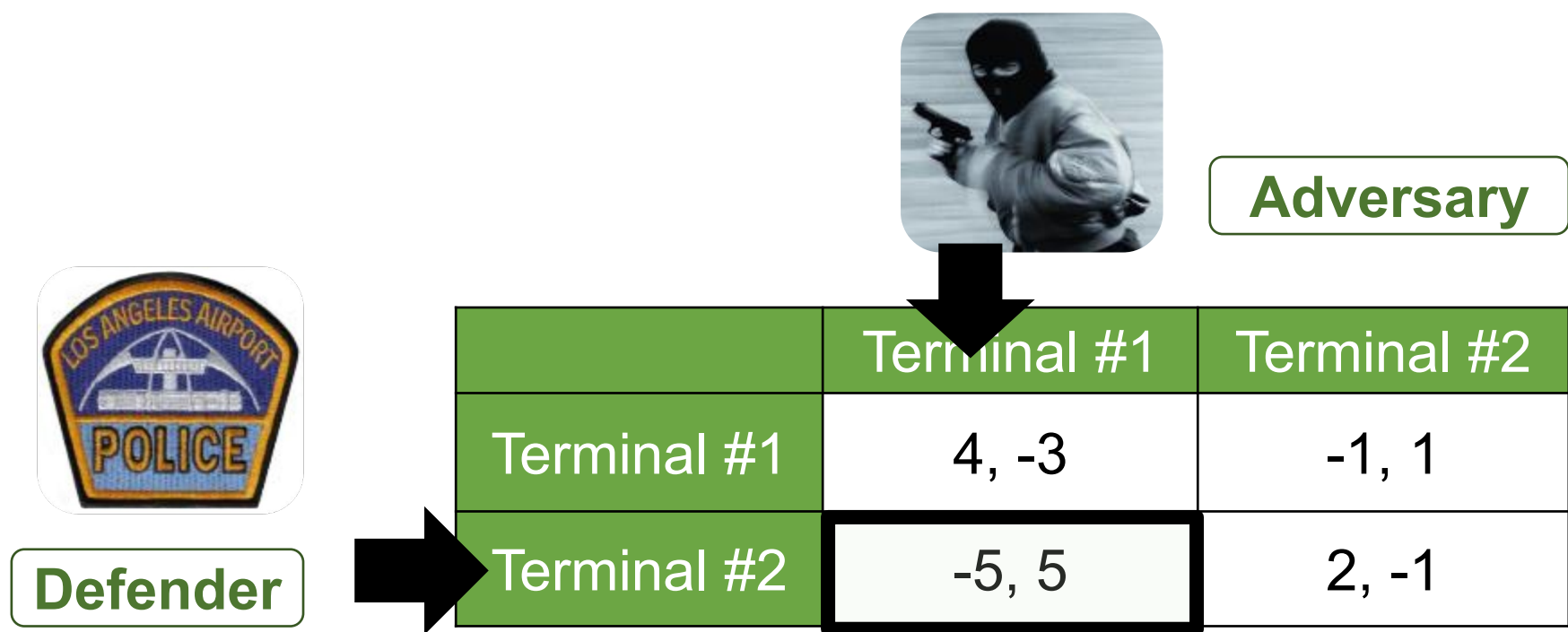


Adversary

	Terminal #1	Terminal #2
Terminal #1	4, -3	-1, 1
Terminal #2	-5, 5	2, -1

Game Theory for Security Resource Optimization

New Model: Stackelberg Security Games



Model: Stackelberg Security Games

Stackelberg: Defender commits to randomized strategy, adversary responds

Security optimization: Not 100% security; increase cost/uncertainty to attackers

Challenges faced: Massive scale games



Defender



Adversary

	Terminal #1	Terminal #2
Terminal #1	4, -3	-1, 1
Terminal #2	-5, 5	2, -1

ARMOR at LAX

Basic Security Game Operation [2007]



Kiekintveld



Pita



	Target #1	Target #2	Target #3
Defender #1	2, -1	-3, 4	-3, 4
Defender #2	-3, 3	3, -2	
Defender #3			



Mixed Integer Program



$\Pr(\text{Canine patrol, 8 AM @Terminals 2,5,6}) = 0.17$

Canine Team Schedule, July 28

	Term 1	Term 2	Term 3	Term 4	Term 5	Term 6	Term 7	Term 8
8 AM		Team1			Team3	Team5		
9 AM			Team1	Team2				Team4
...	...	...	...	...	...	...	...	...

OK IF YOU DO NOT FOLLOW THIS SLIDE

Mixed Integer Program [2007]



Kiekintveld



Pita



j →

i ↓	Target #1	Target #2	Target #3
Defender #1	2, -1	-3, 4	-3, 4
Defender #2	-3, 3	3, -2	
Defender #3			

$$\max \sum_{i \in X} \sum_{j \in Q} R_{ij} \times x_i \times q_j$$

Maximize defender expected utility

$$s.t. \sum_i x_i = 1$$

Defender mixed strategy

$$\sum_{j \in Q} q_j = 1$$

Adversary response

$$0 \leq (a - \sum_{i \in X} C_{ij} x_i) \leq (1 - q_j) M$$

Adversary best response

We are trying to
Find x_i

SECURITY GAME PAYOFFS [2007]

Previous Research Provides Payoffs in Security Games



	Target #1	Target #2	Target #3
Defender #1	2, -1	-3, 4	-3, 4
Defender #2	-3, 3	3, -2	
Defender #3			

+ Handling
Uncertainty

$$\max \sum_{i \in X} \sum_{j \in Q} R_{ij} \times x_i \times q_j$$

Maximize defender
expected utility



ARMOR: Optimizing Security Resource Allocation [2007]

First application: Computational game theory for operational security



January 2009

- January 3rd *Loaded 9/mm pistol*
- January 9th *16-handguns,
1000 rounds of ammo*
- January 10th *Two unloaded shotguns*
- January 12th *Loaded 22/cal rifle*
- January 17th *Loaded 9/mm pistol*
- January 22nd *Unloaded 9/mm pistol*

ARMOR AIRPORT SECURITY: LAX [2008]

Congressional Subcommittee Hearings



**Commendations
City of Los Angeles**



**Erroll Southern testimony
Congressional subcommittee**



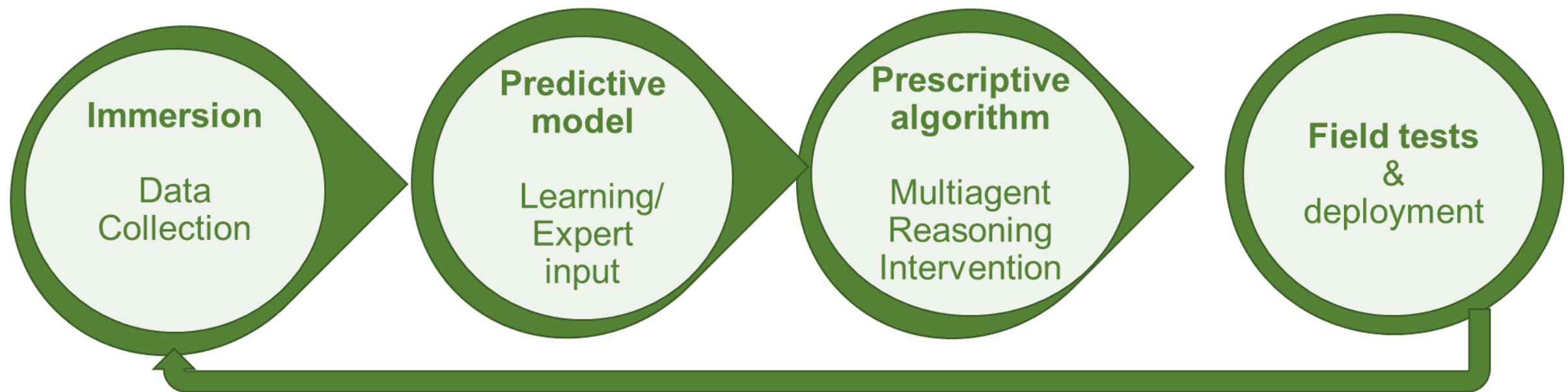
ARMOR...throws a digital cloak of invisibility....



