Imitative Follower Deception in Stackelberg Games

Jiarui Gan, Haifeng Xu, Qingyu Guo, Long Tran-Thanh, Zinovi Rabinovich, Michael Wooldridge

Oxford, Harvard, NTU, Southampton

Background: Stackelberg Games & Learning

- A leader (L) vs. a follower (F)
- Stackelberg equilibrium $\langle x^*, y^* \rangle$ --- the optimal leader commitment:
  - $\langle \mathrm{x}^*, \mathrm{y}^* \rangle = \mathrm{argmax}_{\mathrm{x},\, \mathrm{y} \in BestResp(\mathrm{x})}\ U_L(\mathrm{x}, \mathrm{y})$
  - $BestResp(\mathrm{x}) \coloneqq \underset{\mathrm{y}}{\mathrm{argmax}}\ U_F(\mathrm{x}, \mathrm{y})$

👍 Efficient computation of optimal leader commitment

👍 Applications: security, exam design, contract design, mechanism design

When Follower Type (Payoffs) is Uncertain…

👉 **Learn** the optimal commitment by observing **follower best responses** [Letchford et al., 2009; Blum et al., 2014; Haghtalab et al., 2016; Roth et al., 2016; Peng et al., 2019]

	1^F	2^F
1^L	-1	1/3
2^L	3	-1

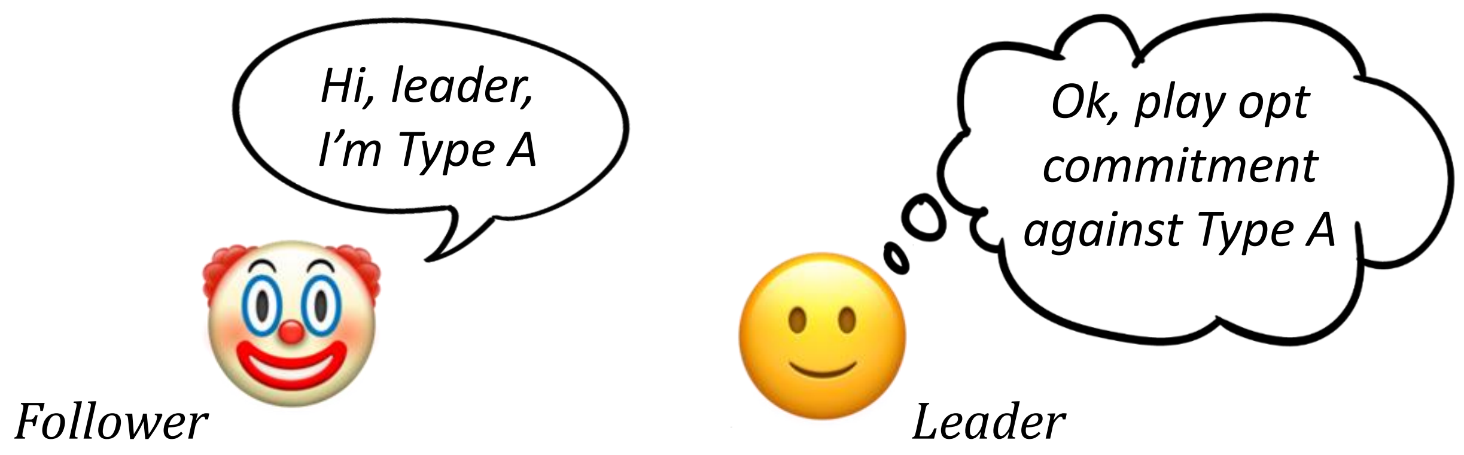
Type A

	1^F	2^F
1^L	-1	1
2^L	1	-1

Type B

Our Model: Play Against Follower Deception

- A naïve playbook when deception is ignored



Leader Policy: a Better Playbook

- A policy-based framework
  - **Stage 1**: **Leader** commits to a policy that specifies the strategy he will play for each reported (learned) follower type.

LEADER POLICY

Hi Follower, I'll play:

$\mathbf{x}_1$ *if you behave like type **A**,*

$\mathbf{x}_2$ *if you behave like type **B**,*

…

$\mathbf{x}_{26}$ *if you behave like type **Z**.*

Feel free to choose how to behave.

