NEURAL INFORMATION PROCESSING SYSTEMS

AETDICE: Unified Framework and Offline Optimization for Nonlinear Multi-Objective RL

Woosung Kim[1*], Youngjun Suh[1*], Jinho Lee[1*], Jongmin Lee[2†], Byung-Jun Lee[1,3†]

[1]Korea University [2]Yonsei University [3]Gauss Labs Inc.

*Equal contribution †Corresponding authors

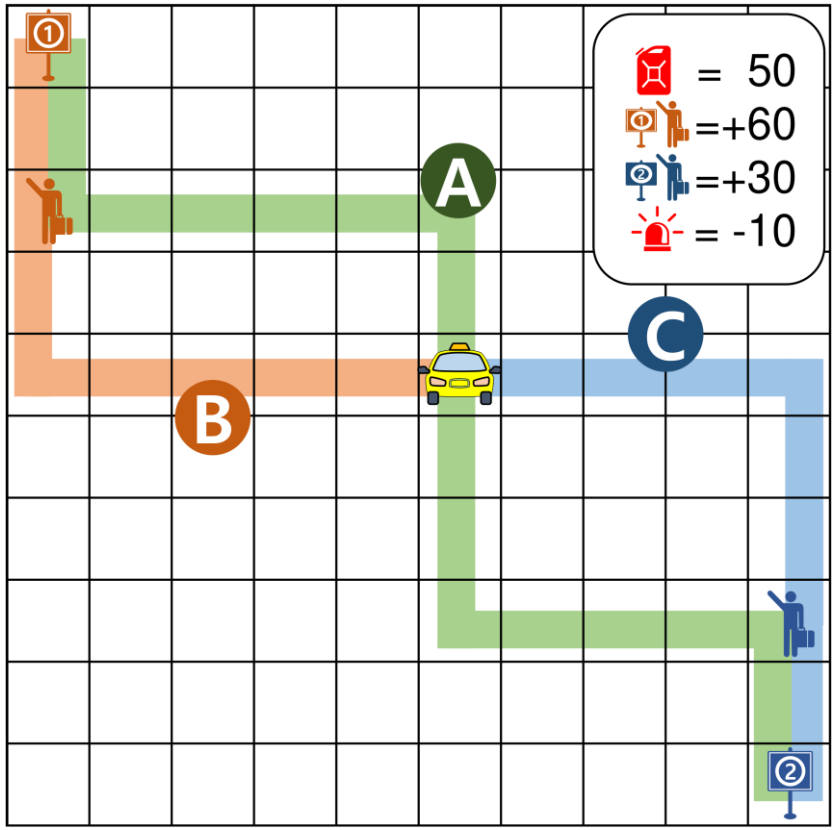






SER and ESR Lead to Different Optimal Policies

$$J_{\mathrm{SER}}(\pi; u) = u\big(\mathbb{E}_{\tau\sim\pi}[\mathbf{R}(\tau)]\big) \qquad J_{\mathrm{ESR}}(\pi; u) = \mathbb{E}_{\tau\sim\pi}\big[u(\mathbf{R}(\tau))\big]$$



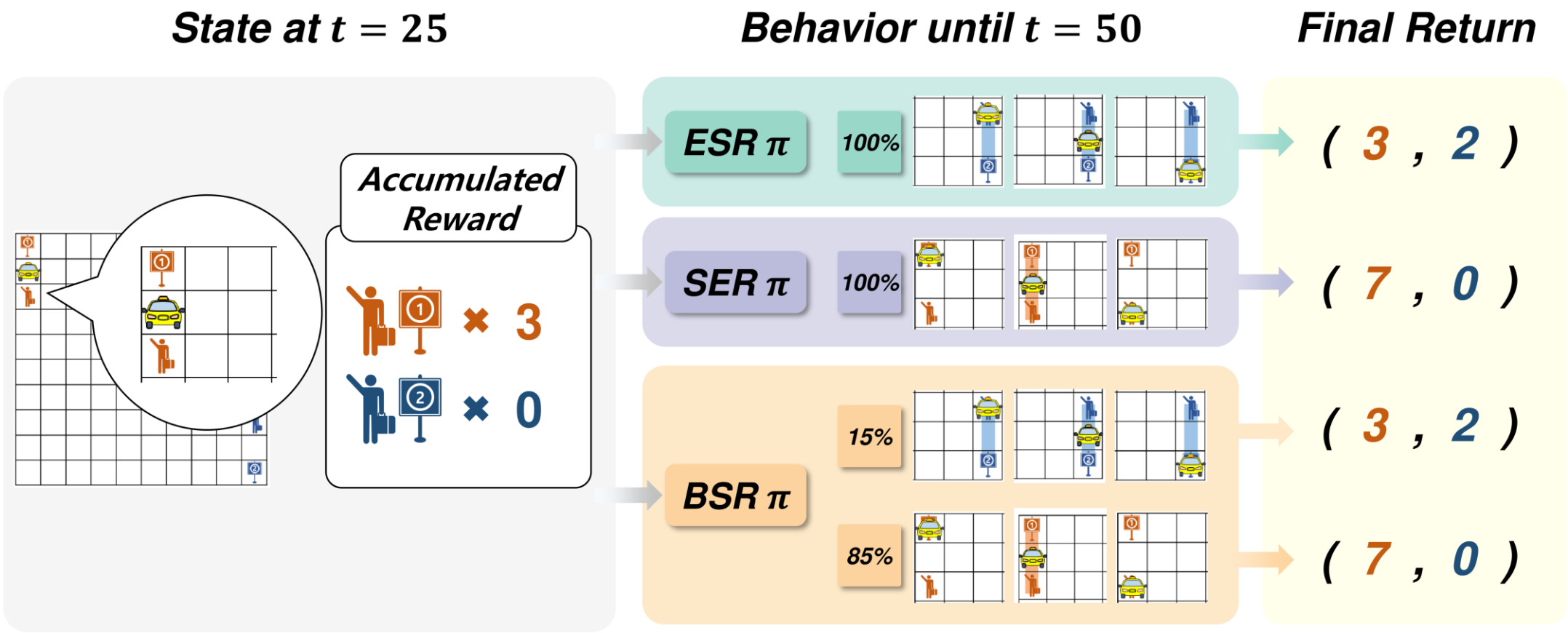
Fair-Taxi: two passenger groups, fairness utility u

