Empirical Game-Theoretic Analysis and the Behavior of Software Agents

Michael P. Wellman
University of Michigan

Previously at AIPS/ICAPS…

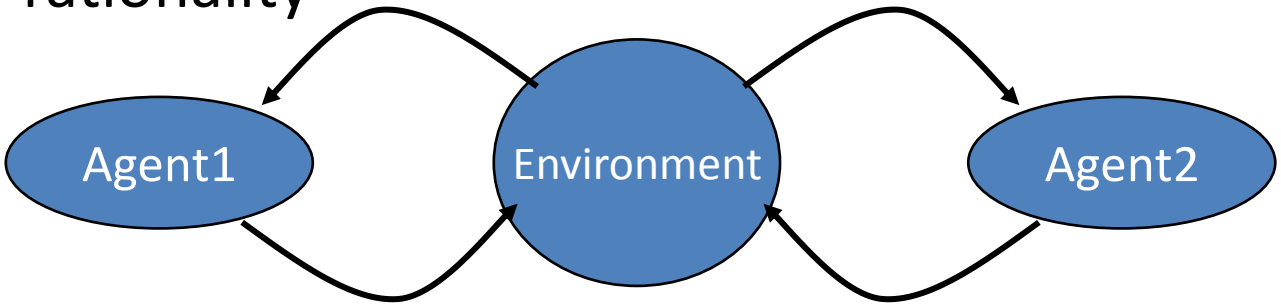
92: Modular utility representation for decision-theoretic planning (Wellman & Doyle)

04: Price prediction strategies for market-based scheduling (MacKie-Mason et al.)

04: Distributed feedback control for decision making on supply chains (Kiekintveld et al.)

Planning in Strategic Environments

- Planning problem
  - find agent behavior satisfying/optimizing objectives wrt environment
  - strives for rationality



- When environment contains other agents
  - model them as *rational planners* as well
  - problem is a game
  - search now multi-dimensional, different (global) objective

Real-World Games

complex dynamics and uncertainty

- rich strategy space
  - *strategy*: obs* × time → action
- severely incomplete information
  - interdependent types (signals)
  - info partially revealed over time

→ analytic game-theoretic solutions few and far between

two approaches

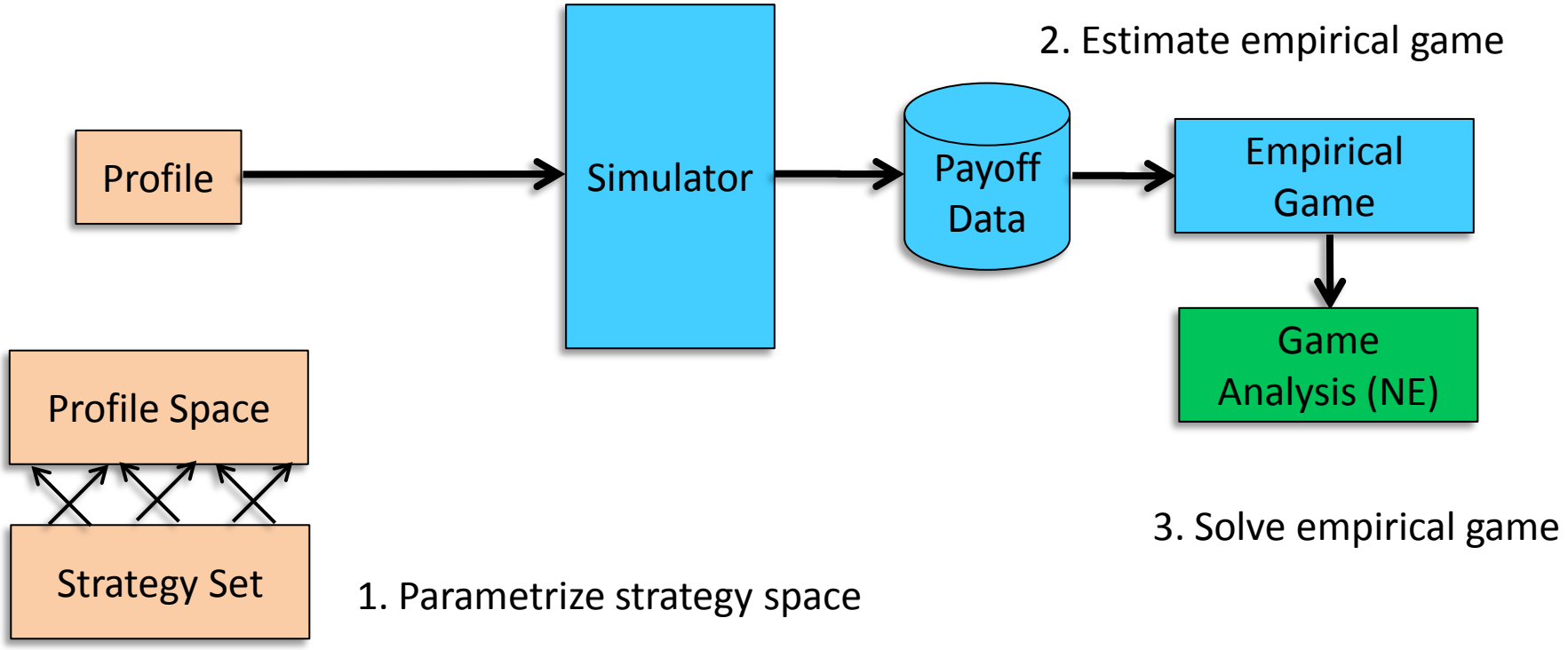
1. analyze (stylized) approximations
   - one-shot, complete info…
2. simulation-based methods
   - search
   - empirical: statistics, machine learning,…

Empirical Game-Theoretic Analysis (EGTA)

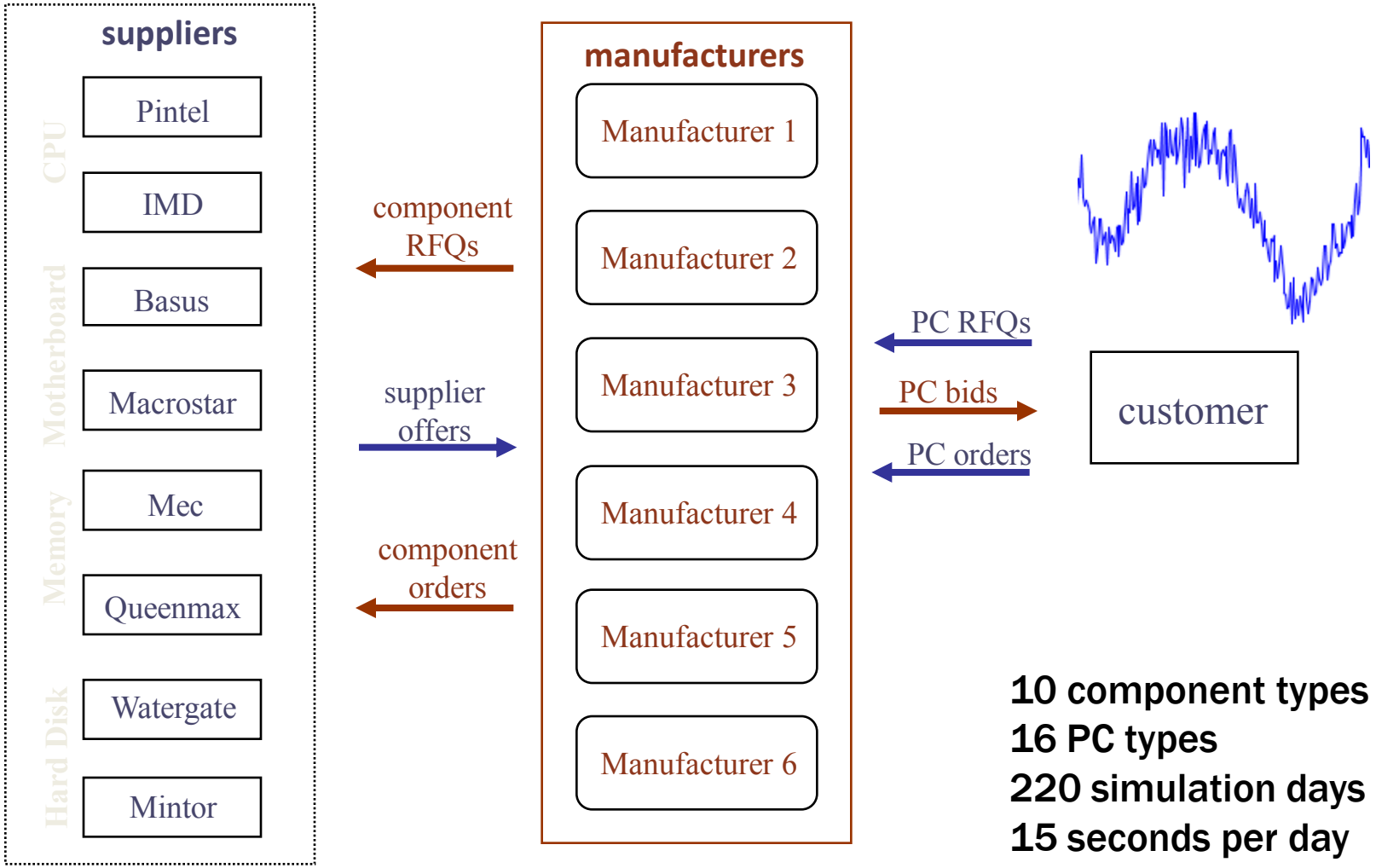
- Game described *procedurally*, no directly usable analytical form
- Parametrize strategy space based on agent architecture
- Selectively explore strategy/profile space
- Induce game model (payoff function) from simulation data