New applications: cybersecurity, protecting of endangered wildlife and fisheries, protecting forests, audit games, drug design against viruses, traffic enforcement, software code testing, adversarial machine learning

Questions?

Behind the scenes



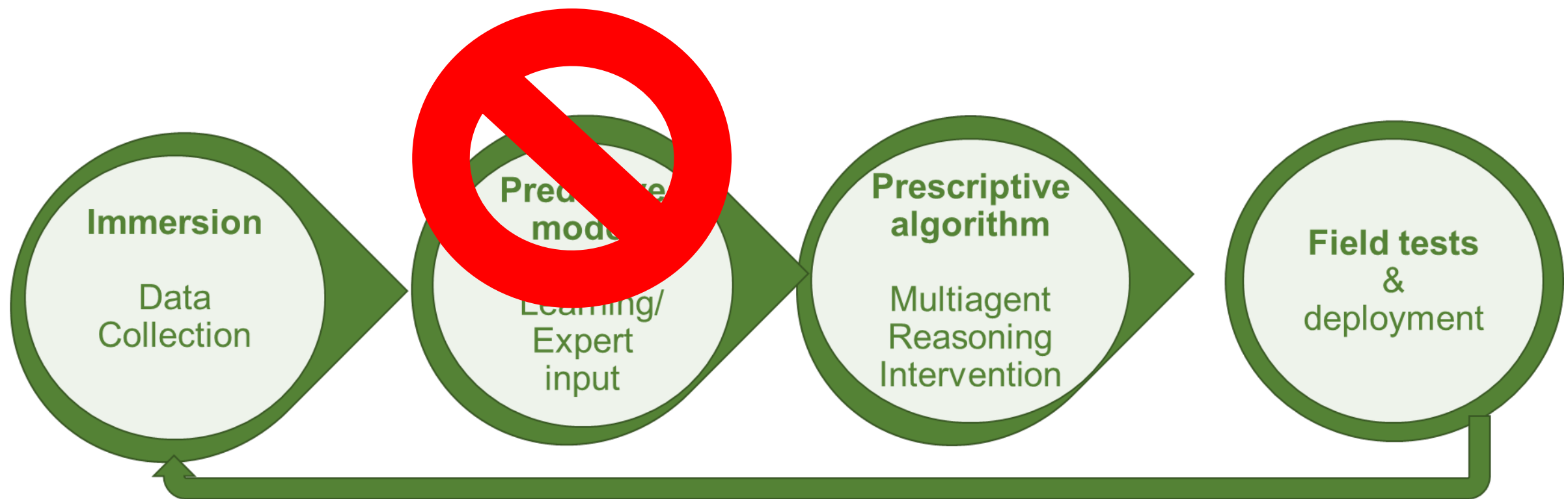
Lessons in Immersion



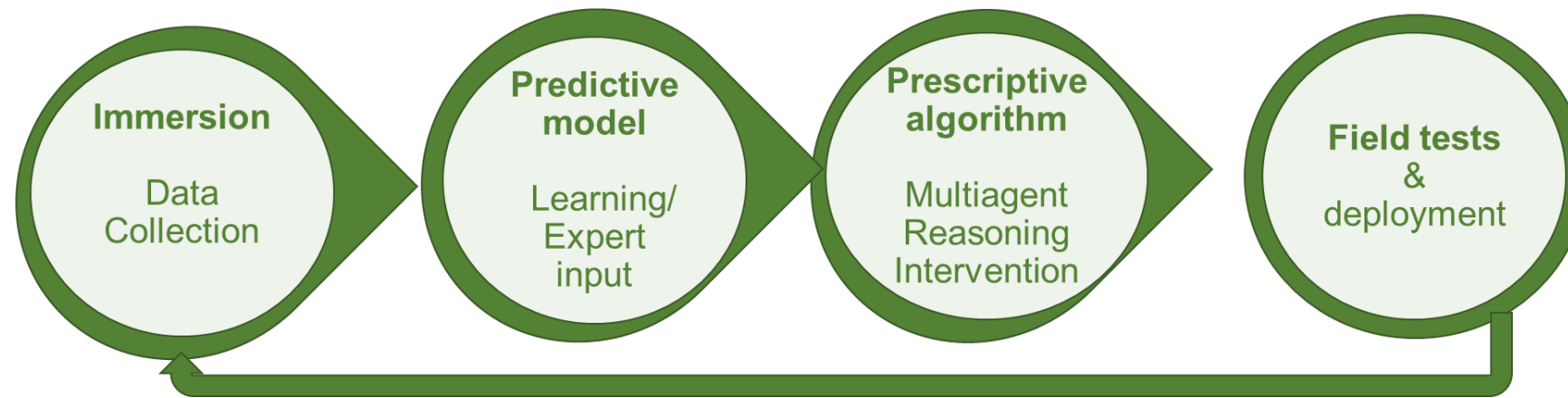
- Understanding their counter-terrorism experience



No data-driven predictive model



AI ADVANCES?

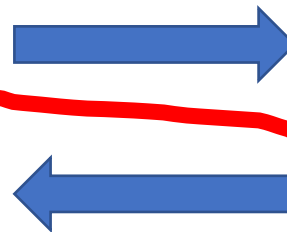


Advancing AI + MAS

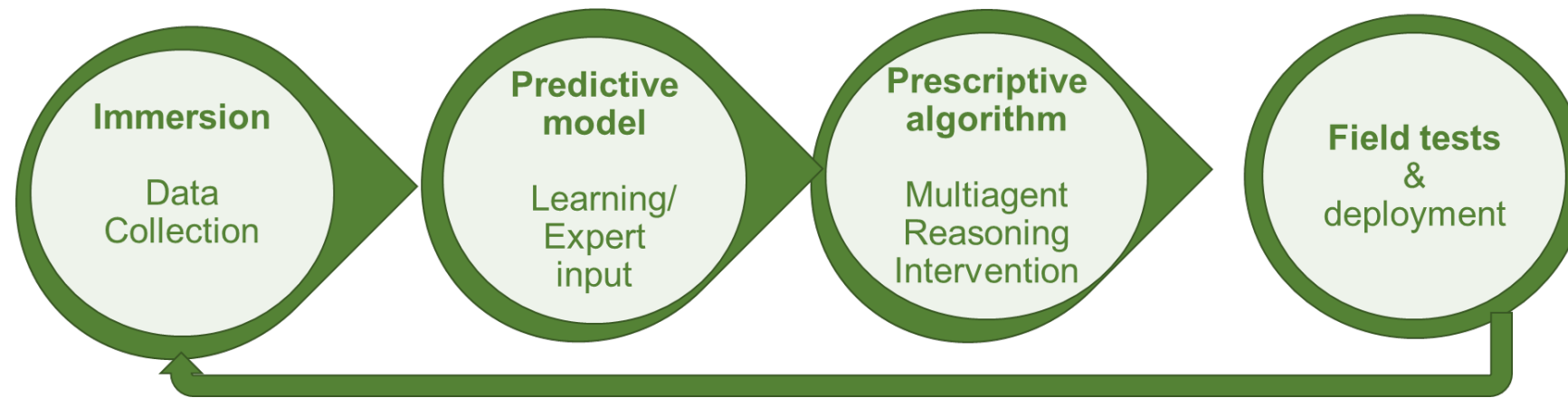
- Game theory
- Social networks
- Bandits
- POMDPs
- RL
- Decision-focused Learning

Achieving impact

- Public health
- Conservation
- Public safety



AI ADVANCES?



