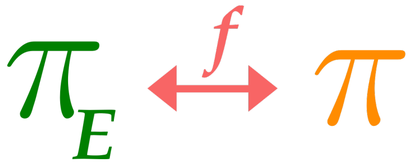
Causal Imitation Learning under Temporally Correlated Noise

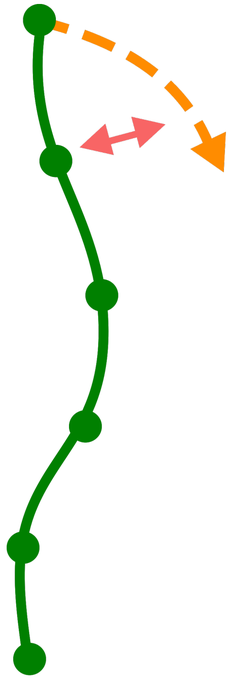
Gokul Swamy

(joint work with Sanjiban Choudhury, Drew Bagnell, Steven Wu)

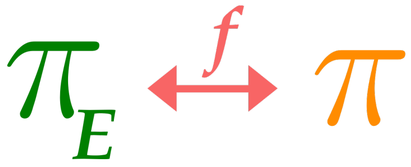

π_E
f
π

$\pi_E \overset{f}{\longleftrightarrow} \pi$

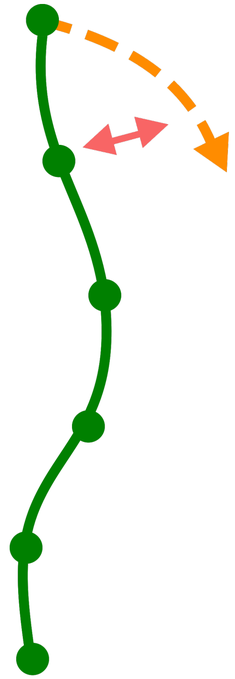
Offline



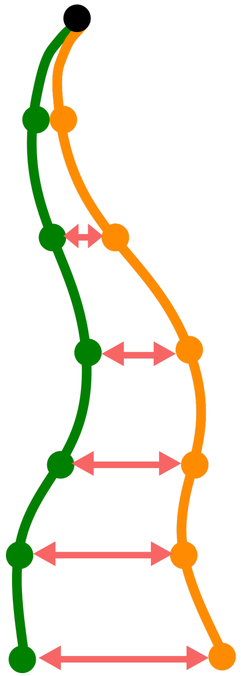
Behavioral Cloning …



Offline



Online

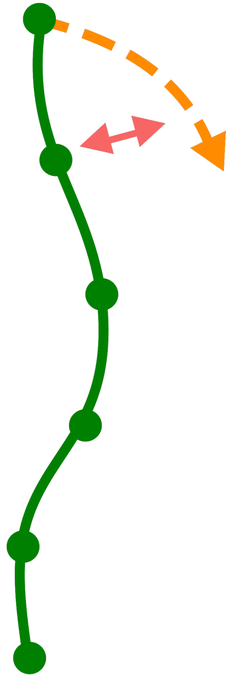


Behavioral Cloning …

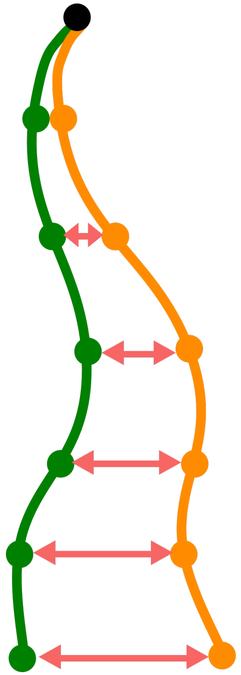
GAIL, MaxEnt IRL …

$\pi_E \overset{f}{\longleftrightarrow} \pi$

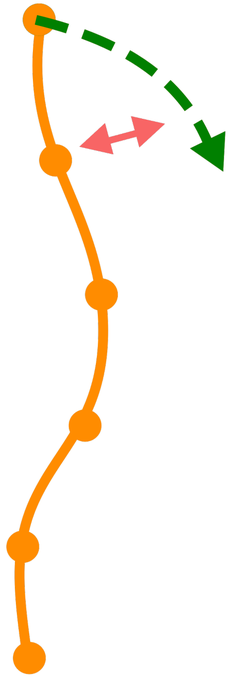
Offline



Online



Interactive



Behavioral Cloning …

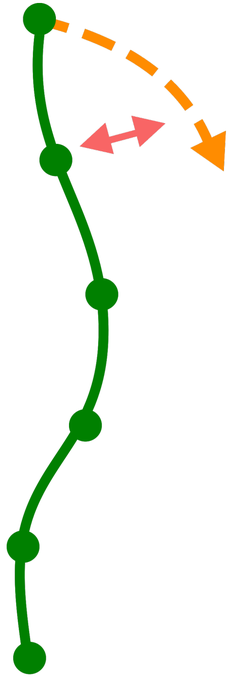
GAIL, MaxEnt IRL …

DAgger …

$\pi_E \overset{f}{\longleftrightarrow} \pi$

[Swamy et al, 2021]