Empirical game

EGTA Process

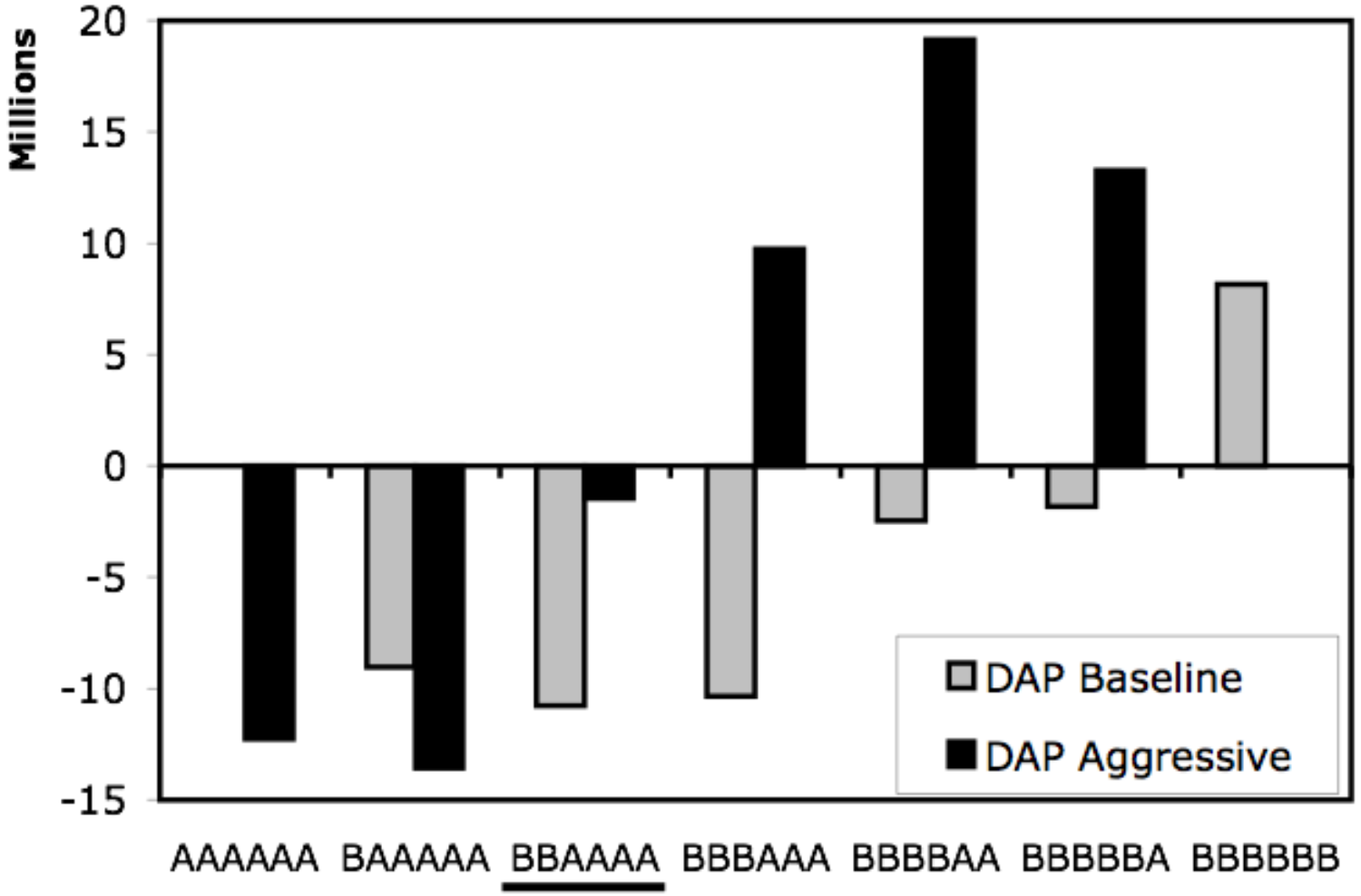


TAC Supply Chain Mgmt Game

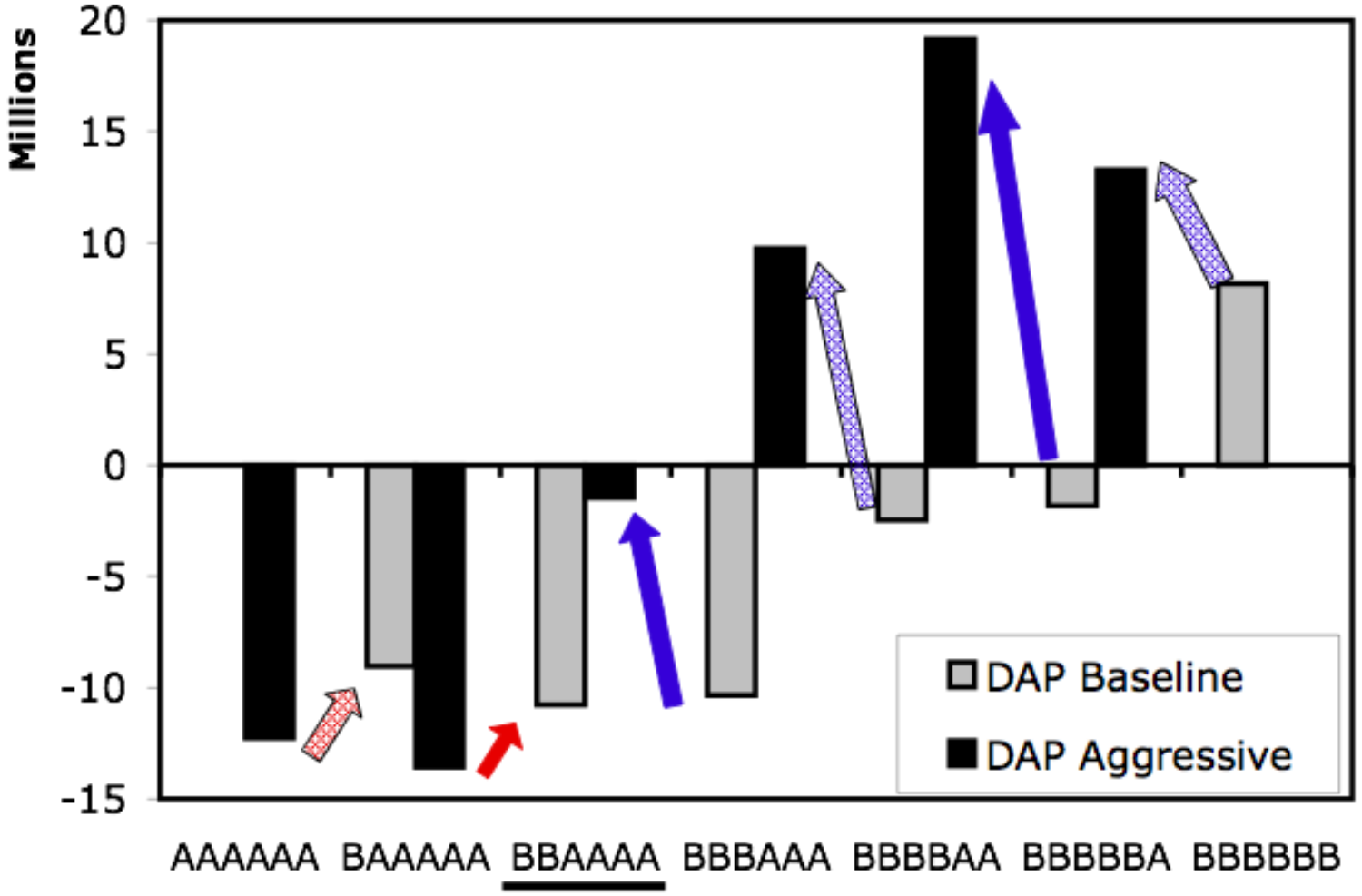


10 component types
16 PC types
220 simulation days
15 seconds per day

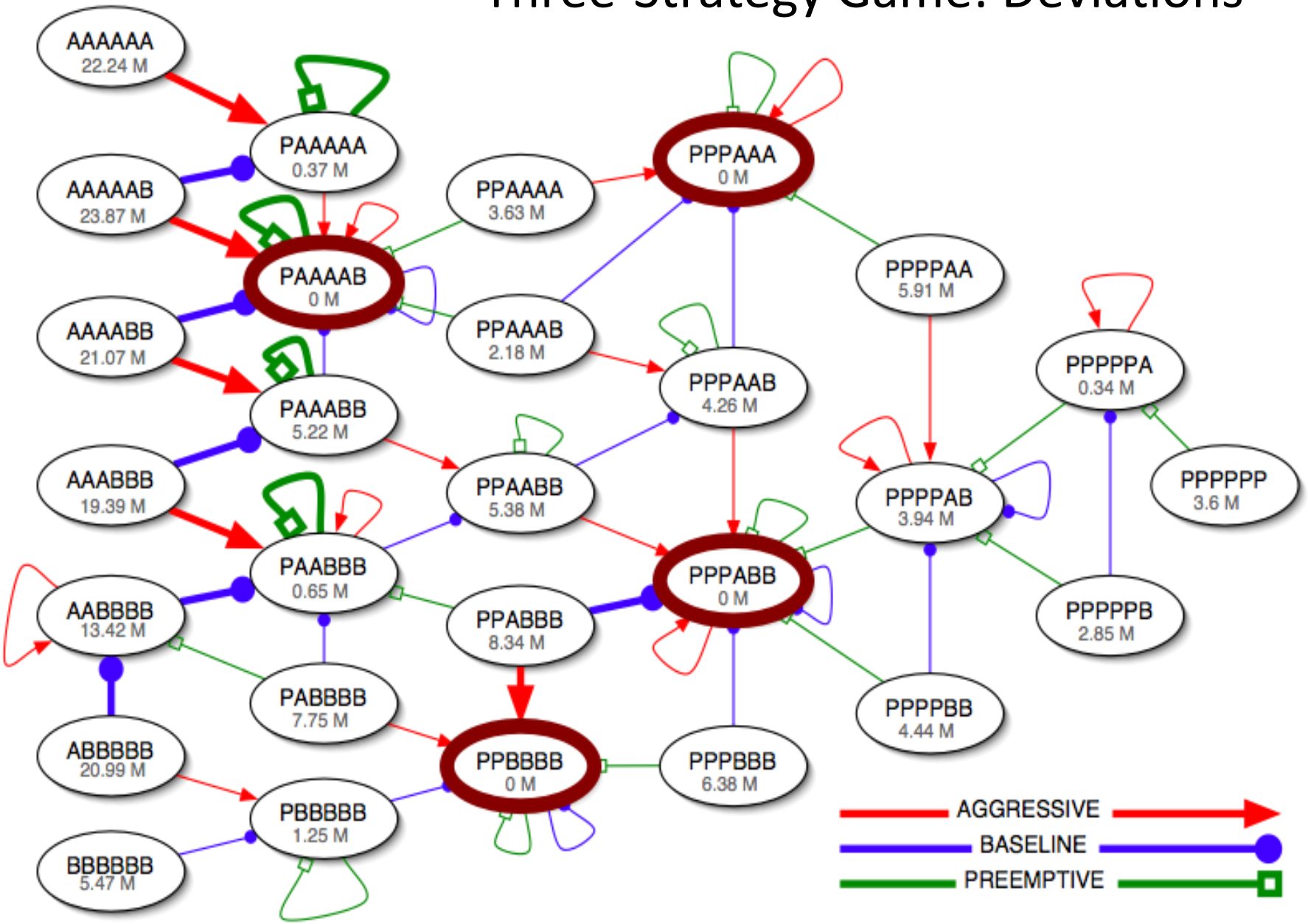
Two-Strategy Game (Unpreempted)