- E.G.: Stackelberg security games from ARMOR
- There wasn't as much interest in Stackelberg
- NOT JUST Stackelberg, BUT SECURITY GAME

Simplify Interaction

ARMOR - Checkpoint

FileRestrictionsReports

LAX
Los Angeles World Airports

CREATE

HOMELAND SECURITY CENTER

TEAMCORE

USC

☐ Must Be Scheduled

☐ Must Not Be Scheduled

☒ At Least One Scheduled

☐ Unrestrict

Apply

Manually adjust the generated schedule

Add

or

Remove

<November, 2007>

Sun	Mon	Tue	Wed	Thu	Fri	Sat
28	29	30	31	1	2	3
4	5	6	7	8	9	10
11	12	1				
	9	20	21	22	23	24
25	26	27	28	29	30	1
2	3	4	5	6	7	8

Today: 11/13/2007

Set First Day of Schedule

Days to Schedule7

Patrols to Schedule0

Simultaneous Patrols1

Randomness: Uncalculated

	Tuesday						Wednesday						Thursday						Friday						Saturday						Sunday						Monday						At-Leasts
Checkpoint #:	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	none								
8:00-10:00 AM	█																																	1									
10:00-12:00 AM																																		2									
12:00-2:00 PM			█																																								
2:00-4:00 PM																																											
4:00-6:00 PM																																											
6:00-8:00 PM																																											
8:00-10:00 PM																																											

Remove At-Least

ARMOR Transition



ARMOR at LAX: *Six Month Evaluation (Aug 2007 to Feb 2008)*



Cost-benefit analysis

Savings

- \$30 Million in ARMOR
- \$35 Million in PROTECT
- > benefit of IRIS

Risk Analysis, Vol. 40, No. 3, 2020

DOI: 10.1111/risa.13403

Assessing the Benefits and Costs of Homeland Security Research: A Risk-Informed Methodology with Applications for the U.S. Coast Guard

Detlof von Winterfeldt,^{1,*} R. Scott Farrow,² Richard S. John,¹ Jonathan Eyer,¹ Adam Z. Rose,¹ and Heather Rosoff¹

J. Benefit Cost Anal. 2020; 1–22 © The Author(s), 2020. Published by Cambridge University Press on behalf of the Society for Benefit-Cost Analysis
doi:[10.1017/bca.2020.24](https://doi.org/10.1017/bca.2020.24)

Scott Farrow* and Detlof von Winterfeldt

Retrospective Benefit–Cost Analysis of Security-Enhancing and Cost-Saving Technologies

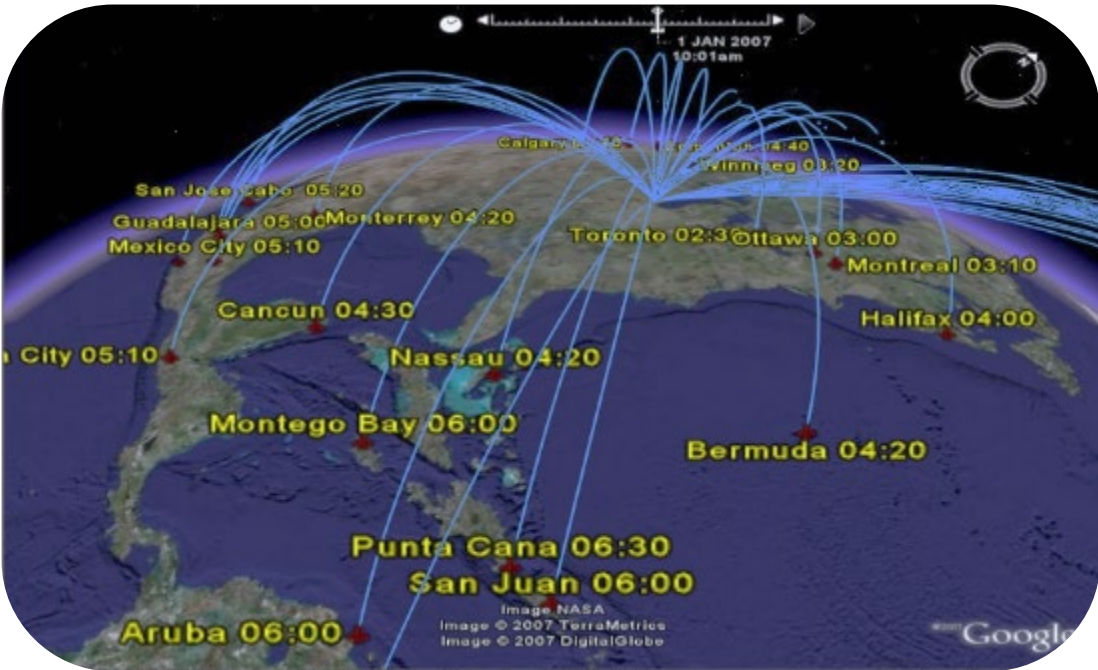
Some lessons

- Impact evaluation is complicated: Domain experts may not provide scientific evaluation
- Must respect others with other areas of expertise
- No zero-risk application
- Pressure to publish AI advances: if its not a methodological advance AI conferences don't care
 - *Problematic for AI for social impact*
- Did not set an end date! There must be an end date

Questions

Federal Air Marshals Service [2009]

Visiting Freedom Center: Home of Federal Air Marshals Service



	Strategy 1	Strategy 2	Strategy 3	Strategy 4
Strategy 1	IRIS 1000 flights/day Actions: ~10 ⁴¹			
Strategy 2				
Strategy 3				
Strategy 4				

Scale Up Difficulty [2009]



Kiekintveld



Jain

x_i Defender mixed strategy

1000 flights, 20 air marshals:

10^{41} combinations

$$\max_{x,q} \sum_{i \in X} \sum_{j \in Q} R_{ij} x_i q_j$$

$$s.t. \sum_i x_i = 1, \sum_{j \in Q} q_j = 1$$

$$0 \leq (a - \sum_{i \in X} C_{ij} x_i) \leq (1 - q_j) M$$

	Attack 1	Attack 2	Attack ...	Attack 1000
1, 2, 3 ..	5,-10	4,-8	...	-20,9
1, 2, 4 ..	5,-10	4,-8	...	-20,9
1, 3, 5 ..	5,-10	-9,5	...	-20,9
...				
...	← 10^{41} rows			

Scale Up [2009]

Exploiting Small Support Size



Kiekintveld



Jain

Small support set size:
Most x_i variables zero

1000 flights, 20 air marshals:
 10^{41} combinations

		Attack 1	Attack 2	Attack ...	Attack 1000
$X_{123} = 0.0$	1, 2, 3 ..	5, -10	4, -8	...	-20, 9
$X_{124} = 0.239$	1, 2, 4 ..	5, -10	4, -8	...	-20, 9
$X_{135} = 0.0$	1, 3, 5 ..	5, 10	0, 5	...	20, 0
$X_{378} = 0.123$	...				
	... ← 10^{41} rows				

New Exact Algorithm for Scale up



Kiekintveld



Jain

Incremental strategy generation: First for Stackelberg Security Games

Primary

	Attack 1	Attack 2	...	Attack 6
1,2,4	5,-10	4,-8	...	-20,9

	Attack 1	Attack 2	...	Attack 6
1,2,4	5,-10	4,-8	...	-20,9

3,7,8	-8,10			
-------	-------	--	--	--

	Attack 1	Attack 2	...	Attack 6
1,2,4	5,-10	4,-8	...	-20,9
3,7,8	-8,10	-8,10	...	-8,10
...	...	...	...	...