- **A** serve both groups in one episode
- **B** serve group 1 only
- **C** serve group 2 only

ESR-optimal: always **A**, fair **in every episode**

SER-optimal: B or **C** at random, fair **only on average**

ESR Depends on Reward History



Same physical state, but group 1 was served three times already.

ESR uses that history; SER need not.

The Missing Cases in Offline MORL

Offline MORL so far

	Linear G	Concave G
Linear F	**Linear MORL** OptiDICE	**SER** FairDICE
Nonlinear F	**?**	**?**

No existing offline method covers ESR or general nonlinear F.

AET Unifies SER and ESR

$$J_{\mathrm{AET}}(\pi; F, G) = \underbrace{G}_{\text{Aggregation}}\left(\underbrace{\mathbb{E}_{\tau\sim\pi}}_{\text{Expectation}}\left[\underbrace{F}_{\text{Transformation}}\left(\mathbf{R}(\tau)\right)\right]\right)$$

F **= identity**
→ SER

G **= identity**
→ ESR

F can be non-convex—allowing flexible efficiency and safety objectives.

AETDICE: Reward Shaping + DICE, Fully Offline

① **Augment the state** with the accumulated reward $\mathbf{R}^{\text{acc}}$ (handles non-Markovian policies)

② **Shaped reward** = increment of ***F***

$$\tilde{\mathbf{r}}(\tilde{s}_t, a_t) = F\left(\mathbf{R}_t^{\text{acc}} + \mathbf{r}(s_t, a_t)\right) - F\left(\mathbf{R}_t^{\text{acc}}\right)$$

Summed over an episode: r̃ adds up to ***F***(**R**(τ))

③ **Optimize concave G**

Convex DICE program → AETDICE



Key: F only changes shaped rewards. Optimization stays **convex** even for non-convex **F**.

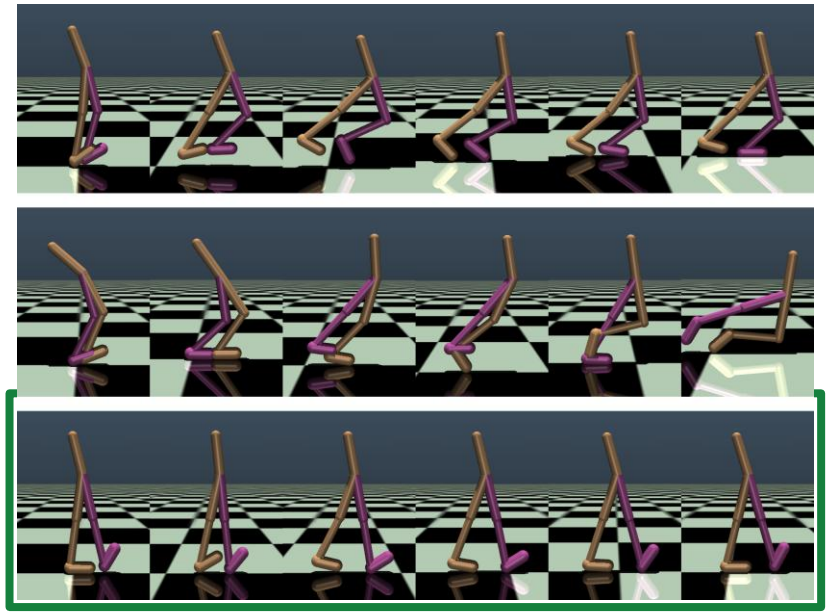
Non-convex *F*: Safety Utility Cuts Falls from 21% to 2%

Energy-first, ρ = 0.3	Efficiency ↑	Fall ↓	Safe ↑
BC	0.730	36.0%	26.3%
ESR-IQL	0.789	33.3%	22.3%
AET-E	0.868	21.3%	21.3%
AET-S	**0.948**	**1.7%**	**99.7%**

Shaded: ours. AET-E = AETDICE with efficiency ***F***. AET-S = AET-E + safety. 3 seeds × 100 ep.