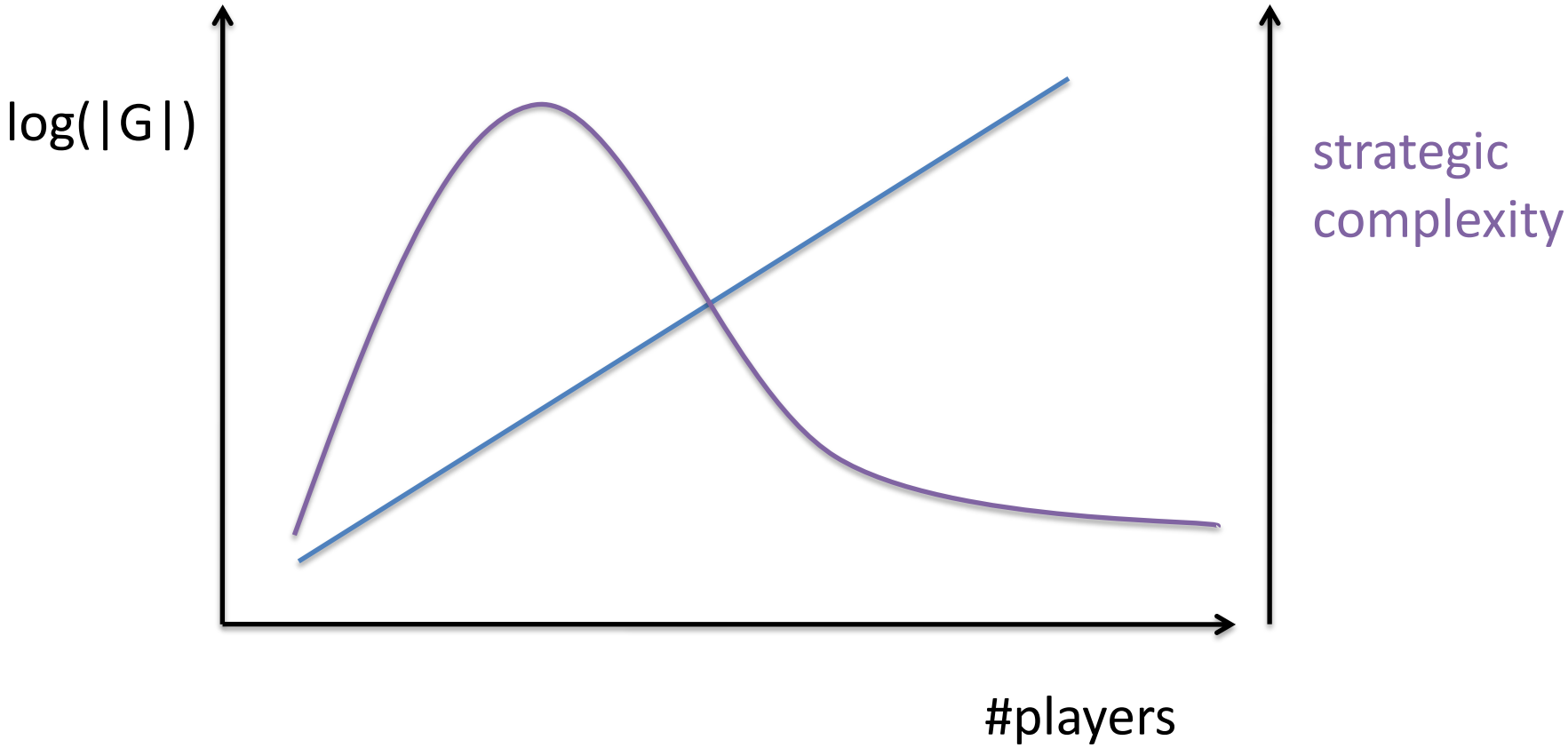
Two-Strategy Game (Unpreempted)



Three-Strategy Game: Deviations



Scaling #Players



Improving Scalability

- Exploit locality of interaction
  - graphical games, MAIDs, action-graph games, …
- Aggregate agents
  - hierarchical reduction (Wellman et al. AAAI-05)
  - clustering (Ficici et al. UAI-08)

Hierarchical Game Reduction

- *p*-player reduced version of symmetric game Γ

$$\Gamma\downarrow_p = \langle p, S, \hat{u}() \rangle$$

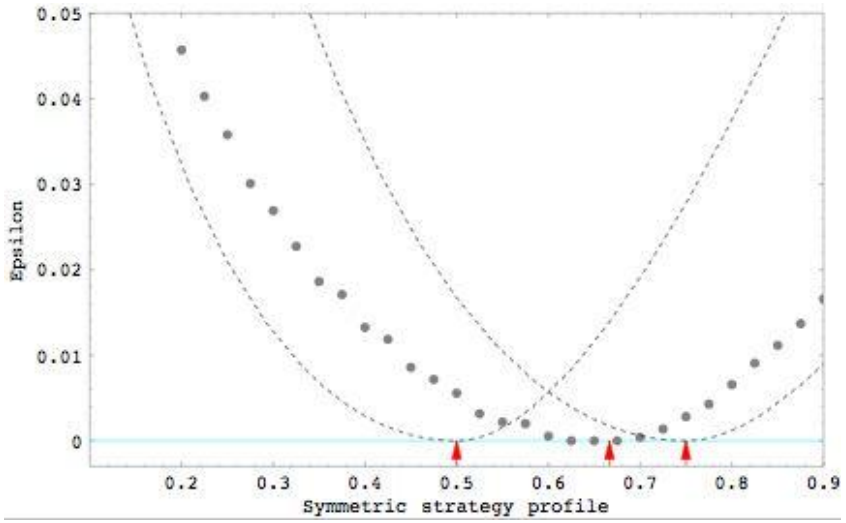
where

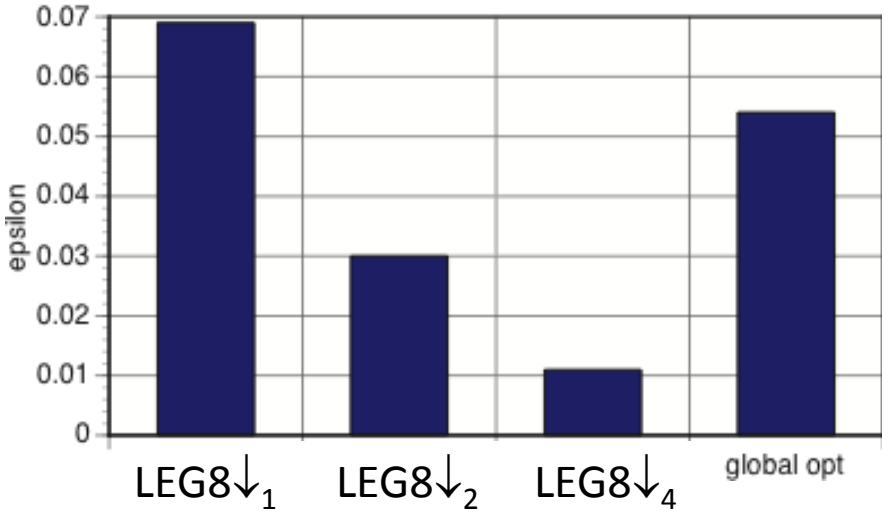
$$\hat{u}_i(s_1,\ldots,s_p) = u_{q\cdot i}(\underbrace{s_1}_{q},\ldots,\underbrace{s_2}_{q},\ldots,\ldots,\underbrace{s_p}_{q},\ldots)$$

Premise: Reduced game often a good approximation of original, with dramatically smaller profile space.

Why Trust Reduced-Game Results?

- Claim: Equilibria in reduced game likely to be relatively stable in full game
- Evidence:
  - Random instances of local-effect games (LEGs)
  - FPSB auctions