Secondary (LP Duality Theory)
Best new pure strategy

GLOBAL OPTIMAL
1000 defender strategies
NOT 10^{41}

ality Theory)
re strategy

IRIS: Deployed FAMS [2009-]

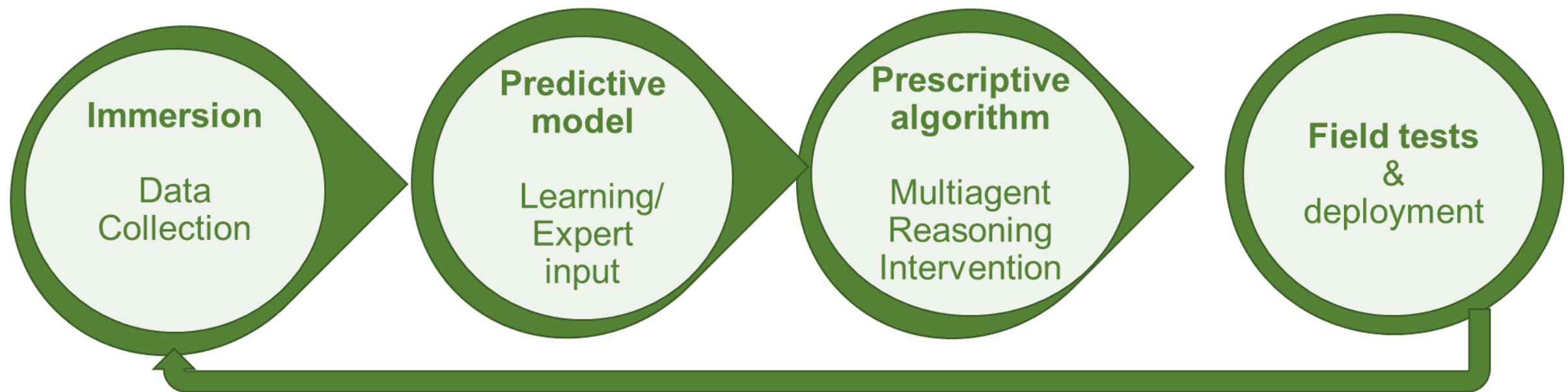


Significant change in FAMS operations



September 2011: Certificate of Appreciation (Federal Air Marshals)

Behind the scenes



26 Nov 2008, Mumbai

Police Checkpoints: Network Security Game



Jain

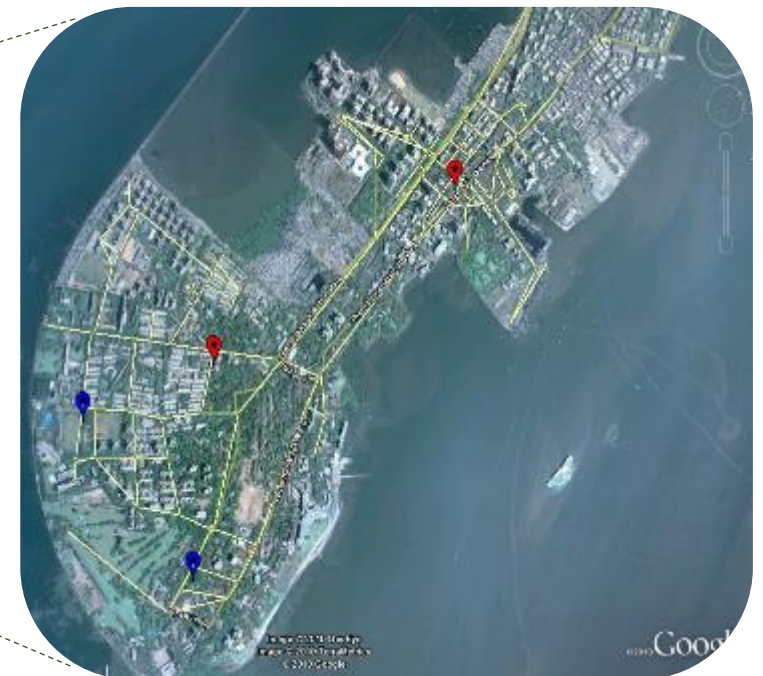


Road networks:

20,000 roads, 15 checkpoints



150 edges
2 Checkpoints
150-choose-2 strategies



Zero-Sum Network Security Game [2013]



Jain

Double oracle: New exact optimal algorithm for scale-up

	Path #1	Path #2	Path #3
Checkpoint strategy #1	5, -5	-1, 1	-2, 2
Checkpoint strategy #2	-5, 5	1, -1	-2, 2



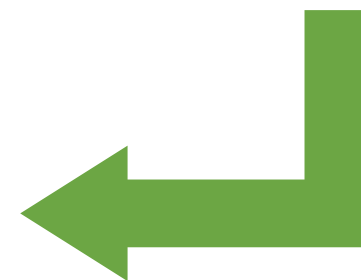
Defender oracle

	Path #1	Path #2
Checkpoint strategy #1	5, -5	-1, 1
Checkpoint strategy #2	-5, 5	2, -1



Attacker oracle

	Path #1	Path #2	Path #3
Checkpoint strategy #1	5, -5	-1, 1	-2, 2
Checkpoint strategy #2	-5, 5	1, -1	-2, 2



Presentation at the Indian National Police Academy: Network Security Game [2016]

Road networks:

20,000 roads, 15 checkpoint:
Solved under 20 min



Some lessons

- No “immersion” meant no ability to build up trust

PROTECT: Port and Ferry Protection Patrols [2011]



Shieh

An

Boston



Los Angeles



New York



PROTECT: Port and Ferry Protection Patrols [2011]



Shieh



An



PROTECT: Ferry Protection Deployed [2013]



Fang



Jiang



PROTECT: Ferry Protection Deployed [2013]



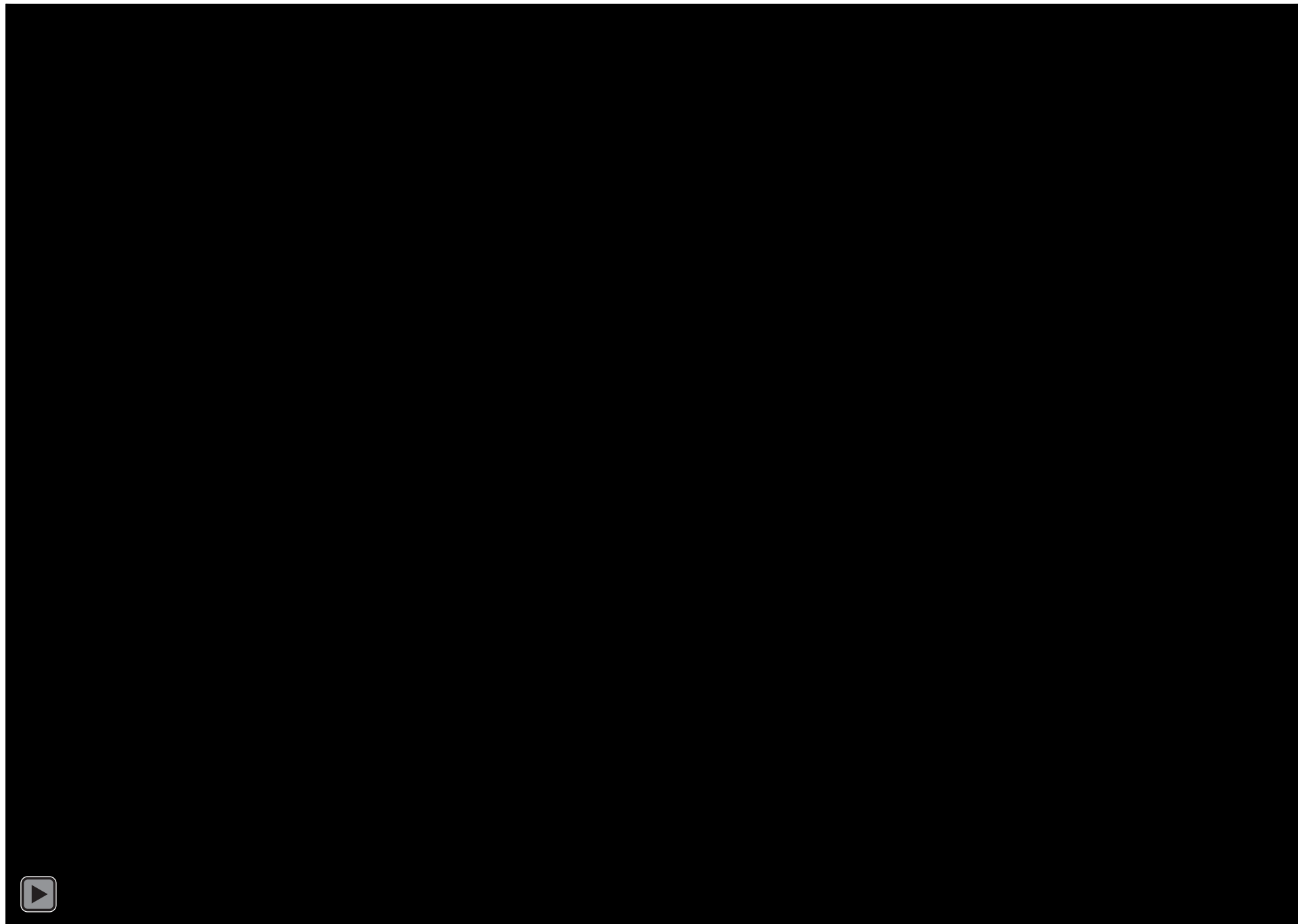
Fang



Jiang



PROTECT: Ferry Protection Deployed [2013]



