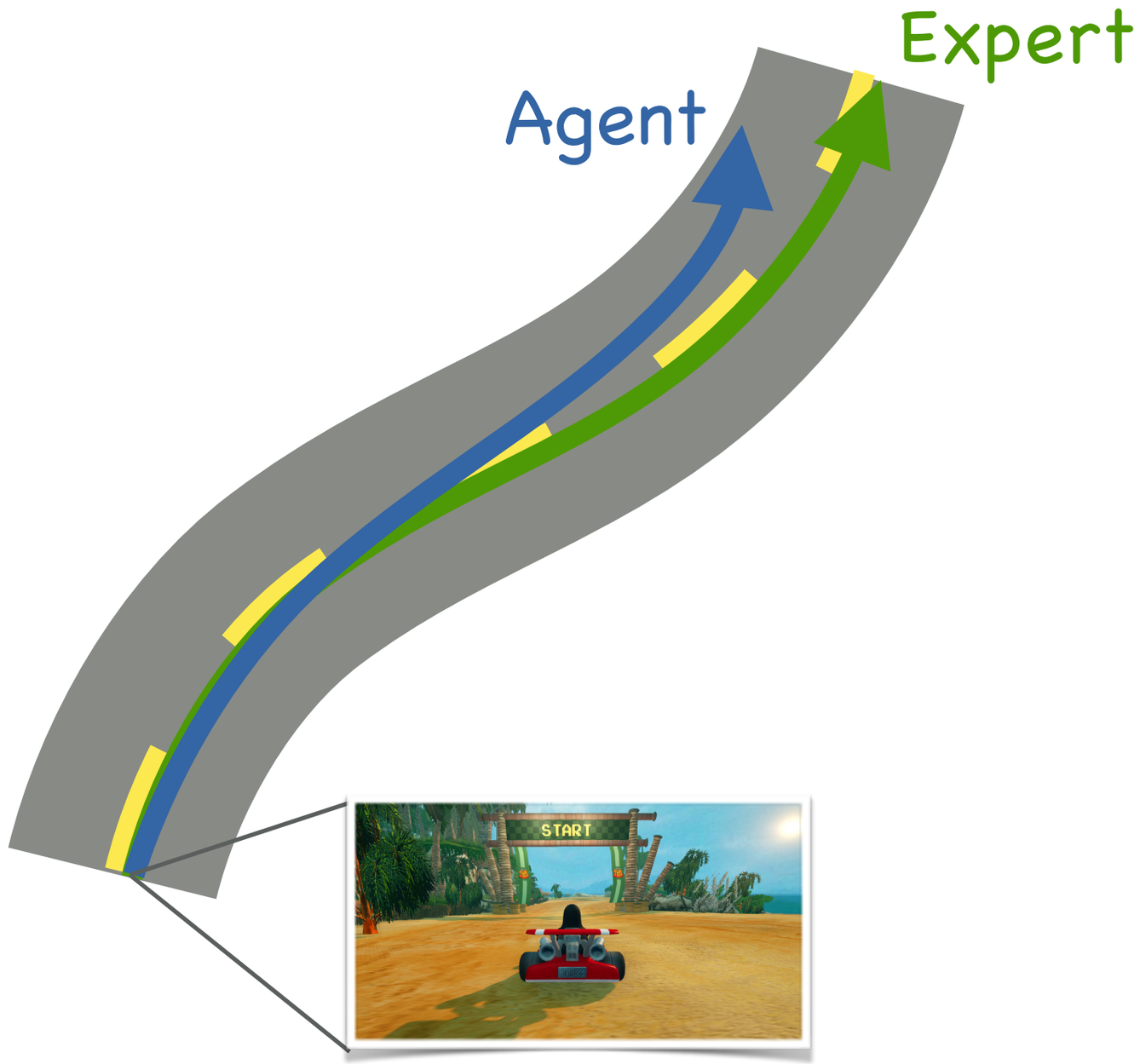
DAgger

Imitation learning

- Drift
  - Mismatch in training and testing distribution



Imitation learning - Alternative interpretation

- Expert policy π_E
- Agent policy π
- Iterate
  - Take action $a_t^E \sim \pi_E(\cdot \mid s_t)$
  - Imitate $\log \pi(a_t^E \mid s_t)$
  - State update $s_{t+1} \sim T(\cdot \mid a_t^E, s_t)$