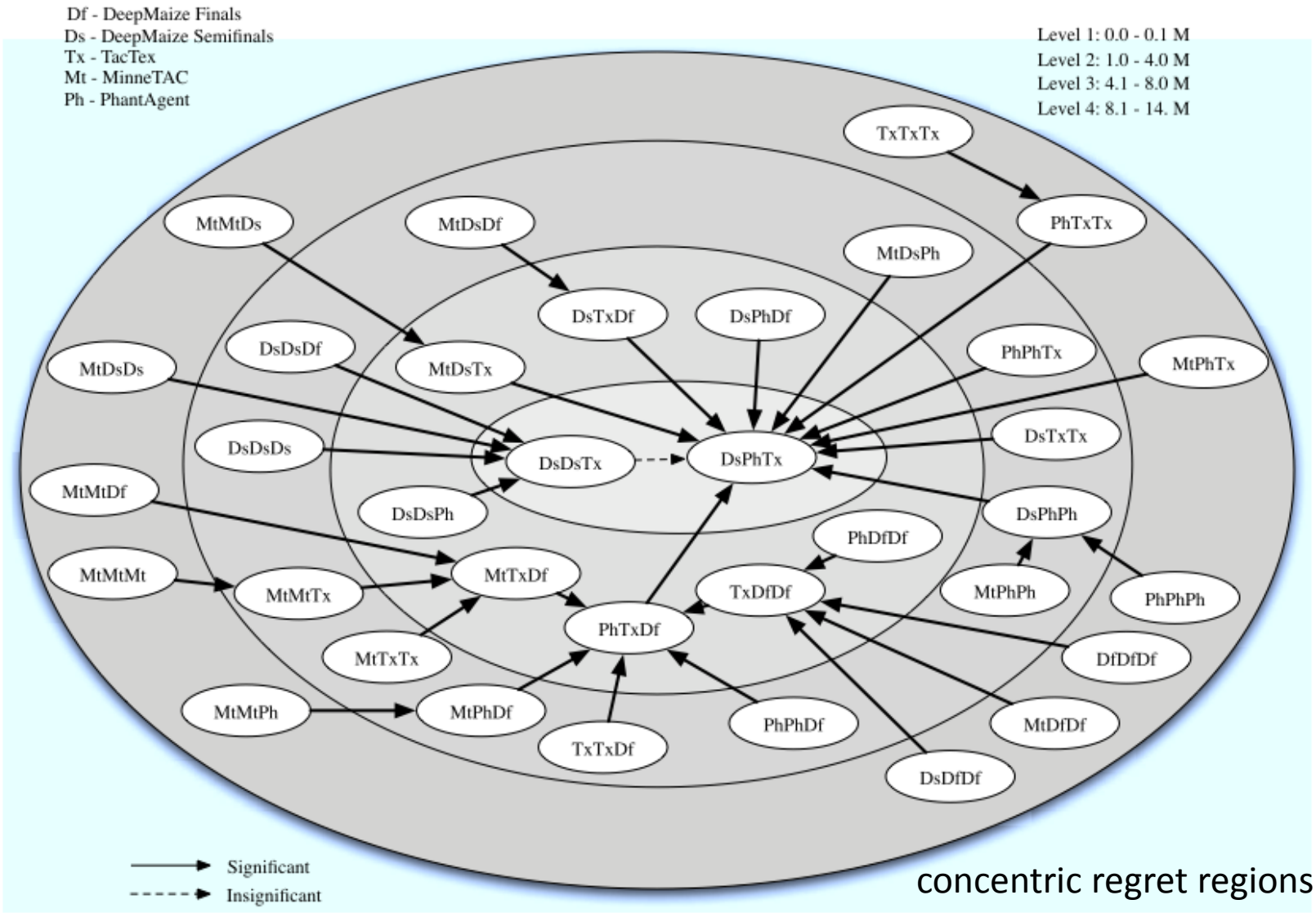




Research Questions

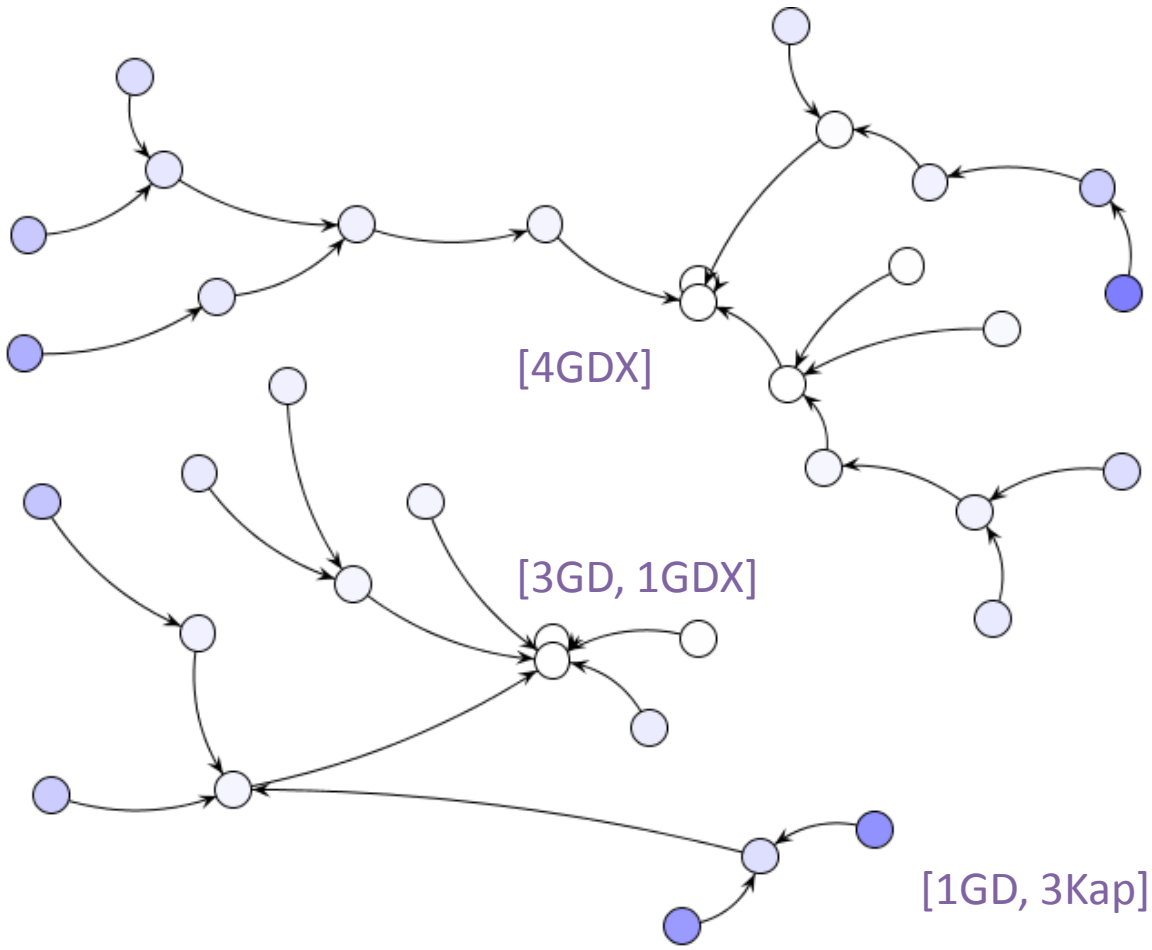
- What is the space of reduction aggregations?
  - and which are most effective for what classes of games?
- How to adjust for systematic biases of reduction?
- How to automatically cluster agents in non-symmetric games?
- How to reason at multiple levels of aggregation?

TAC/SCM-06 Deviation Graph



concentric regret regions

CDA Deviation Graph



4 strategies:
GD, GDX, ZI, Kap

Ranking Strategies: TAC/SCM-07

SCM-07 Tournament

Agent	Finals	Semifinals
PhantAgent	8.67	10.38 [2]
TacTex	6.31	5.75 [2]
DeepMaize	5.45	9.759 [1]
Maxon	1.79	5.631 [1]
Tinhorn	1.34	6.94 [1]
CMieux	1.24	2.66 [2]

SCM-07 EGTA

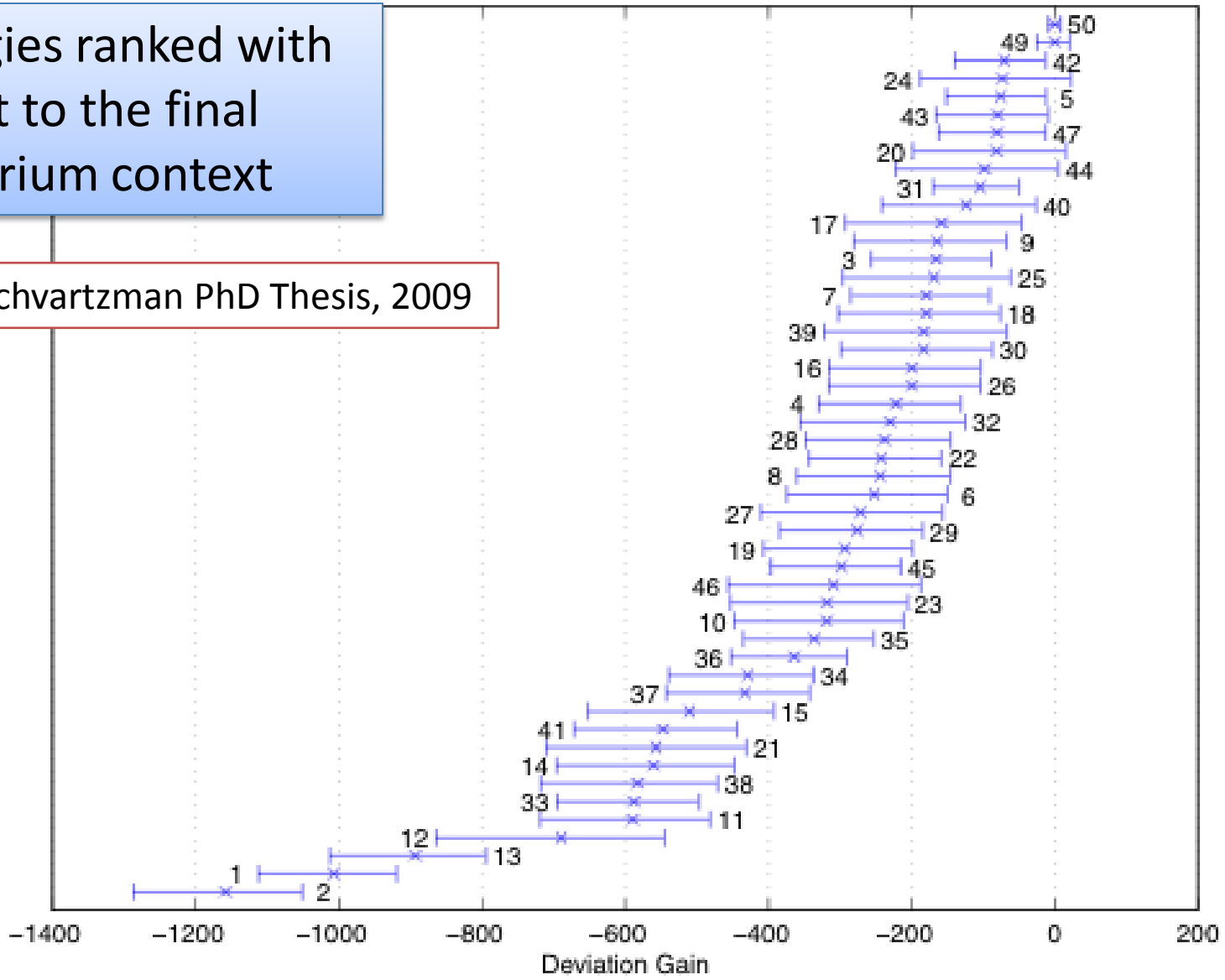
Agent	NE Regret	Max Regret
DM07 S [C07-9]	0.32	3.40
DM07 F [C07-34]	0	2.63
PH07	0	48.84
TT07 S	2.90	16.95
TT07 F	0	10.89
DM06 S	3.21	8.17
PH06	1.31	11.00
TT06	1.03	14.78
MR05	2.98	14.67

from PR Jordan PhD Thesis, 2009

Strategy Ranking (TAC Travel)

Strategies ranked with respect to the final equilibrium context

from LJ Schvartzman PhD Thesis, 2009



Strategy Ranking (CDA)

strategy	NE1 regret	NE2 regret	symm. profile payoff
GDX	0	1.32	247.98
GD	0.49	3.26	248.57
RB	2.20	8.64	248.08
ZIP	2.90	9.86	247.95
Kaplan	4.56	24.55	2.02
ZIbtq	14.67	17.44	247.45
ZI	16.42	16.82	248.07