FERRIES: Mobile Resources & Moving Targets

Transition Graph Representation

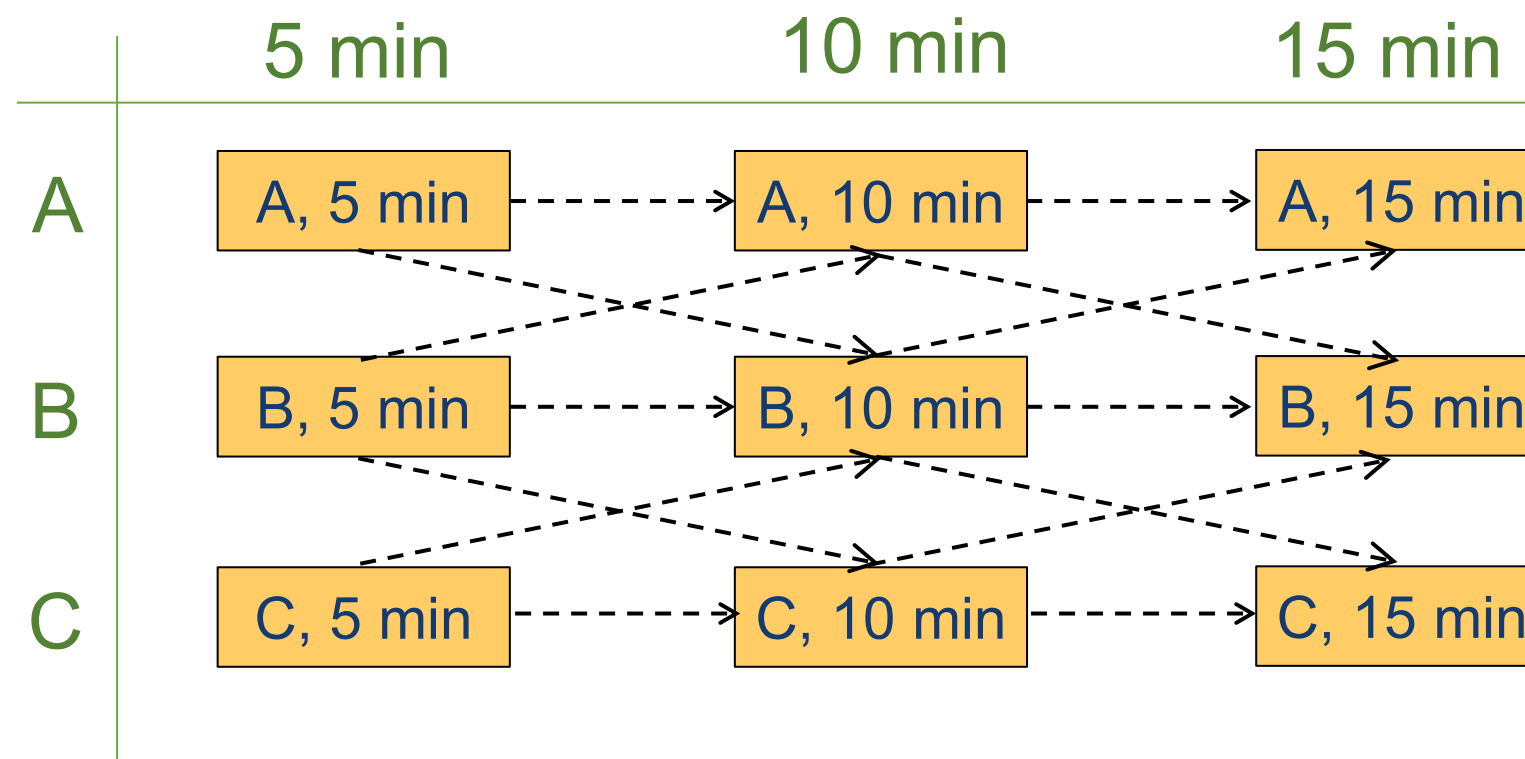


Fang



Jiang

Marginal strategy: New scale-up approach for Stackelberg Security Games



PROTECT: Port Protection Patrols [2013]

Congressional Subcommittee Hearing



**June 2013: Meritorious Team Commendation
from Commandant (US Coast Guard)**



**July 2011: Operational Excellence
Award (US Coast Guard, Boston)**



**US Coast Guard testimony
Congressional subcommittee**

Train Patrols

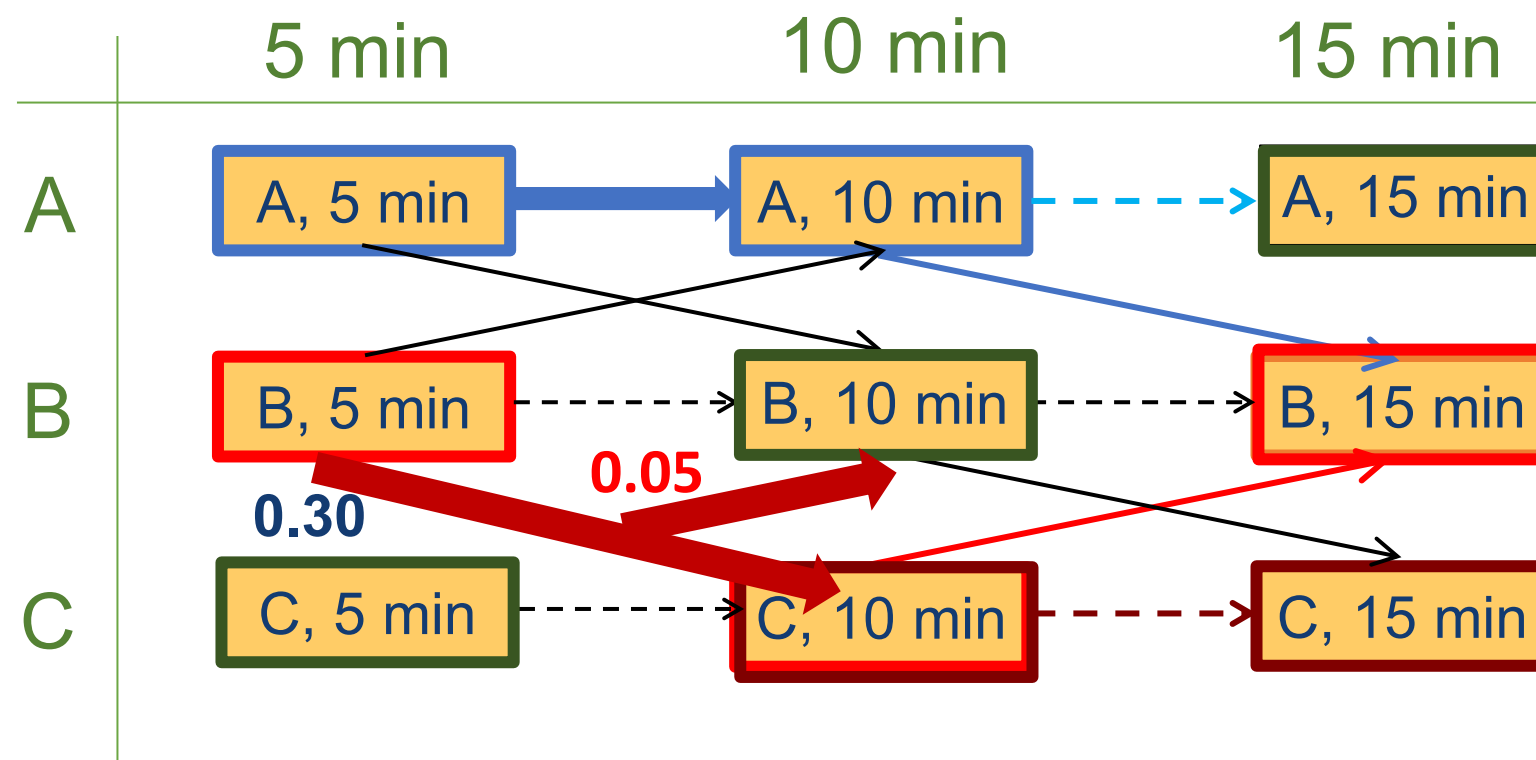
Execution Uncertainty: MDPs



Jiang



Delle Fave



Questions

Evaluation

- Breakout
- “BUT DOES THIS WORK”?