  - **Stage 2**: **Follower** optimally reports (imitates) a type ***T***, so that the strategy the leader will play according to her policy maximizes the follower's utility in Stage 3.
  - **Stage 3**: **Leader** plays a strategy **x** as prescribed by her policy and **Follower** best responds to **x** as if he is of type ***T***.

- **Example:**
  - Play $\left(\frac{3}{4} - \epsilon,\ \frac{1}{4} + \epsilon\right)$ if Follower behaves like *Type A*.
  - Play $\left(\frac{1}{2} + \boldsymbol{\epsilon},\ \frac{1}{2} - \boldsymbol{\epsilon}\right)$ if Follower behaves like *Type B*.

👉 A *Type-A* follower now has **no** incentive to misreport *Type B* !!

Imitative Follower Deception: an Example

	1^F	2^F
1^L	*1, -1*	*-1, 1/3*
2^L	*-1, 3*	*0.99, -1*

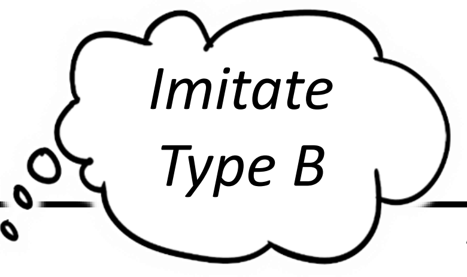
Type A

	1^F	2^F
1^L	*1, -1*	*-1, 1*
2^L	*-1, 1*	*0.99, -1*

Type B

A *defender* (the leader, row player) wants to defend two areas 1 and 2, which a *poacher* (the follower, column player) wants to attack. The poacher may be of Types A or B as his payoffs depend on animal prices on the black market, which fluctuate and are held private by the poacher.

- When the follower is truthful



	Type A	*Type B*
Optimal commitment:	$\left(\frac{3}{4} - \epsilon,\ \frac{1}{4} + \epsilon\right)$	$\left(\frac{1}{2} - \epsilon,\ \frac{1}{2} + \epsilon\right)$
Follower response:	1^F	1^F
Leader utility:	1/2	0
Follower utility:	0	0

- But when the follower is **untruthful**…

👉 A *Type-A* follower has an incentive to imitate *Type B*, which makes the leader play $(1/2 - \epsilon,\ 1/2 + \epsilon)$!

👉 A Type-A follower gets ≈ 1, but the leader only gets ≈ 0 🤔

Computing Optimal Policy: Algorithmic Results

- A complete view of the complexity: **OptPly** is hard to approximate, and hard still under **incentive compatibility (OptPly-IC)**

Theorem. For any $\epsilon > 0$, no poly-time $\frac{1}{(|\Theta|-1)^{1-\epsilon}}$-approximation for **OptPly** unless P=NP, even when the number of follower actions is fixed to 3.

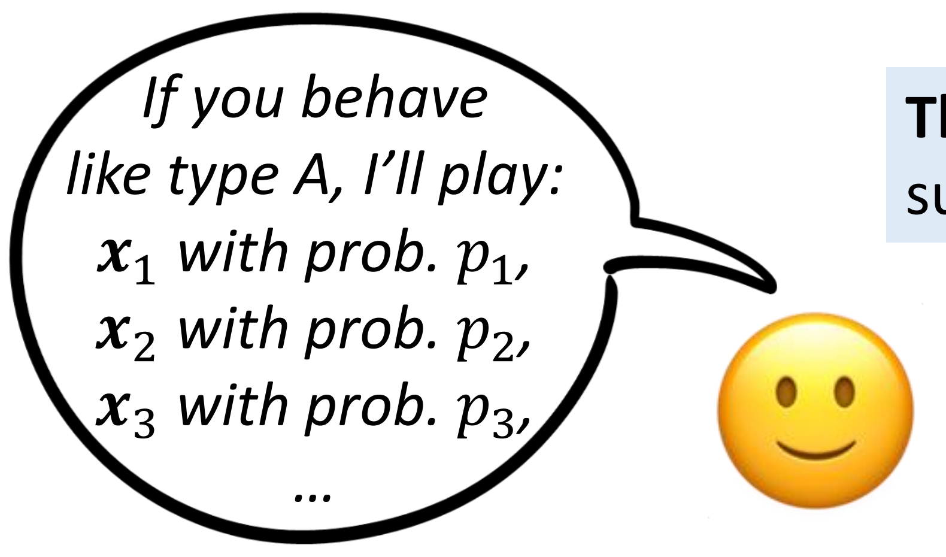
Theorem. For any $\epsilon > 0$, no poly-time $\frac{1}{|\Theta|^{1-\epsilon}}$-approximation for **OptPly-IC** unless P=NP, even when the number of follower actions is fixed to 3.