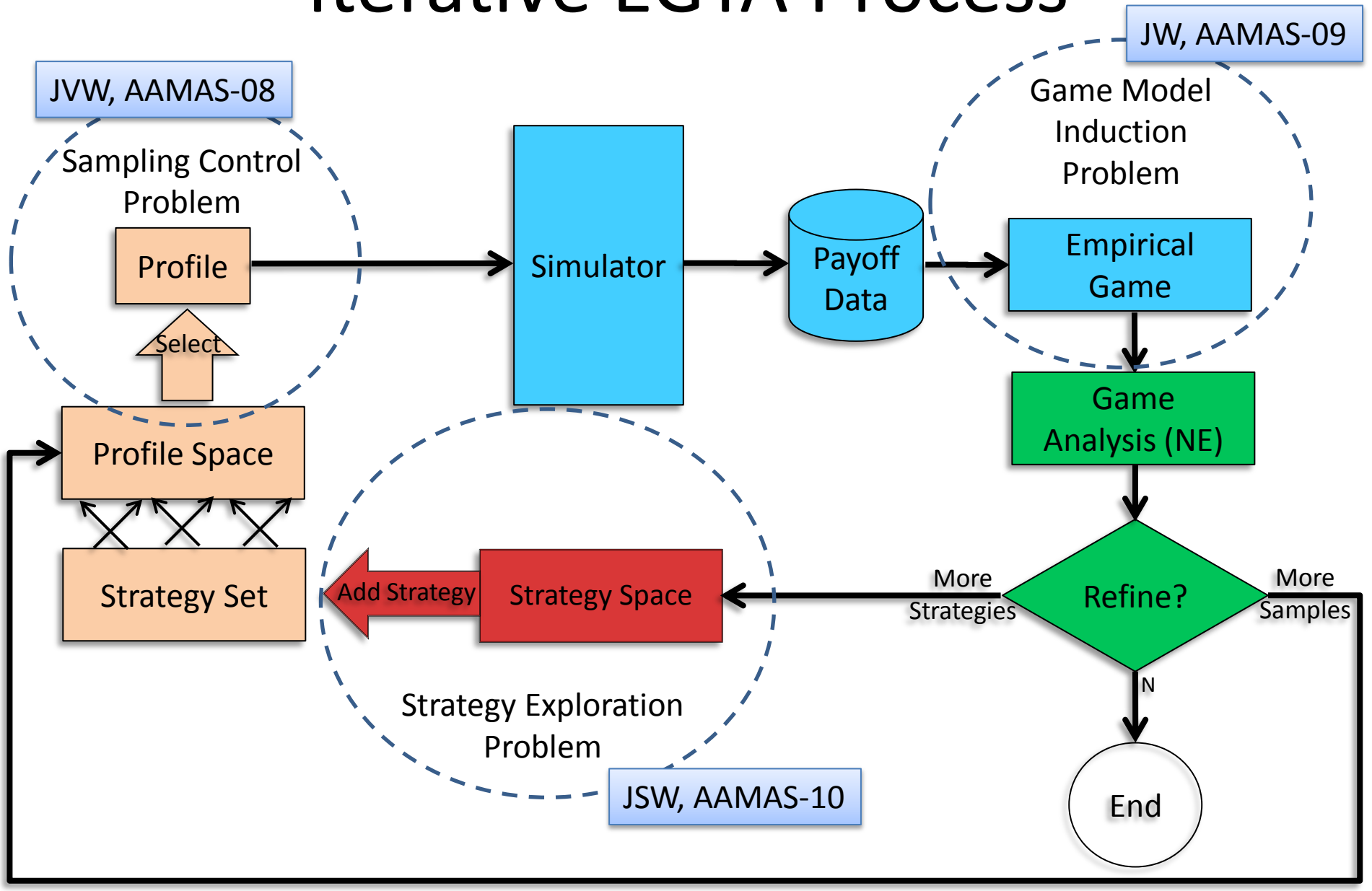
DeepMaize-08 Design Exploration

	Predictions					*Controller*					
	Customer Dataset		Component Horizon Treatment			Bid Improvement			EG Procurement		
ID	SCM05	SCM[06-07]	AIO	INTRP-Bug	INTERP	EQ	SA	GA	07	PH	07+
0	✓		✓			✓			✓		
1	✓			✓		✓			✓		
2		✓	✓			✓			✓		
3-5		✓		✓		✓			✓		
6		✓		✓		✓				✓	
7		✓		✓		✓					✓
8		✓		✓			✓			✓	
9		✓		✓				✓		✓	
10	✓	✓		✓		✓				✓	
11	✓				✓	✓			✓		
13		✓			✓			✓			✓
14		✓			✓	✓				✓	
15		✓	✓			✓				✓	
17-27		✓		✓		✓				✓	
28-29		✓			✓	✓				✓	

Table 5.15: DeepMaize 08 tested feature matrix.

from PR Jordan PhD Thesis, 2009

Iterative EGTA Process



Sampling Control Problem

- Revealed payoff model
  - sample provides exact payoff
  - minimum-regret-first search (MRFS)
    - attempts to refute best current candidate
- Noisy payoff model
  - sample drawn from payoff distribution
  - information gain search (IGS)
    - sample profile maximizing entropy difference wrt probability of being min-regret profile

Min-Regret-First Search

start
(arbitrary)

	c1	*c2*	*c3*	*c4*
r1	9,5			
r2				
r3				
r4				

Profile	*ε-bound*
(r1,c1)	0

Min-Regret Search

evaluated

best

	c1	*c2*	*c3*	*c4*
r1	9,5	3,3		
r2				
r3				
r4				

Profile	*ε-bound*
(r1,c1)	0
(r1,c2)	2

Select random deviation from current best profile

Min-Regret Search

evaluated

best

	c1	*c2*	*c3*	*c4*
r1	9,5	3,3		
r2	6,4			
r3				
r4				

Profile	*ε-bound*
(r1,c1)	0
(r1,c2)	2
(r2,c1)	3

Min-Regret Search

evaluated

best

	c1	*c2*	*c3*	*c4*
r1	9,5	3,3		
r2	6,4			
r3	2,2			
r4				

Profile	*ε-bound*
(r1,c1)	0
(r1,c2)	2
(r2,c1)	3
(r3,c1)	7

Min-Regret Search

evaluated

best

	c1	*c2*	*c3*	*c4*
r1	9,5	3,3		4,8
r2	6,4			
r3	2,2			
r4				

Profile	*ε-bound*
(r1,c1)	3
(r1,c2)	5
(r2,c1)	3
(r3,c1)	7
(r1,c4)	0

Min-Regret Search

evaluated

best

	c1	*c2*	*c3*	*c4*
r1	9,5	3,3		4,8
r2	6,4			5,3
r3	2,2			
r4				

Profile	*ε-bound*
(r1,c1)	3
(r1,c2)	5
(r2,c1)	3
(r3,c1)	7
(r1,c4)	1
(r2,c4)	1

Min-Regret Search

evaluated

best

	c1	*c2*	*c3*	*c4*
r1	9,5	3,3		4,8
r2	6,4	8,8		5,3
r3	2,2			
r4				

Profile	*ε-bound*
(r1,c1)	3
(r1,c2)	5
(r2,c1)	4
(r3,c1)	7
(r1,c4)	1
(r2,c4)	5
(r2,c2)	0

Min-Regret Search

evaluated

best

	c1	*c2*	*c3*	*c4*
r1	9,5	3,3		4,8
r2	6,4	8,8	3,0	5,3
r3	2,2			
r4				

Profile	*ε-bound*
(r1,c1)	3
(r1,c2)	5
(r2,c1)	4
(r3,c1)	7
(r1,c4)	1
(r2,c4)	5
(r2,c2)	0
(r2,c3)	8

Min-Regret Search

evaluated

best

	c1	*c2*	*c3*	*c4*
r1	9,5	3,3		4,8
r2	6,4	8,8	3,0	5,3
r3	2,2	2,1		
r4				

Profile	*ε-bound*
(r1,c1)	3
(r1,c2)	5
(r2,c1)	4
(r3,c1)	7
(r1,c4)	1
(r2,c4)	5
(r2,c2)	0
(r2,c3)	8
(r3,c2)	6

Min-Regret Search

evaluated

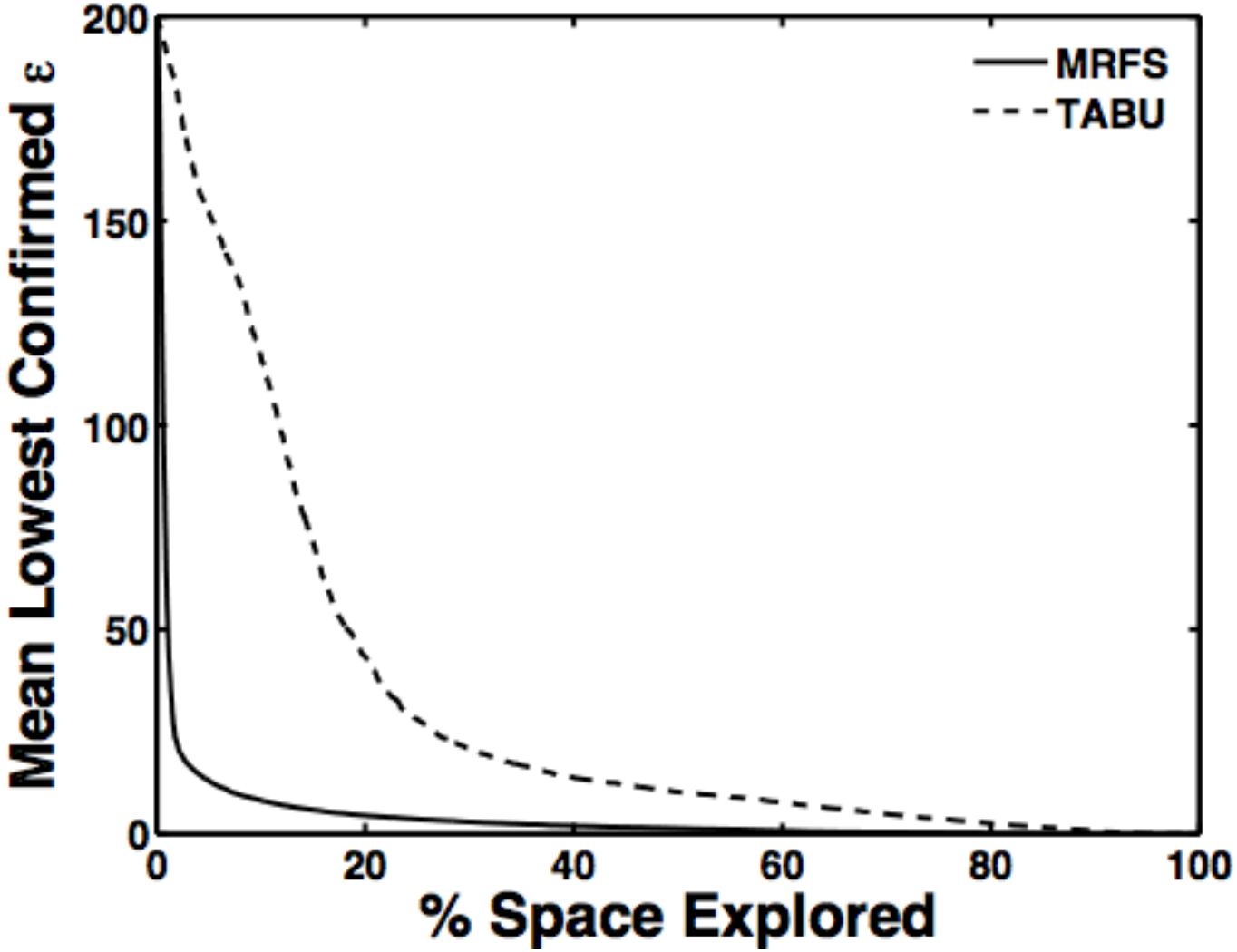
best

NE Confirmed!