Evaluating Deployed Security Systems Not Easy

How Well Optimized Use of Limited Security Resources?

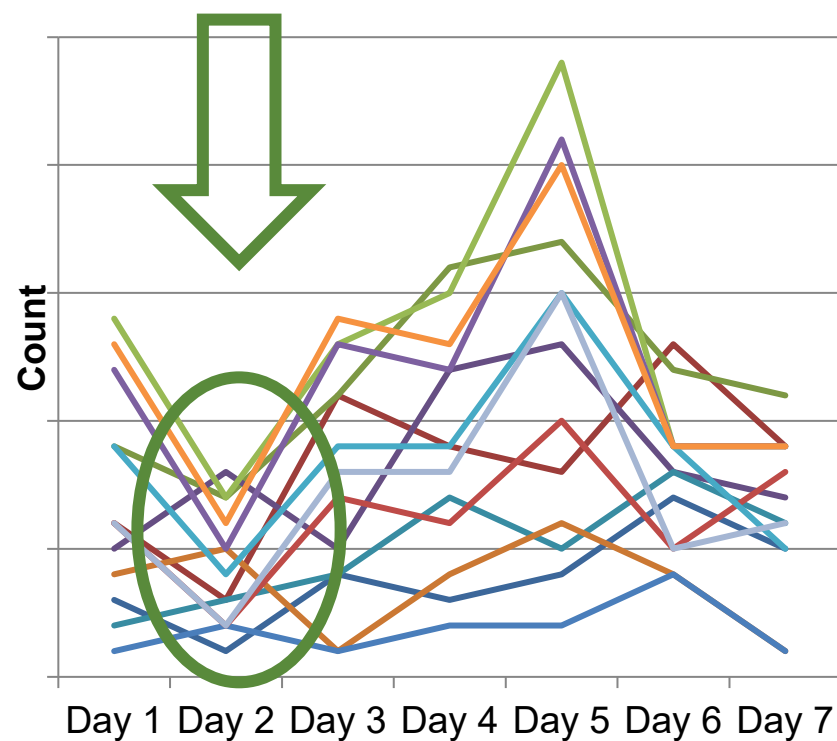
Security Games superior
vs
Human Schedulers/"simple random"

- ❖ Lab evaluation
 - ❖ *Scheduling competitions: Patrol quality unpredictability? Coverage?*
 - ❖ Field evaluation: Tests against real adversaries
 - ❖ *Economic cost-benefit analysis*
 - ❖ ...
-

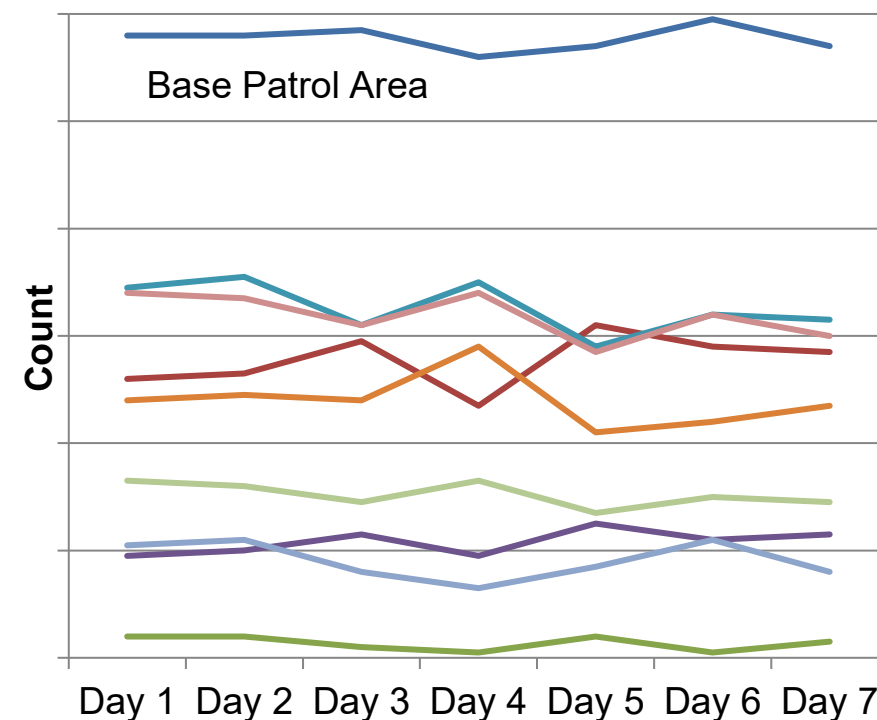
Field Evaluation of Schedule Quality

Improved Patrol Unpredictability & Coverage for Less Effort

Patrols Before PROTECT: Boston



Patrols After PROTECT: Boston



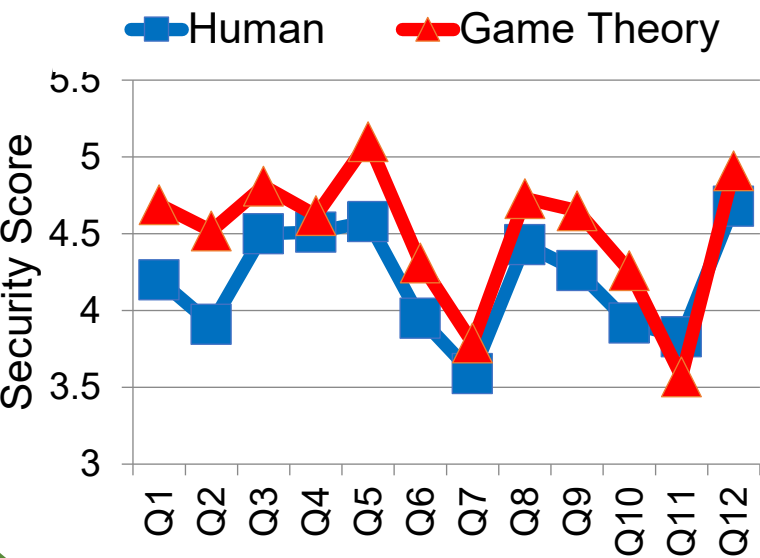
350% increase in defender expected utility

Field Evaluation of Schedule Quality

Improved Patrol Unpredictability & Coverage for Less Effort

FAMS: IRIS Outperformed expert human over six months

Report:GAO-09-903T



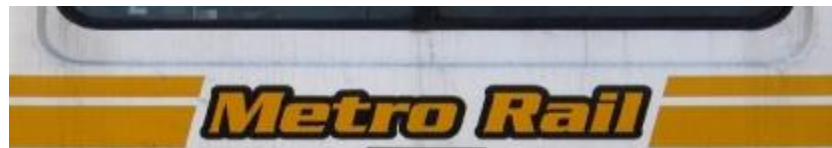
Trains: TRUSTS outperformed expert humans schedule 90 officers on LA trains



