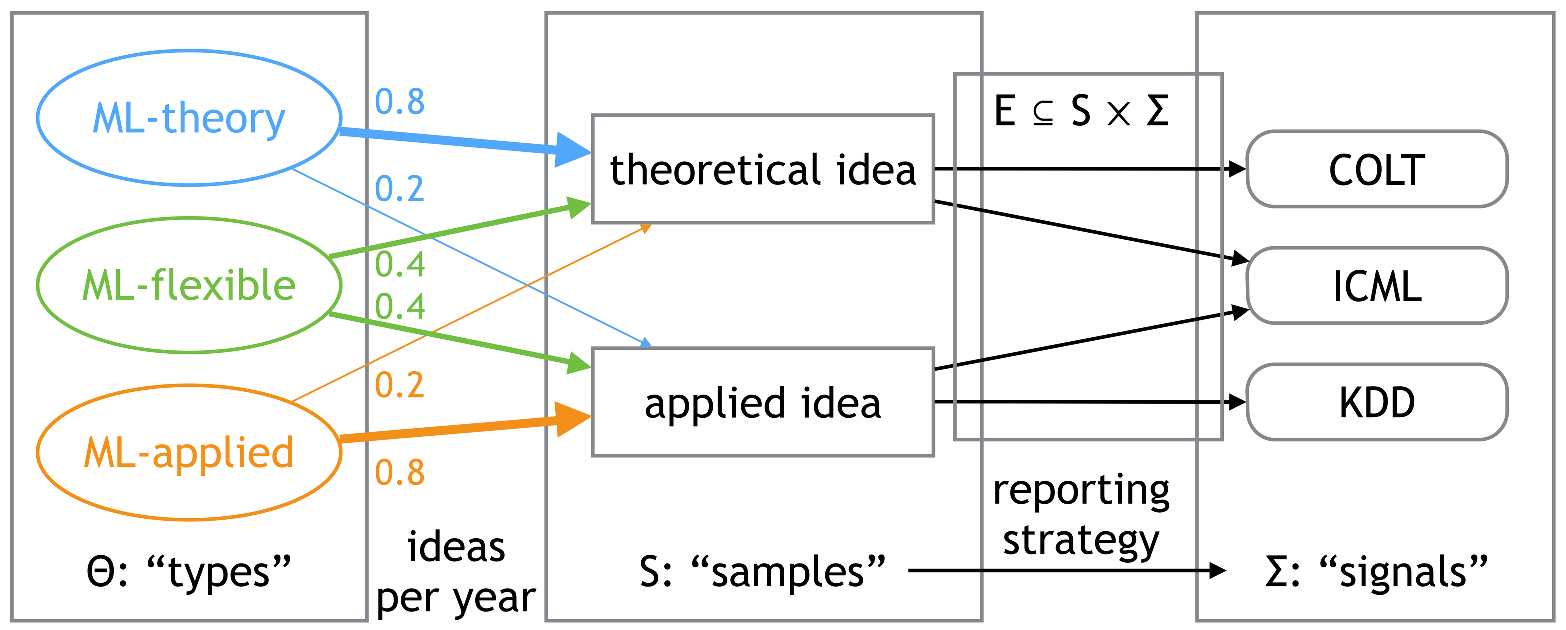
Distinguishing Distributions When Samples Are Strategically Transformed