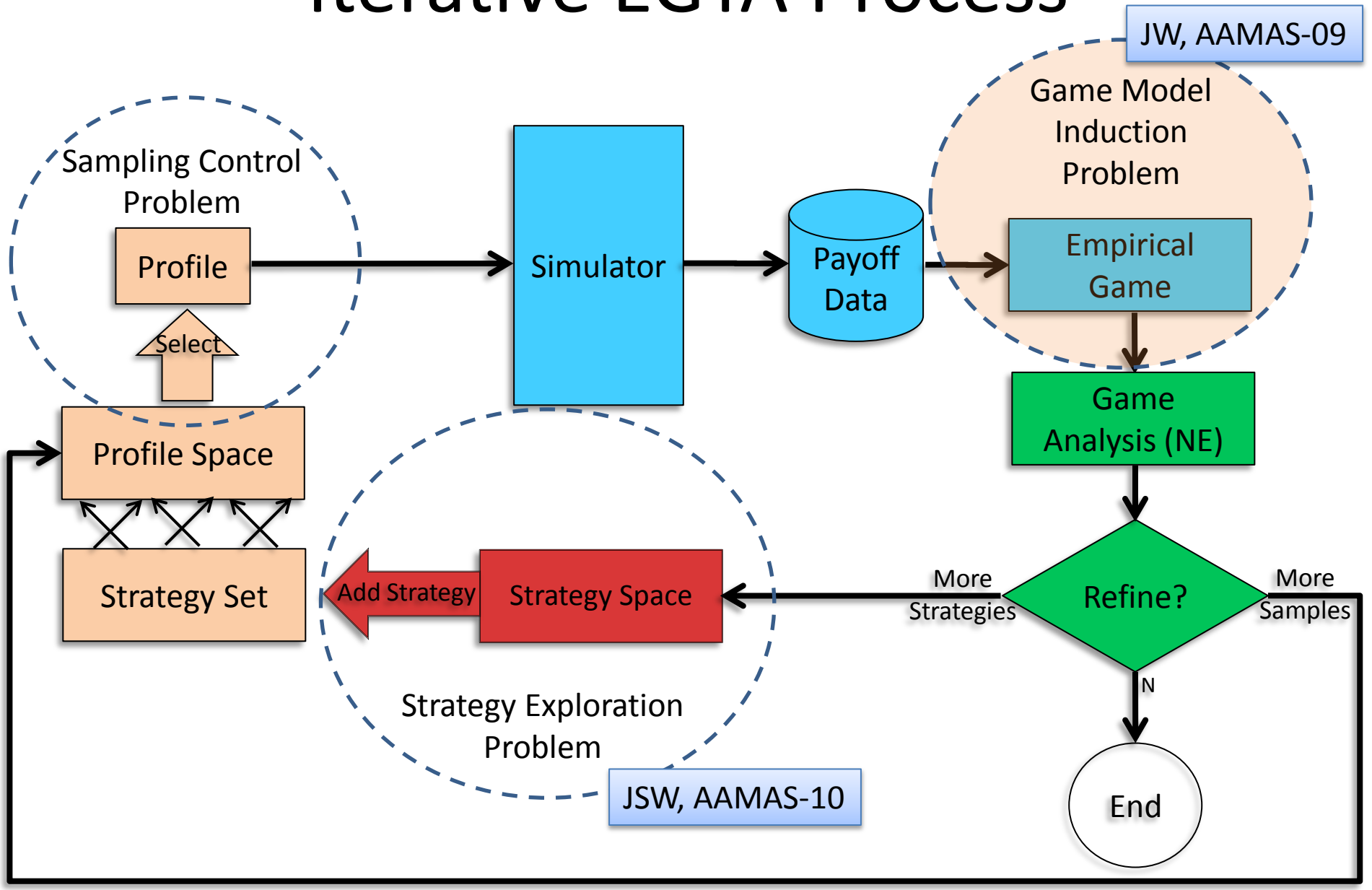
	c1	*c2*	*c3*	*c4*
r1	9,5	3,3		4,8
r2	6,4	8,8	3,0	5,3
r3	2,2	2,1		
r4		2,0		

Profile	*ε-bound*
(r1,c1)	3
(r1,c2)	5
(r2,c1)	4
(r3,c1)	7
(r1,c4)	1
(r2,c4)	5
(r2,c2)	0*
(r2,c3)	8
(r3,c2)	6
(r4,c2)	6

Finding Approximate PSNE

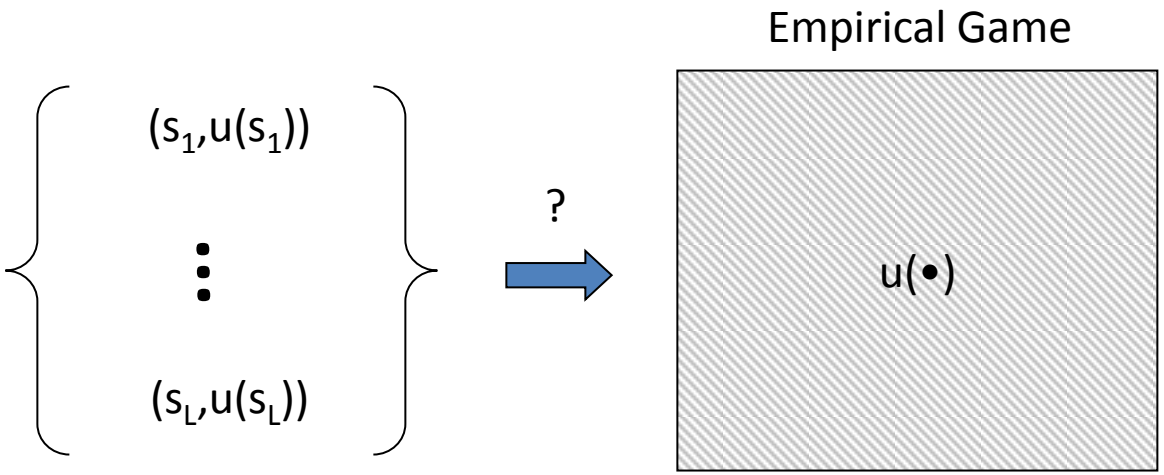


Iterative EGTA Process



Construct Empirical Game

- Simplest approach: direct estimation
  - employ control variates and other variance reduction techniques



Payoff data from selected profiles

Payoff Function Regression

$S_i = [0,1]$

↓ generate data (simulations)

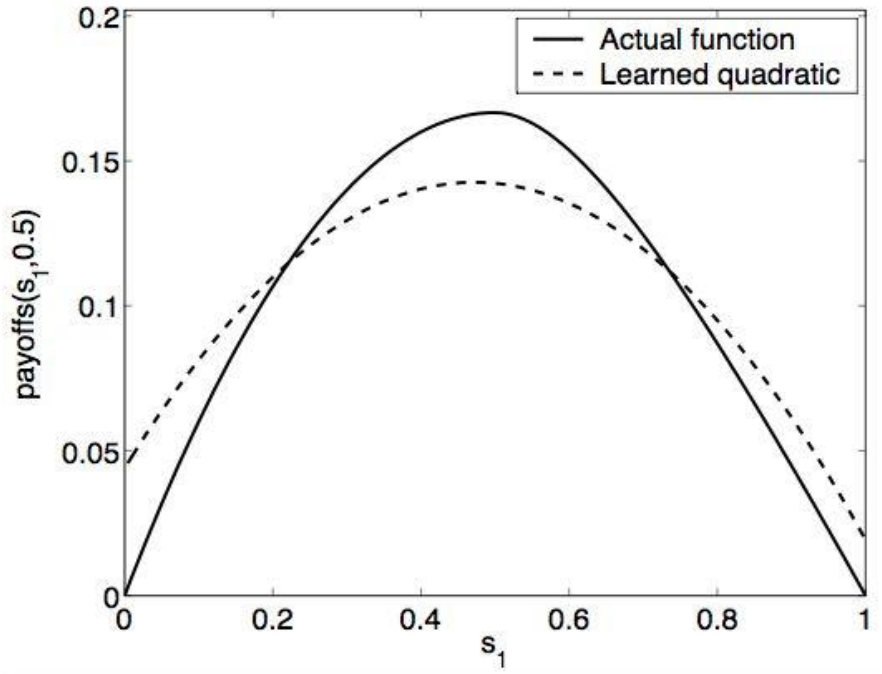
	0	0.5	1
0	3,3	1,4	1,1
0.5	4,1	2,2	4,1
1	1,1	1,0	3,3

↓ learn regression

↓ solve learned game

eq = (0.32,0.32)

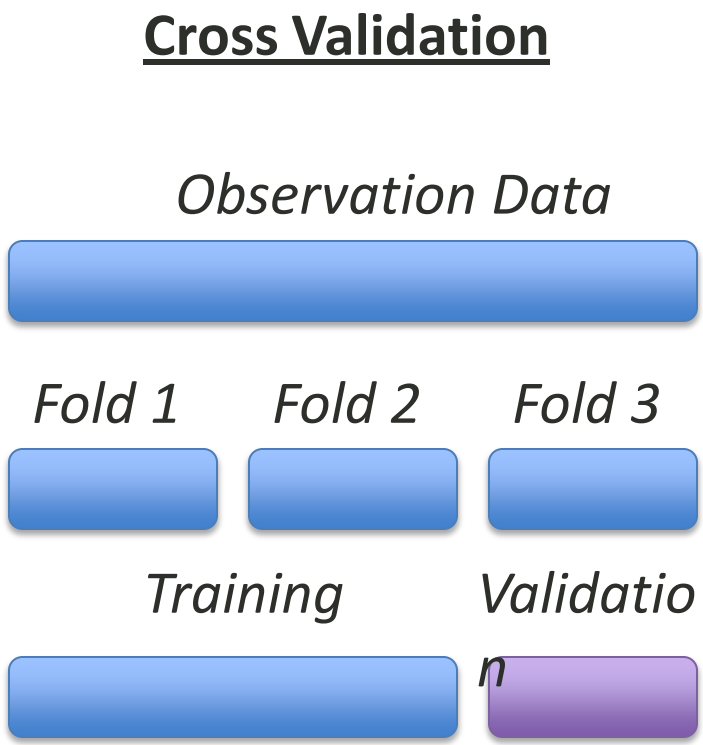
FPSB2 Example



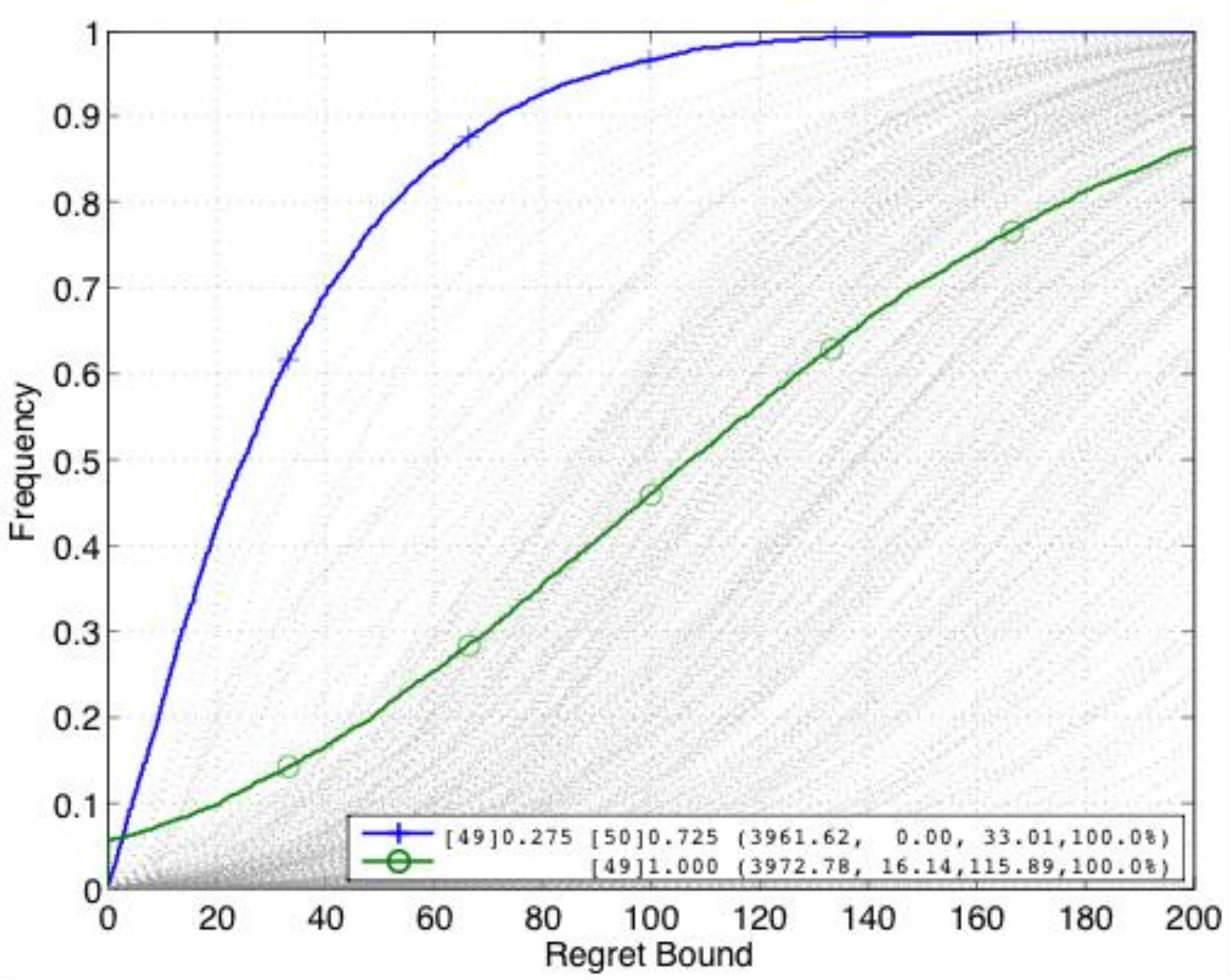
Vorobeychik et al., *ML* 2007

Generalization Risk Approach

- Model variations
  - functional forms, relationship structures, parameters
  - strategy granularity
- Approach:
  - Treat candidate game model as a predictor for payoff data
  - Adopt loss function for predictor
  - Select model candidate minimizing expected loss