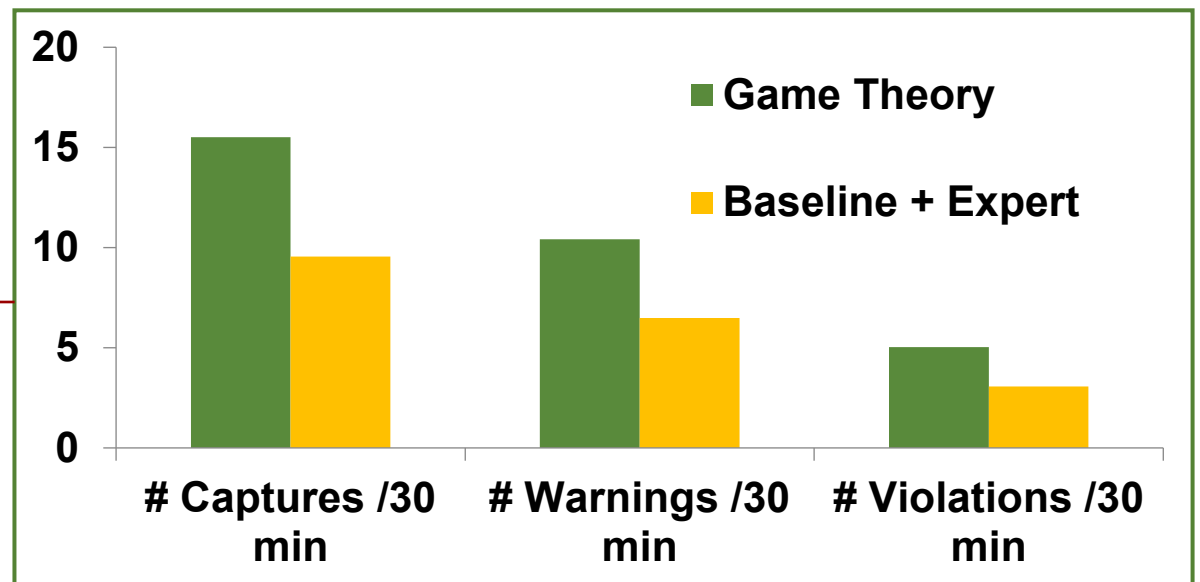
Field Tests Against Adversaries

Computational Game Theory in the Field

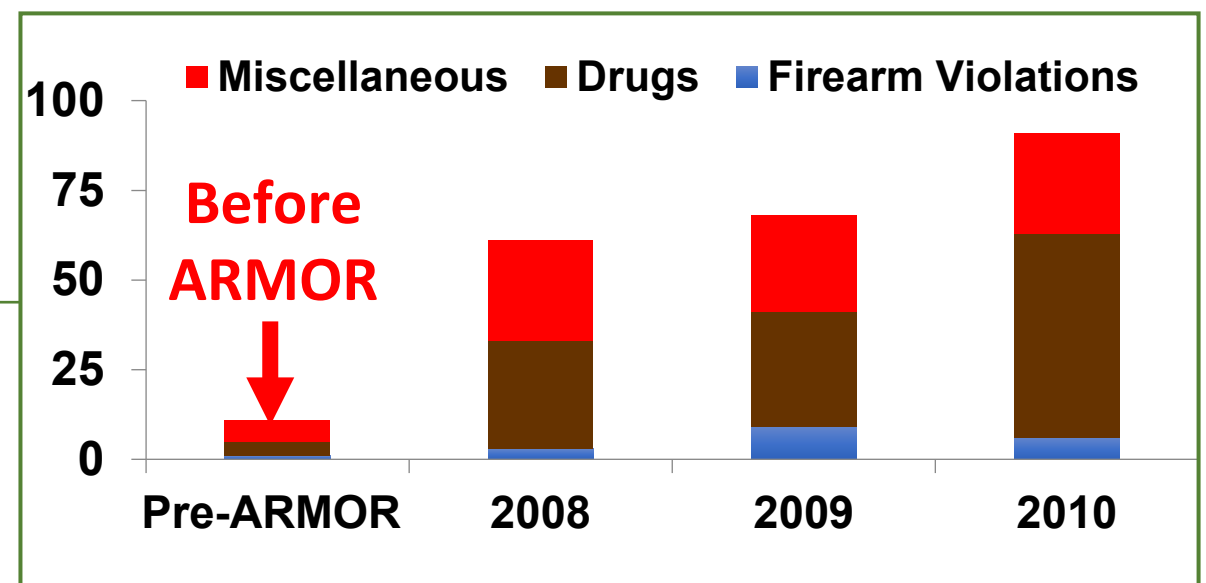
Controlled



- 21 days of patrol, identical conditions
- Game theory vs Baseline+Expert



Not Controlled



Outline

Public Safety and Security:
Stackelberg Security Games



Conservation/Wildlife Protection:
Green Security Games

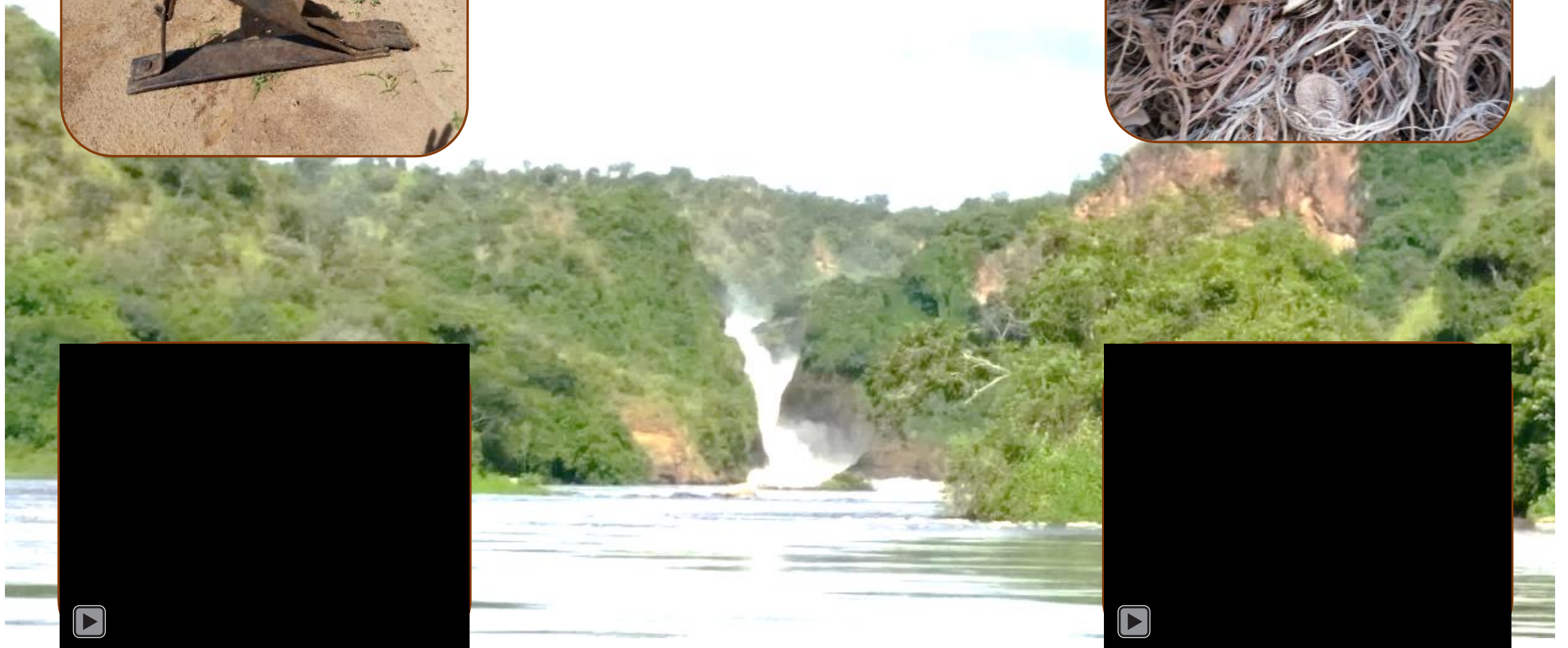
Poaching of Wildlife in Uganda

Limited Intervention (Ranger) Resources to Protect Forests

Snare or Trap

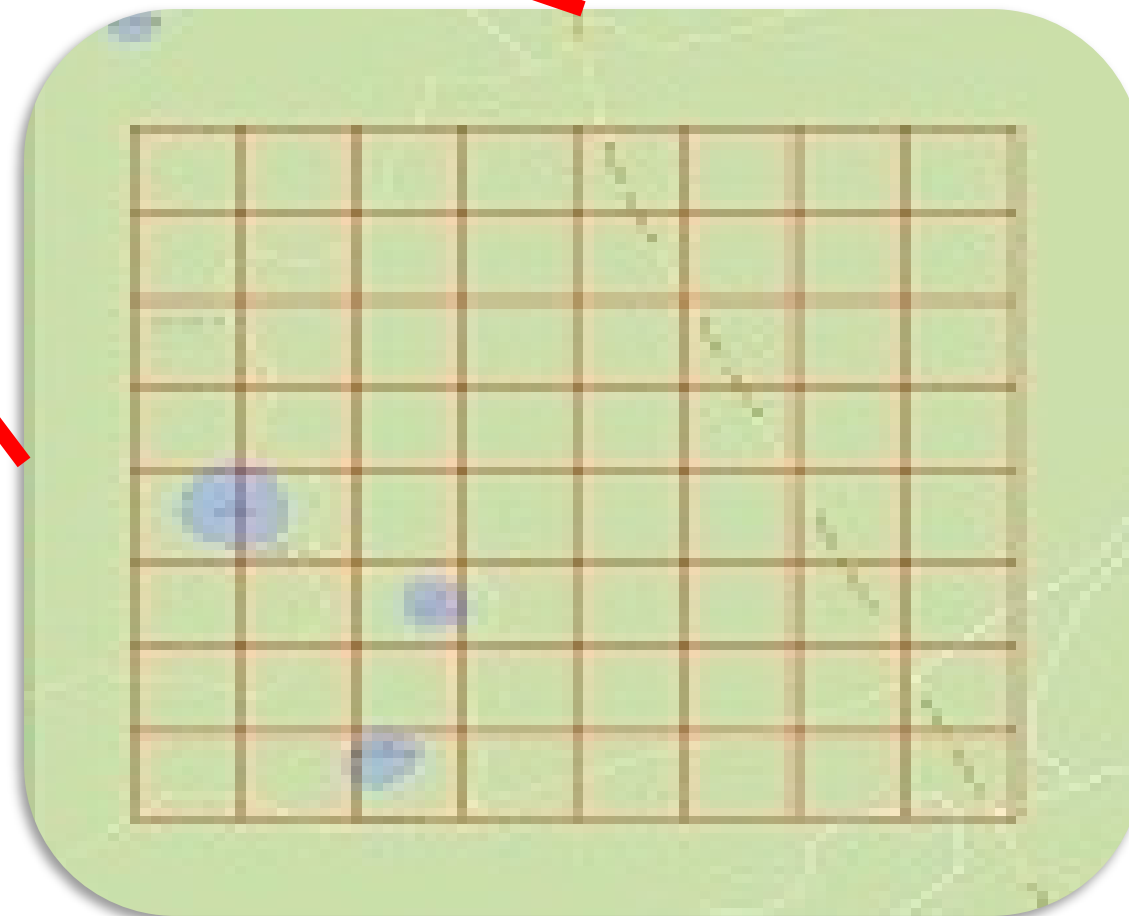