Fall rate

21.3% → 1.7%



Successful walking
AET-E: efficient, but crouched

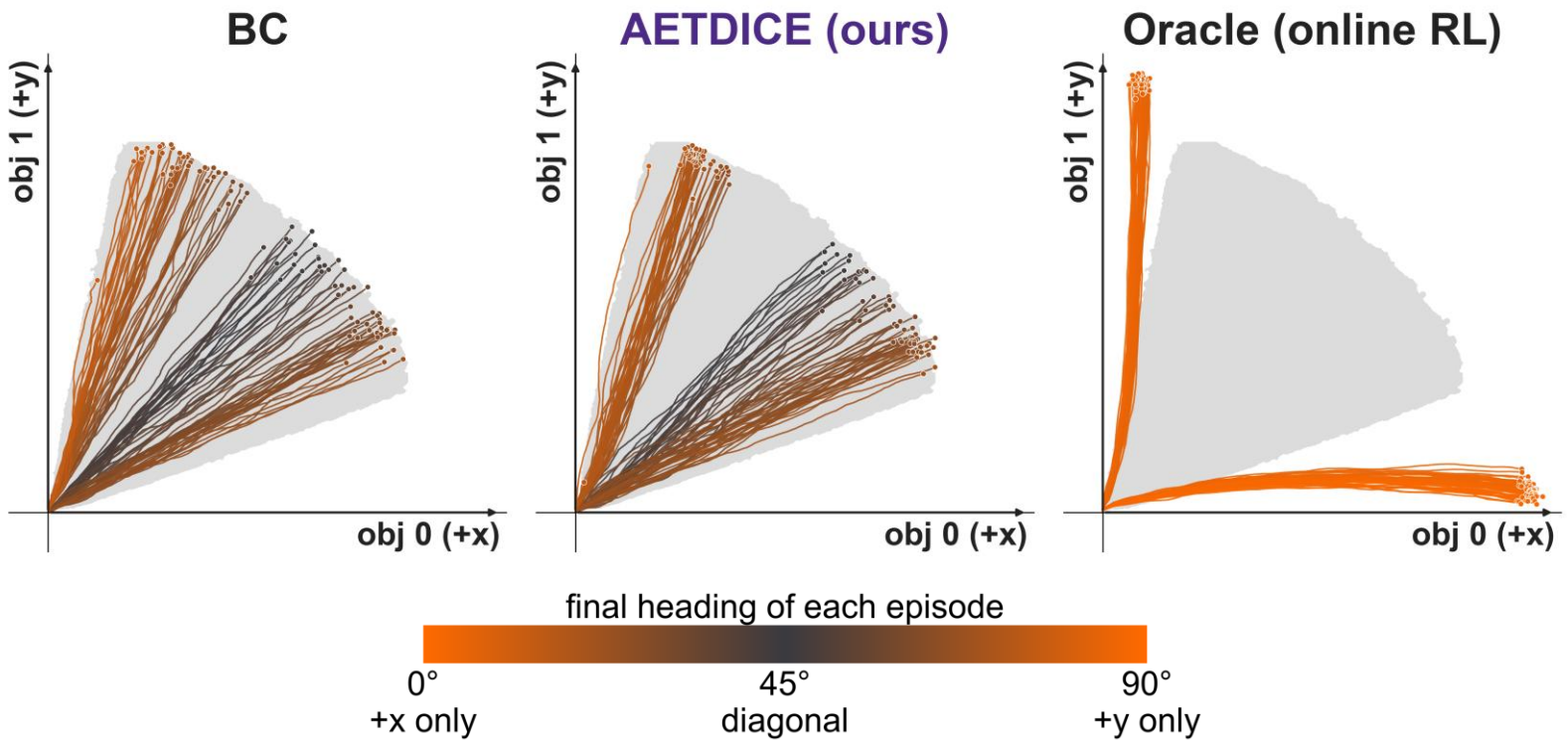
Failed walking
AET-E: falls in **21.3%** of episodes

Safe walking
AET-S: upright; falls **1.7%**, safe **99.7%**, higher efficiency

Add a safety term to F

Same algorithm. Safer walking.

Convex *F* + Concave *G*: Specialize, Yet Cover Both



AETDICE: two direction groups, closer to the oracle.

BC: diagonal behavior from the offline data (gray).

Convex **F**: one direction per episode

Concave **G**: both directions overall

MO-Ant	lin. G	conc. G
OptiDICE-[0,1]	1222.0	12.25
FairDICE (SER)	682.7	13.04
ESR-IQL	**1243.3**	12.23
AETDICE	949.7	**13.69**

The best policy depends on the objective.

Takeaway

- **AET** unifies SER and ESR.
- **AETDICE** learns them fully offline—even for non-convex F.
- **Safety:** falls **21.3% → 1.7%**.

Paper:
arxiv.org/abs/2606.31178

Come see our poster!