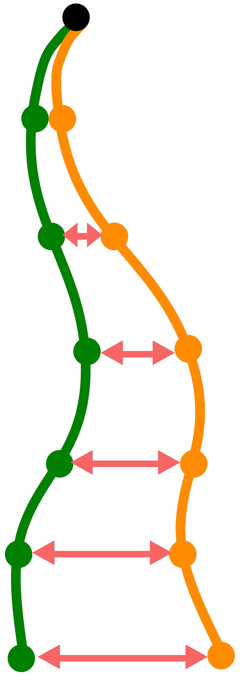
Offline



$J(\pi_E) - J(\pi) \leq O(\epsilon T^2)$

Behavioral Cloning …

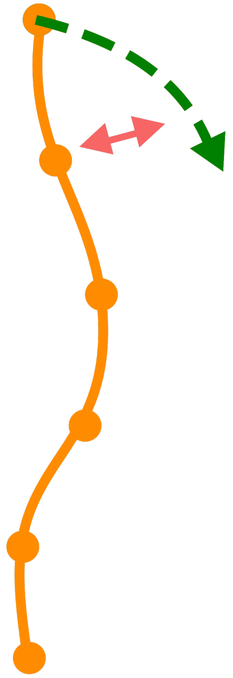
Online



$J(\pi_E) - J(\pi) \leq O(\epsilon T)$

GAIL, MaxEnt IRL …

Interactive



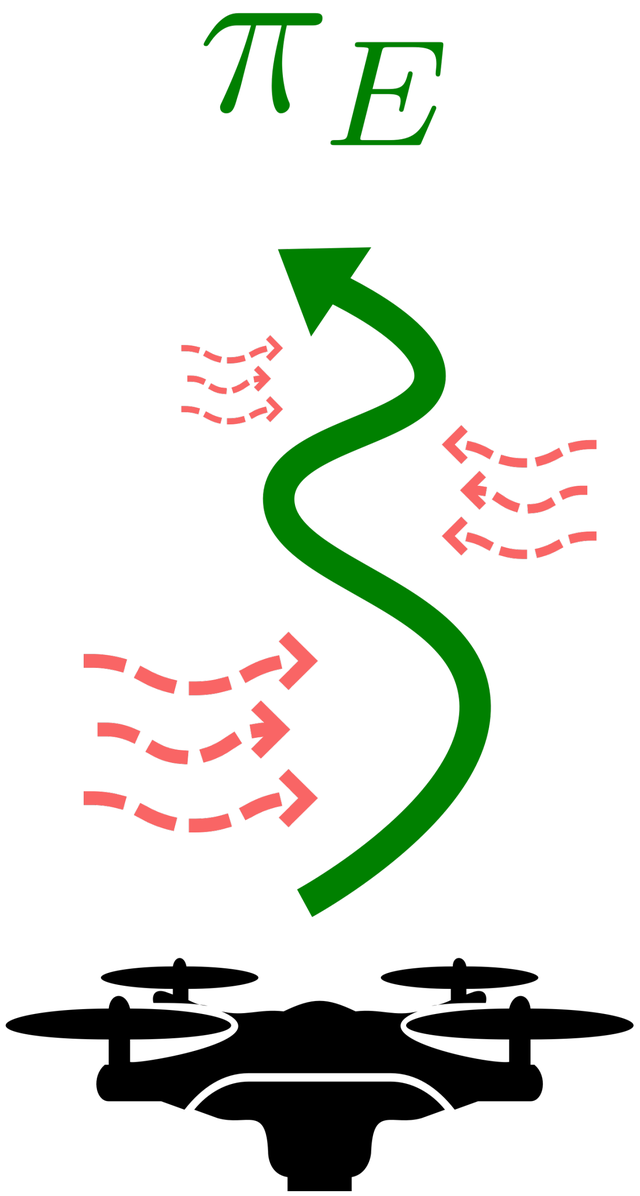
$J(\pi_E) - J(\pi) \leq O(\epsilon H T)$