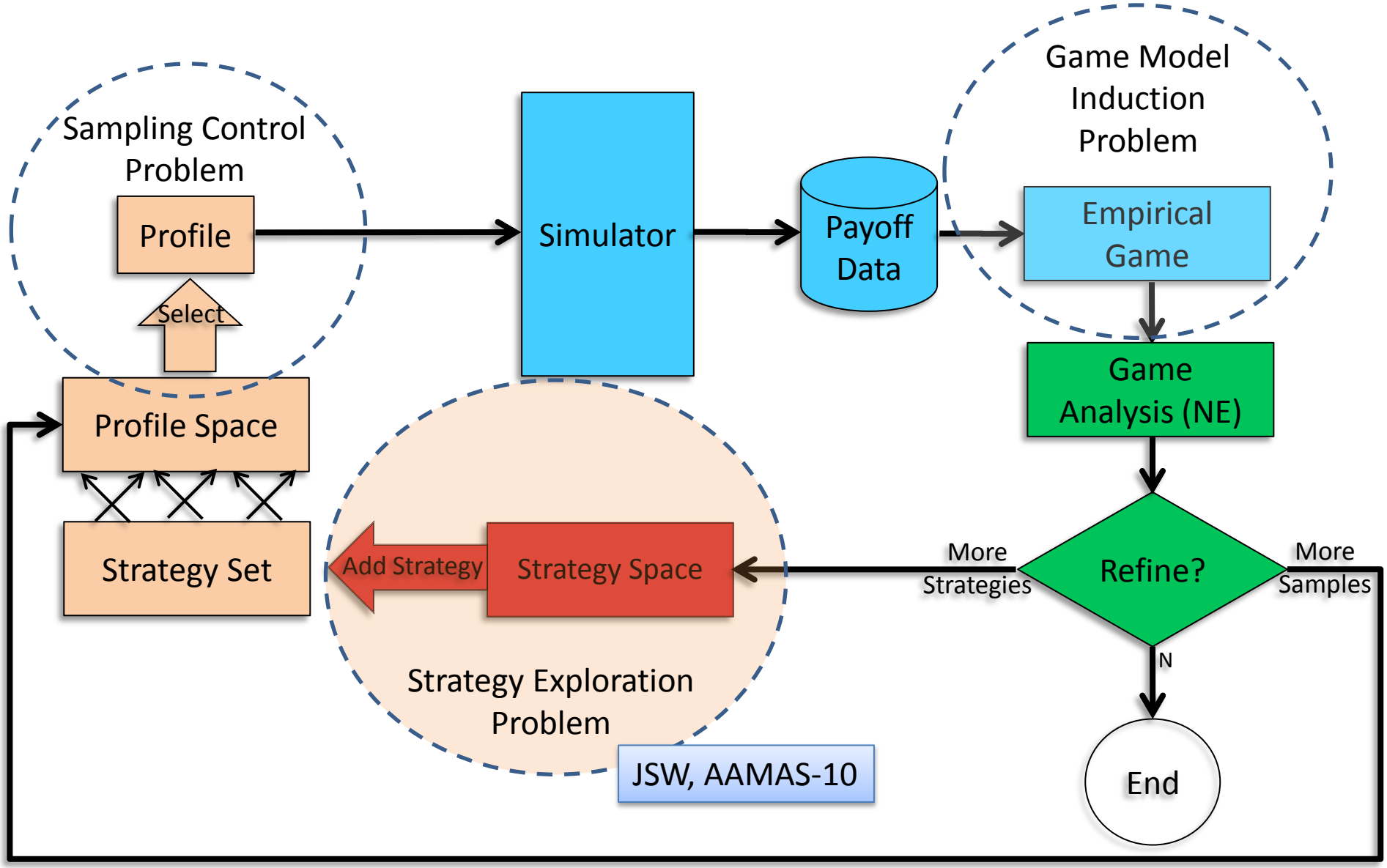


Sensitivity Analysis

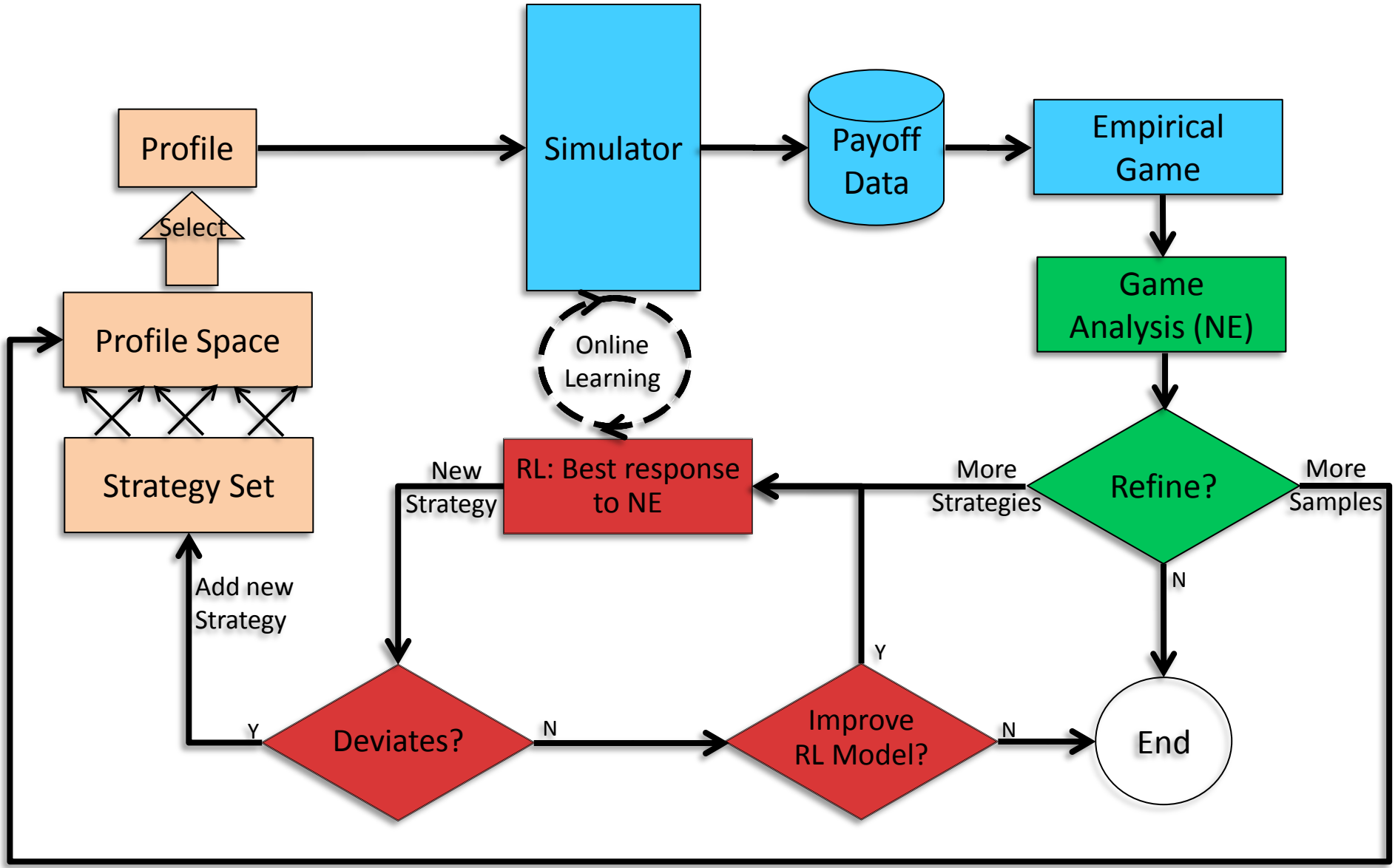


392 two-strategy mixtures

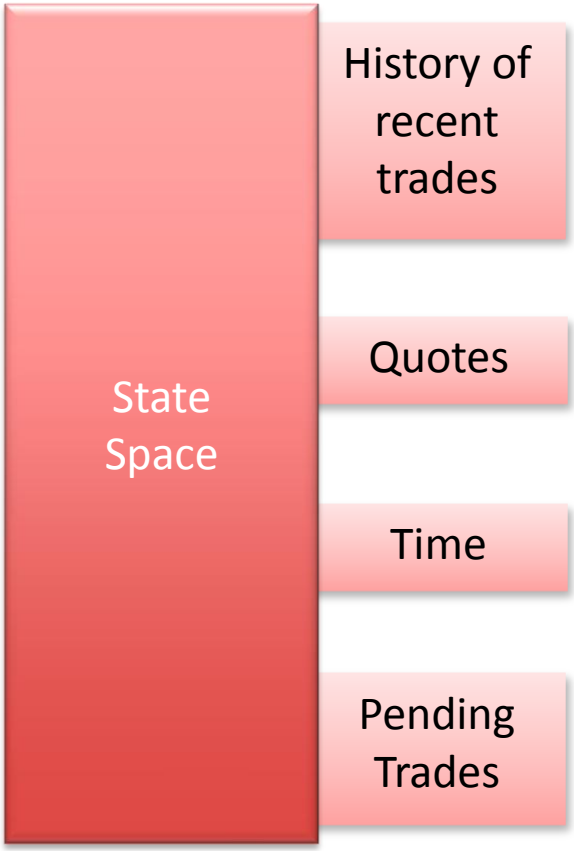
Iterative EGTA Process



Learning New Strategies: EGTA+RL



CDA Learning Problem Setup



H_1: Moving average
H_2: Frequency weighted ratio, threshold= V
H_3: Frequency weighted ratio, threshold= A

Q_1: Opposite role
Q_2: Same role

T_1: Total
T_2: Since last trade

U: Number of trades left
V: Value of next unit to be traded

Actions

A: Offset from V

Rewards

R: Difference between unit valuation and trade price

EGTA/RL Round 1

Strategies	Payoff	NE	Learning	
			Strategy	Dev. Payoff
Kaplan ZI ZIbtq	248.1	1.000 ZI	L1	268.7
L1	242.5	1.000 L1		

EGTA/RL Round 2

Strategies	Payoff	NE	Learning	
			Strategy	Dev. Payoff
Kaplan ZI ZIbtq	248.1	1.000 ZI	L1	268.7
L1	242.5	1.000 L1		
ZIP	248.0	1.000 ZIP		
GD	248.6	1.000 GD	L2-L8 L9	--- 251.8
L9	246.1	0.531 GD 0.469 L9	L10	252.1

EGTA/RL Rounds 3+

Strategies	Payoff	NE	Learning	
			Strategy	Dev. Payoff
…	…	…	…	…
L10	248.0	0.191 GD 0.809 L10	L11	251.0
L11	246.2	1.000 L11		
GDX	245.8	0.192 GDX 0.808 L11	L12	248.3
L12	245.8	0.049 L11 0.951 L12	L13	245.9
L13	245.6	0.872 L12 0.128 L13	L14	245.6
RB	245.6	0.872 L12 0.128 L13		

Final champion

Strategy Exploration Problem

- Premise:
  - Limited ability to cover profile space
  - Expectation to reasonably evaluate all considered strategies
- Need deliberate policy to decide which strategies to introduce
- RL for strategy exploration
  - attempt at best response to current equilibrium
  - is this a good heuristic (even assuming ideal BR calc?)