Hanrui Zhang — Duke
Yu Cheng — UIC
Vincent Conitzer — Duke

Example: Academic Job Market



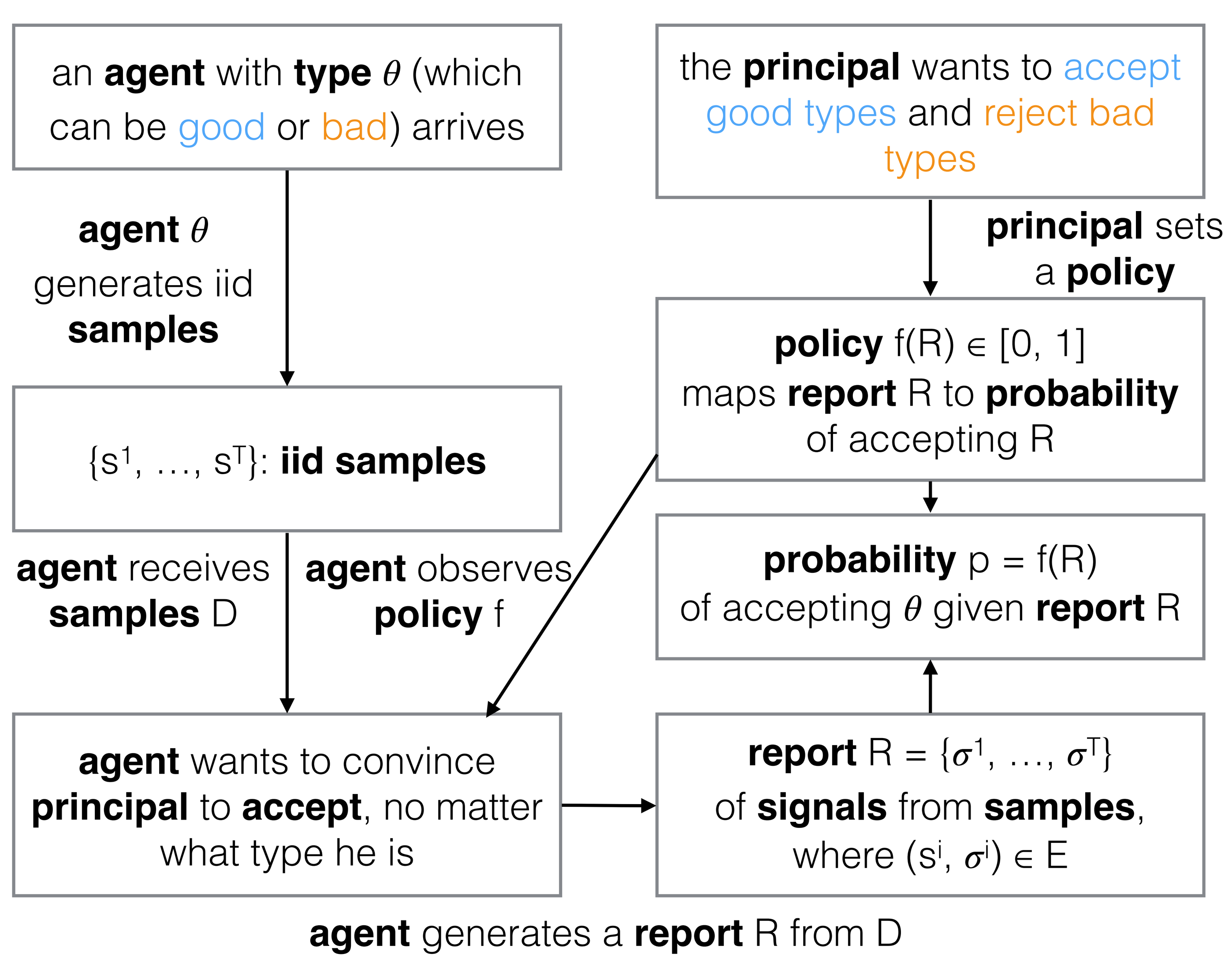
(High-quality) agent's problem:

- **"How do I distinguish myself from other types?"**
- **"How many ideas (years) do I need for that?"**

Principal's problem:

- **"How do I tell ML-flexible agents from others?"**
- **"After how many papers should I feel confident?"**

The Problem



Optimal Policy: the General Case

(Good) agent's perspective:

— "How do I distinguish myself from bad types?"
— Answer: report as far from bad types as possible

$d_{DTV}(x, y) = \max_{x \to \alpha} \min_{y \to \beta} d_{TV}(\alpha, \beta)$
where $x, y \in \Delta(S)$, $\alpha, \beta \in \Delta(\Sigma)$

