~~Persuading a Learning Agent~~
Generalized Principal-Agent Problem with a Learning Agent

Tao Lin, Yiling Chen
Harvard University

Introduction

Many economic problems have a *principal-agent* structure, where a principal **commits** to a strategy first, then an agent **best responds**:

- *Contract Design*
- *Bimatrix Stackelberg Game*
- *Information Design (Bayesian Persuasion)*
- …

However:

- Oftentimes, the principal cannot commit,
- and the agent does not best respond.
- Nowadays, we have machine learning agents.

This work studies ***principal-agent problems with a learning agent***: Can the principal do better than the classical problem where the agent best responds?

(Classical) Generalized Principal-Agent Problem

Proposed by Myerson (1982) & Gan-Han-Wu-Xu (2024):

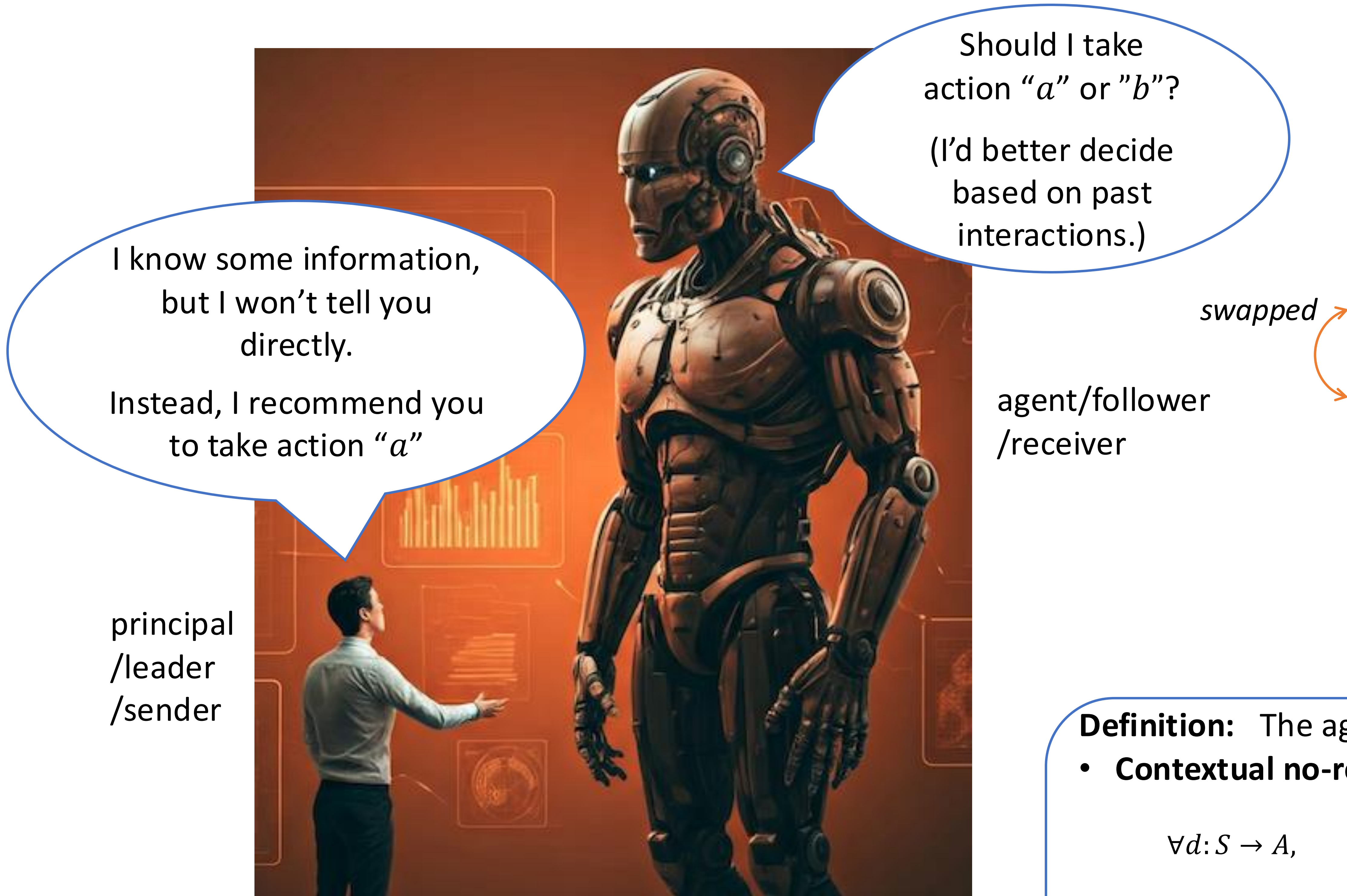
- The principal commits to a strategy $\pi = (q_s, x_s)_{s \in S}$:
  - S is a finite set of signals/recommendations.
  - $(q_s)_{s \in S}$ is a distribution over S: $\sum_{s \in S} q_s = 1$
  - $x_s \in \mathcal{X}$ is a decision associated with signal s.
- The agent chooses a strategy $\rho: \mathcal{X} \to A$
  - A is a finite set of actions of the agent.
  - **Best response:**