Wire snares



Focus on Threat prediction

Diving up the area into 1 Km x 1 Km Grid cells



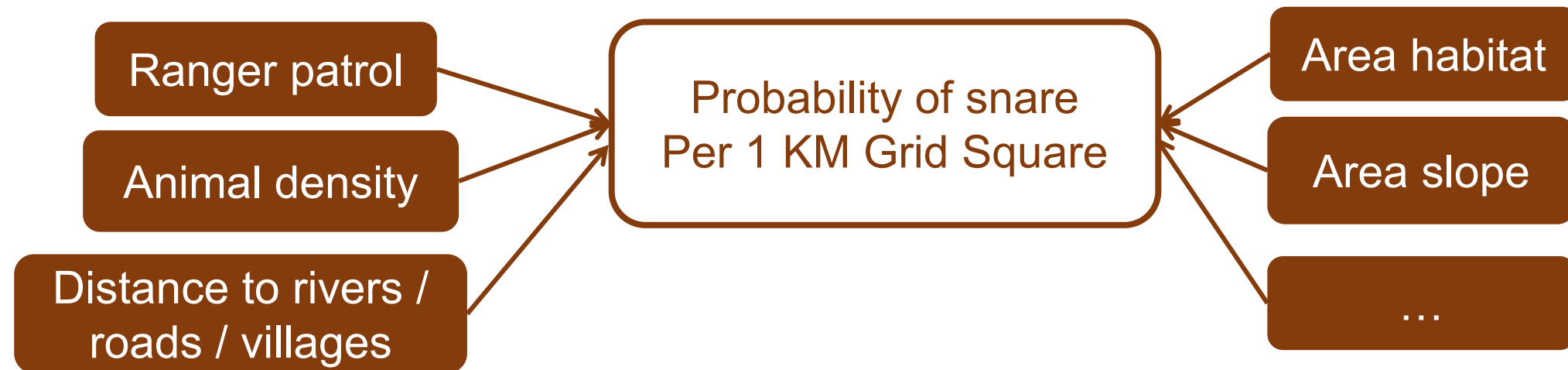
Learning Poacher Model: Uncertainty in Observations



Nguyen



Gholami



What uncertainty might you expect as rangers walk to recover snares and report that data?

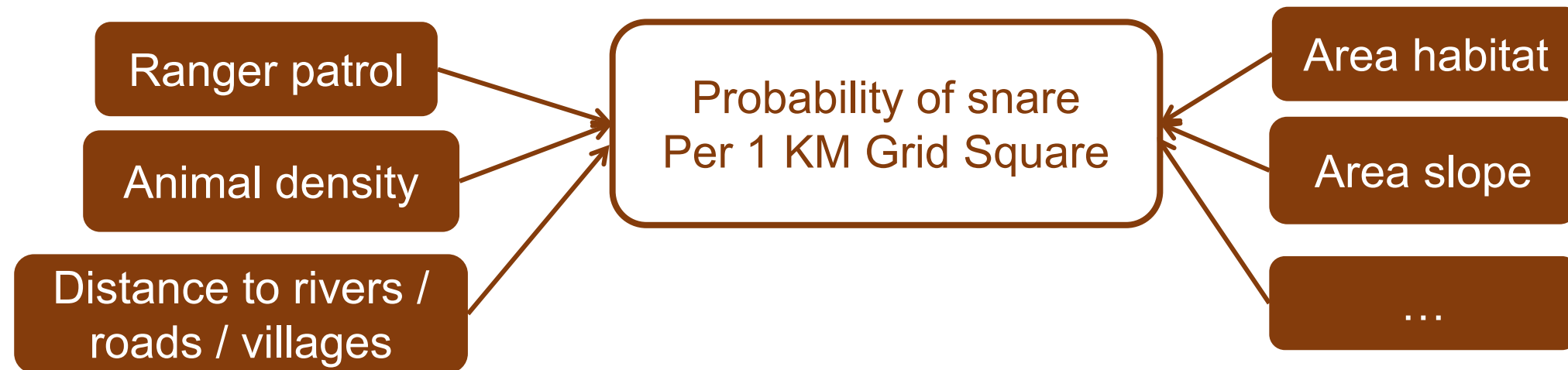
Learning Poacher Model: Uncertainty in Observations



Nguyen



Gholami



Positive and unlabeled datasets



Learning Adversary Response Model: Uncertainty in Observations



Nguyen



Gholami