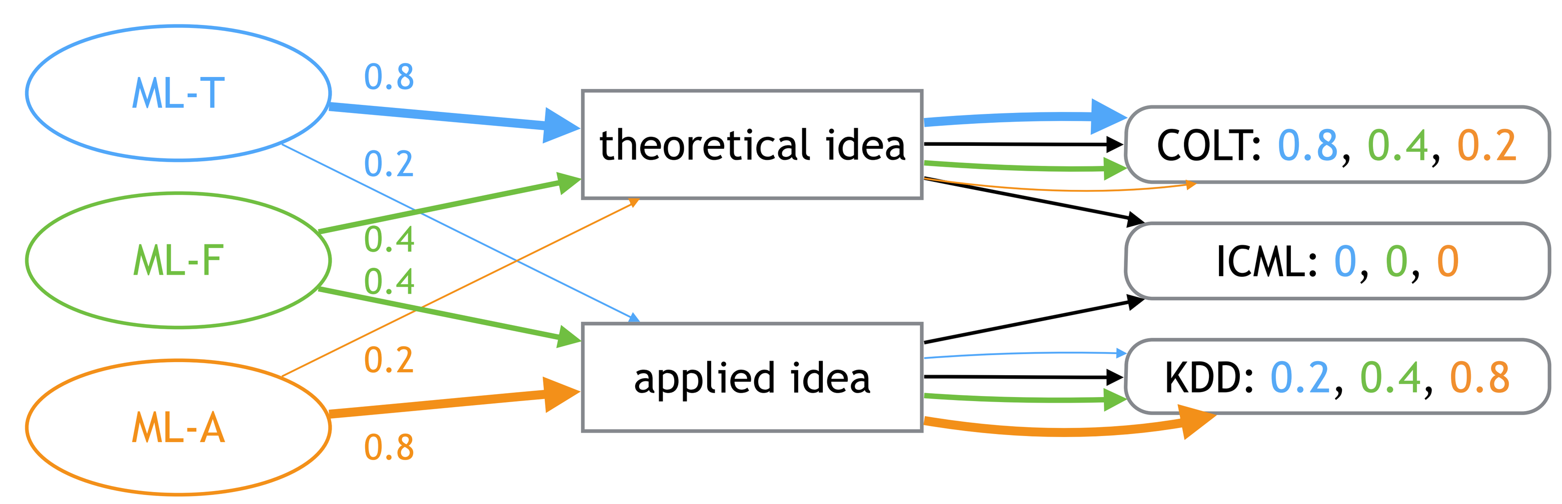
Principal's perspective:

— "How do I tell good agents from others?"
— Answer: ask for a set of signals that maximally separates good types from bad types

$\mathrm{MaxSep}(x, y) = \max_{A \subseteq \Sigma} (\max_{x \to \alpha} \alpha(A) - \max_{y \to \beta} \beta(A))$

Theorem.

- For any x and y, $d_{DTV}(x, y) = \mathrm{MaxSep}(x, y)$
- To distinguish x from y, we need $\Theta(d_{DTV}(x, y)^{-2})$ signals
- Optimal policy: ask for "enough" signals in separating set, which in general is NP-hard to find



- d_{DTV}(ML-F, ML-T) = 0.2, separating set: {KDD}
- d_{DTV}(ML-F, ML-A) = 0.2, separating set: {COLT}
- d_{DTV}(ML-T, ML-F) = 0.4, separating set: {COLT}
- d_{DTV}(ML-T, ML-A) = 0.6, separating set: {COLT}

Theorem.

To distinguish multiple good and bad types:

- Run a classifier for each good / bad pair
- Each good type needs to prove herself against all bad types
- Blow-up in sample complexity by a factor of (# bad types)

Take-home message:

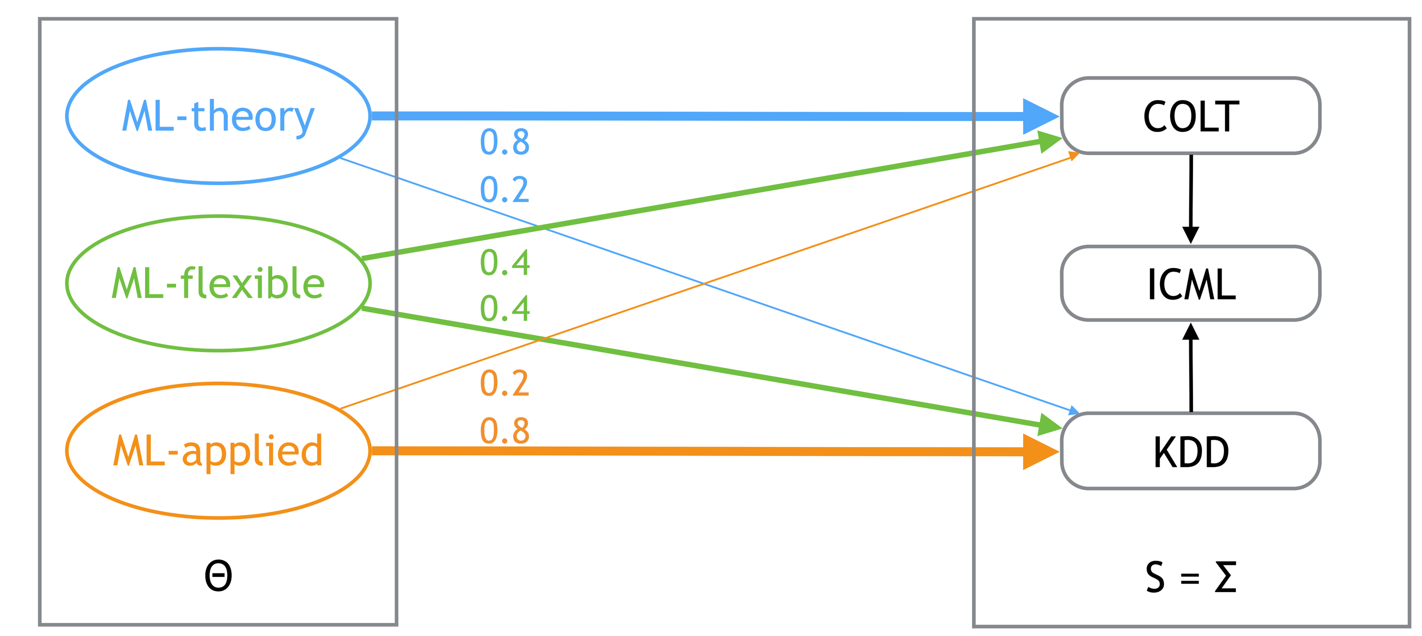
With strategic transformation, d_{DTV} = MaxSep plays the role of d_{TV} in classical tasks