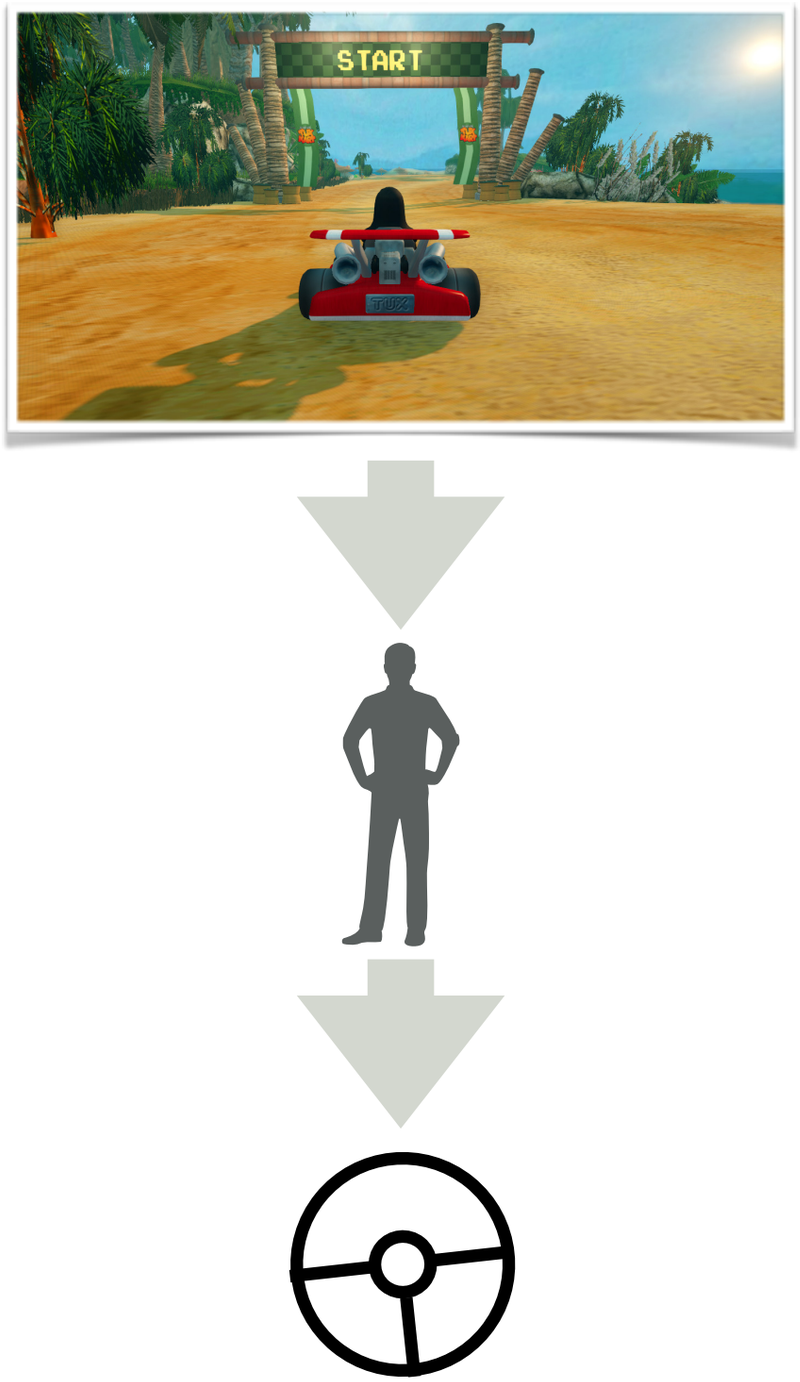
Dataset Aggregation

- Expert policy π_E
- Agent policy π
- Iterate
  - Take action $a_t^E \sim \pi_E(\cdot \mid s_t)$
  - Take action $a_t \sim \pi(\cdot \mid s_t)$
  - Imitate $\log \pi(a_t^E \mid s_t)$
  - State update $s_{t+1} \sim T(\cdot \mid a_t, s_t)$

A Reduction of Imitation Learning and Structured Prediction to No-Regret Online Learning, Ross et al., AISTATS 2011

DAgger - Issues

- Requires expert oracle
  - Very hard to humans



DAgger

- On-policy imitation learning
- Guaranteed to work for agents with enough capacity and good enough expert