Theorem. There exists a poly-time $\frac{1}{|\Theta|}$-approximation algorithm for both w/o IC.

Theorem. Both **OptPly** and **OptPly-IC** are tractable for a fixed $|\Theta|$.

Generalization to Mixed Policies

- A higher level of randomization, able to improve leader utility further

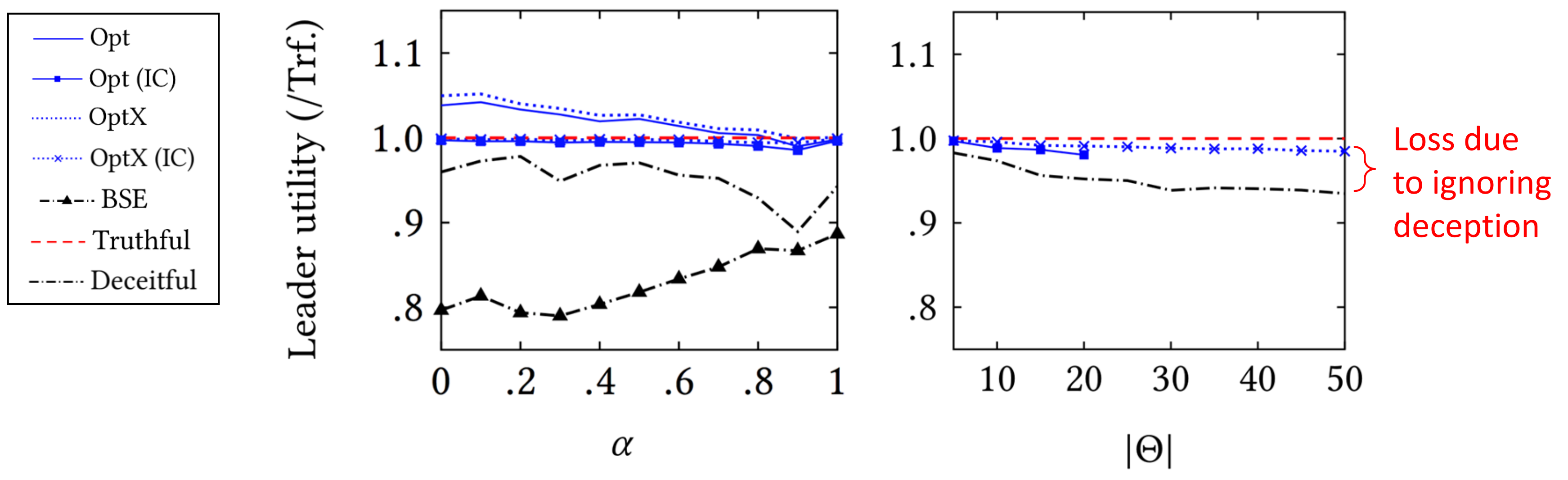


Theorem. Mixed policies with support size m suffice for achieving the optimality.

Theorem. With mixed policy, **OptPoly** remains hard to approximate, but **OptPoly-IC** becomes tractable.

Experiments

- Comparison of leader utility obtained with different approaches


Opt
Opt (IC)
OptX
OptX (IC)
BSE
Truthful
Deceitful
Leader utility (/Trf.)
1.1
1.0
.9
.8
0
.2
.4
.6
.8
1
α
10
20
30
40
50
$|\Theta|$
Loss due to ignoring deception