When Signals Are Partially Ordered

Observation: no one ever wants to publish an ICML paper (which is by no means true in reality)

— But, why?

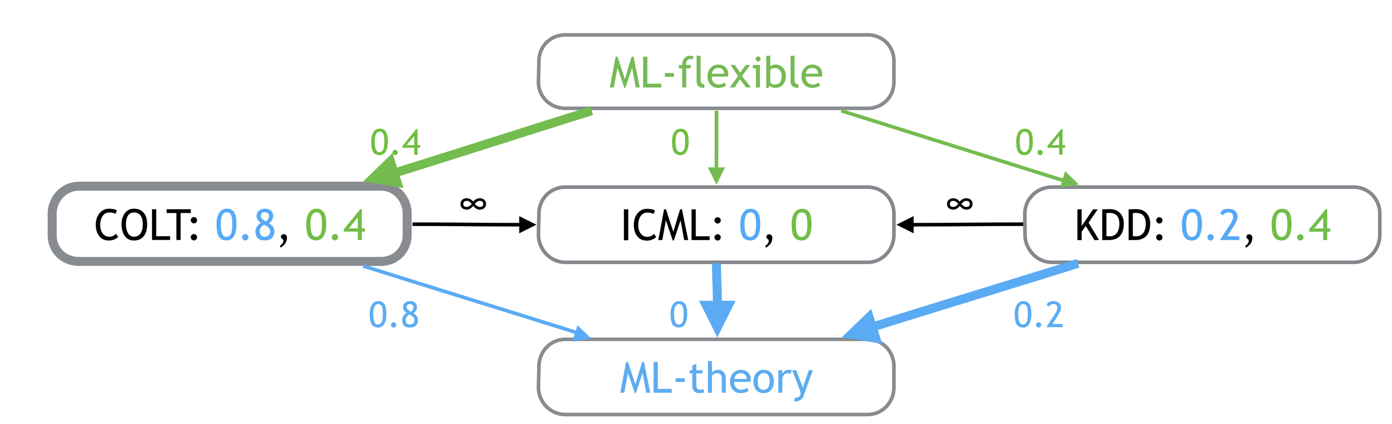
— Answer: because the signal space is partially ordered!



Theorem.

When signals are partially ordered:

- The revelation principle holds
- $d_{DTV}(x, y) = \min_{y \to z} d_{TV}(x, z) = 1 - \mathrm{MaxFlow}(y, x)$



Theorem.

To distinguish x from any y where $d_{DTV}(x, y) \ge \varepsilon$:

- Collect $T = \tilde{O}(\rho / \varepsilon^2)$ signals, where ρ is the width of Σ
- Accept iff $d_{DTV}(x, z) = 1 - \mathrm{MaxFlow}(z, x) < \varepsilon / 2$, where z = empirical distribution of signals
- Policy is independent of y

Take-home message:

Efficient learner & identity tester exist; width measures complexity of the space

See Also

When Samples Are Strategically Selected. Z-C-C. In ICML'19