Example

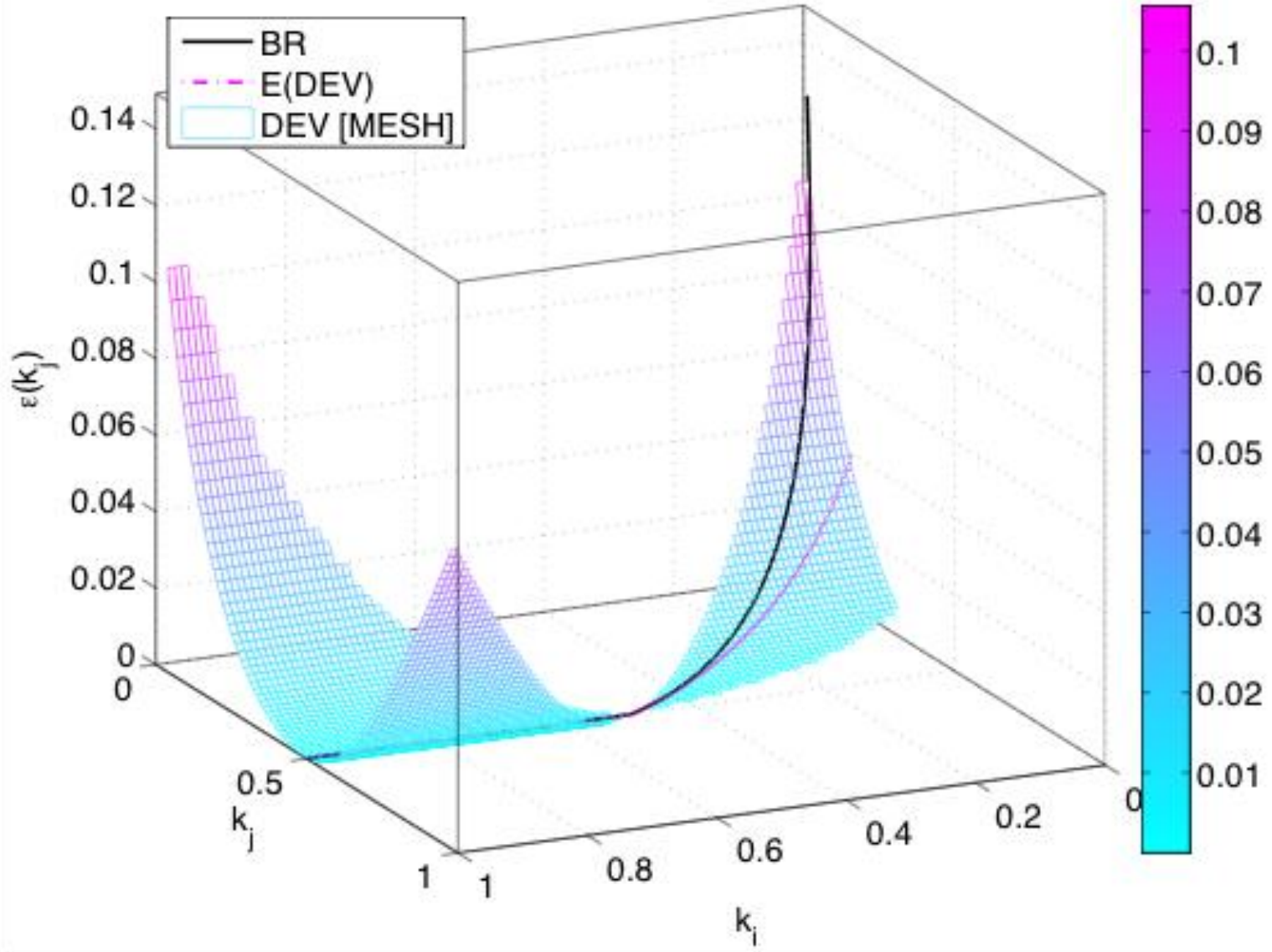
Introduce strategies in order:
A1, A2, A3, A4

Regret may *increase* over subsequent steps!

	A1	A2	A3	A4
A1	1, 1	1, 2	1, 3	1, 4
A2	2, 1	2, 2	2, 3	2, 6
A3	3, 1	3, 2	3, 3	3, 8
A4	4, 1	6, 2	8, 3	4, 4

Strategy Set	Candidate Eq.	Regret wrt True Game
{A1}	(A1,A1)	3
{A1,A2}	(A2,A2)	4
{A1,A2,A3}	(A3,A3)	5
{A1,A2,A3,A4}	(A4,A4)	0

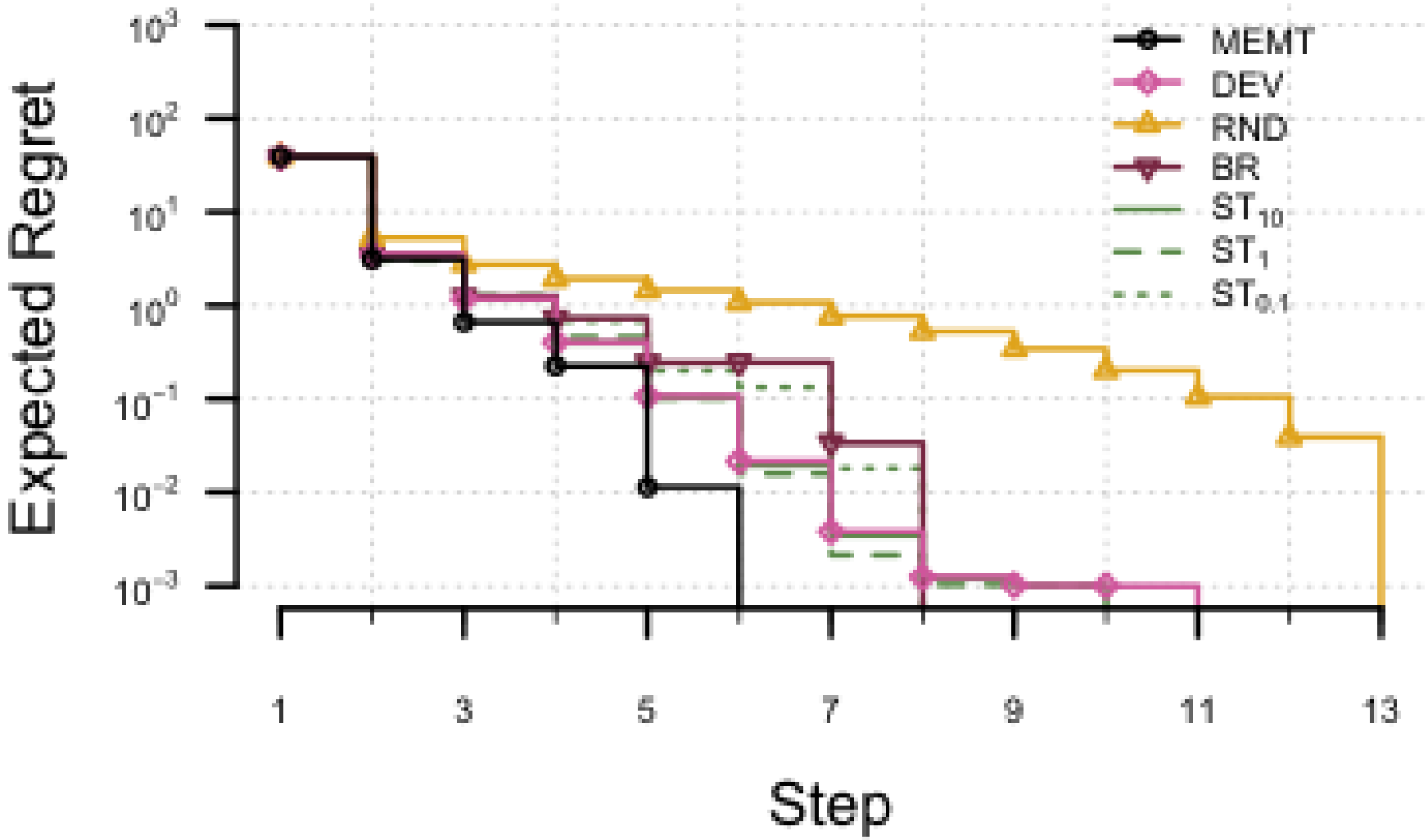
FPSB2 Regret Surface



Exploration Policies

- RND: Random (uniform) selection
- Deviation-Based
  - DEV: Uniform among strategies that deviate from current equilibrium
  - BR: Best response to current equilibrium
  - BR+DEV: Alternate on successive iterations
  - ST(τ): Softmax selection among deviators, proportional to gain
- MEMT:
  - Select strategy that maximizes the gain (regret) from deviating to a strategy outside the set from ***any*** mixture over the set.

CDA↓4


Expected Regret
10^{3}
10^{2}
10^{1}
10^{0}
10^{-1}
10^{-2}
10^{-3}
1
3
5
7
9
11
13
Step
MEMT
DEV
RND
BR
ST_{10}
ST_{1}
$ST_{0.1}$

EGTA Applications

- Market games
  - TAC: Travel, Supply Chain, Ad Auction
  - Canonical auctions: SAAs, CDAs, SSPSBs,…
  - Equity premium in financial trading
- Other domains
  - Privacy: information sharing attacks
  - Networking: routing, wireless AP selection
  - Credit network formation
- Mechanism design

Conclusion: EGTA Methodology

- Extends scope of GT to procedurally defined scenarios
- Embraces statistical underpinnings of strategic reasoning
- Search process:
  - GT for establishing salient strategic context
  - Strategy exploration:
    - e.g., RL to search for best response to that context

  $\rightarrow$Principled approach to evaluate complex strategy spaces
- Growing toolbox of EGTA techniques