$$\rho(x_s) \in \mathrm{argmax}_{a \in A}\, v(x_s, a)$$

- Principal and agent obtain (expected) utility

$$\mathbb{E}_{s \sim q}\left[u(x_s, \rho(x_s))\right], \qquad \mathbb{E}_{s \sim q}\left[v(x_s, \rho(x_s))\right]$$

- $u(x, a), v(x, a)$ are assumed to be linear in $x \in \mathcal{X}$



Examples

- In *Contract Design*,
  - Action a leads to one of n outcomes.
  - $x_s = (p_1, \dots, p_n)$ is a payment vector (contract). $\mathcal{X} = \mathbb{R}_+^n$
- In *Bimatrix Stackelberg Game*,
  - $x_s \in \mathcal{X} = \Delta(\text{rows})$ is the leader's mixed strategy. Follower chooses a column a.
- In *Information Design*,
  - There is an unknown state of the world $\omega \sim$ prior μ
  - $x_s \in \mathcal{X} = \Delta(\Omega)$ is the posterior distribution of ω induced by signal s
  - Constraint: $\sum_{s \in S} q_s x_s = \mu$

Our Problem: Learning Agent

Instead of best-responding, we consider an agent who *learns* which action to take for each signal.

T rounds of interactions. In each round t,

- Based on history, the agent chooses a (randomized) strategy $\rho^t: S \to \Delta(A)$
- The principal chooses a strategy $\pi^t = (q_s^t, x_s^t)_{s \in S}$
- A signal $s^t \sim q^t$ is sampled, then:
  - the principal makes decision $x^t = x_{s^t}^t$
  - the agent samples action $a^t \sim \rho^t(s^t)$
- The two players' total expected utility: $\mathbb{E}\left[\sum_{t=1}^T u(x^t, a^t)\right]$ and $\mathbb{E}\left[\sum_{t=1}^T v(x^t, a^t)\right]$

Definition: The agent's learning algorithm satisfies

- **Contextual no-regret** if

$$\forall d: S \to A, \quad \mathbb{E}\left[\sum_{t=1}^{T}\left(v(x^t, d(s^t)) - v(x^t, a^t)\right)\right] \le \mathrm{CReg}(T) = o(T).$$

- **Contextual no-swap-regret** if

$$\forall d: S \times A \to A, \quad \mathbb{E}\left[\sum_{t=1}^{T}\left(v(x^t, d(s^t, a^t)) - v(x^t, a^t)\right)\right] \le \mathrm{CSReg}(T) = o(T).$$

Main Results: Under some regularity conditions (e.g., *agent has no dominated actions*),

- Against a **contextual no-regret** learning agent, the principal can obtain average utility at least $U^* - \Theta\left(\sqrt{\frac{\mathrm{CReg}(T)}{T}}\right)$; U^* is the principal's optimal utility against a best-responding agent.
- Against a **contextual no-swap-regret** learning agent, the principal ***cannot*** obtain more utility than $U^* + O\left(\frac{\mathrm{CSReg}(T)}{T}\right)$ (even if the principal can adapt to the agent's learning algorithm).
- For *some* **contextual no-regret** agent (MWU), the principal can obtain more than $U^* + \Omega(1)$.

Intuition: Consider the principal's signal s^t, together with the agent's algorithm's choice of action a^t, as a recommendation strategy $\tilde{\pi}$. No-swap-regret learning $\Rightarrow$ the agent (approximately) best responds to $\tilde{\pi}$. No-regret learning does not always have this property.