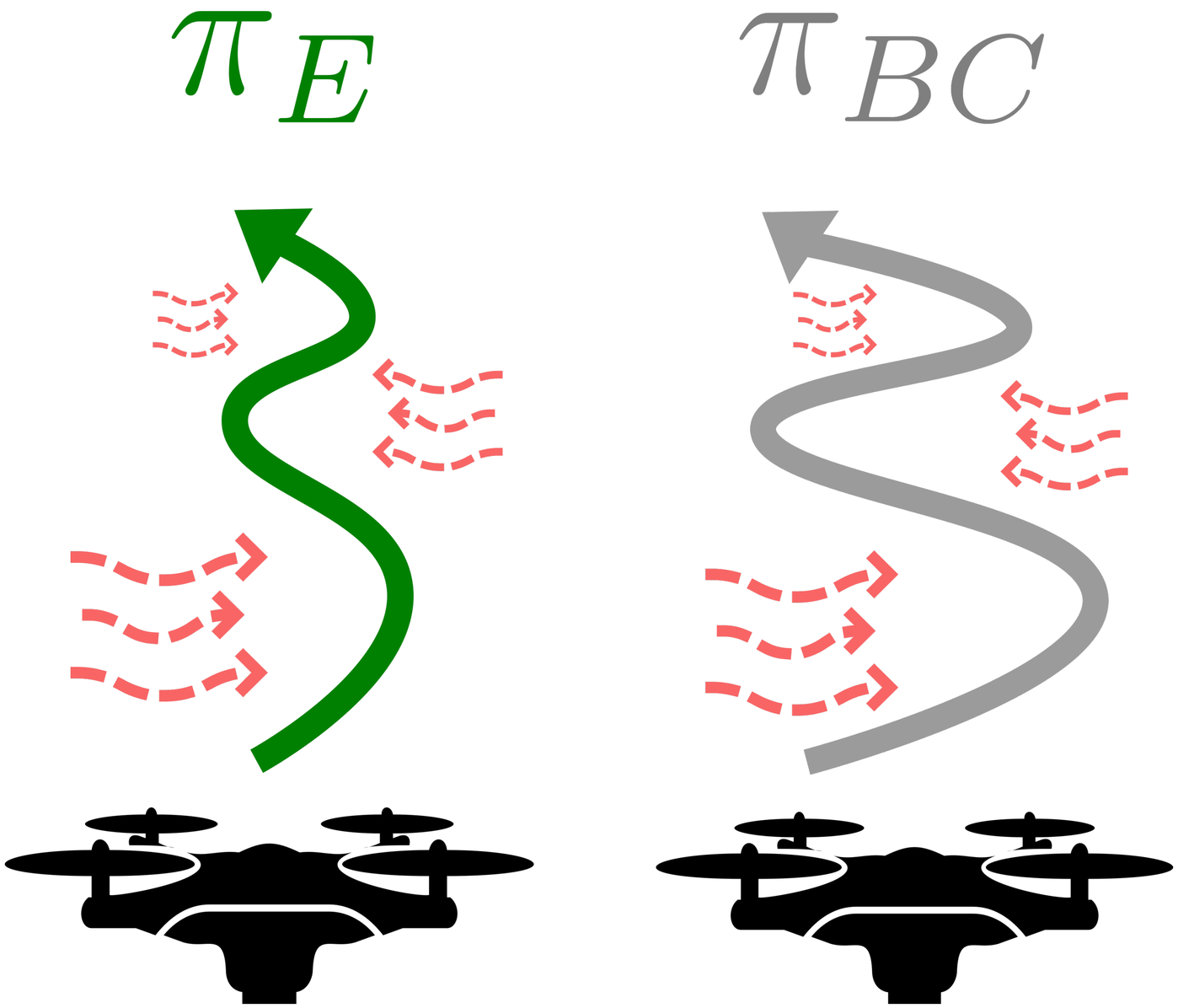
DAgger …

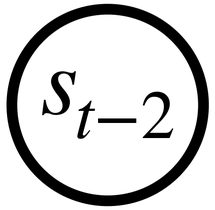
"Hence, a system trained with multiple frames would merely predict a steering angle equal to the current rate of turn as observed through the camera. This would lead to catastrophic behavior in test mode. The robot would simply turn in circles."
— Muller et al., 2006

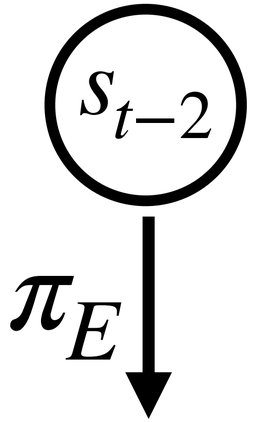
"Hence, a system trained with multiple frames would merely predict a steering angle equal to the current rate of turn as observed through the camera. This would lead to catastrophic behavior in test mode. The robot would simply turn in circles."
— Muller et al., 2006

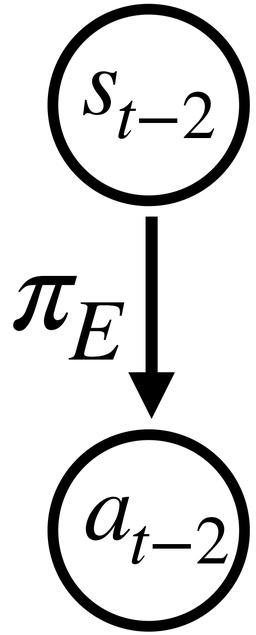
"Actually, since we were fitting a model to a time-series, samples tend to be correlated in time […] Thus, when leaving out a sample in cross validation, we actually left out a large window (16 seconds) of data around that sample, to diminish this bias."
— Ng et al., 2003

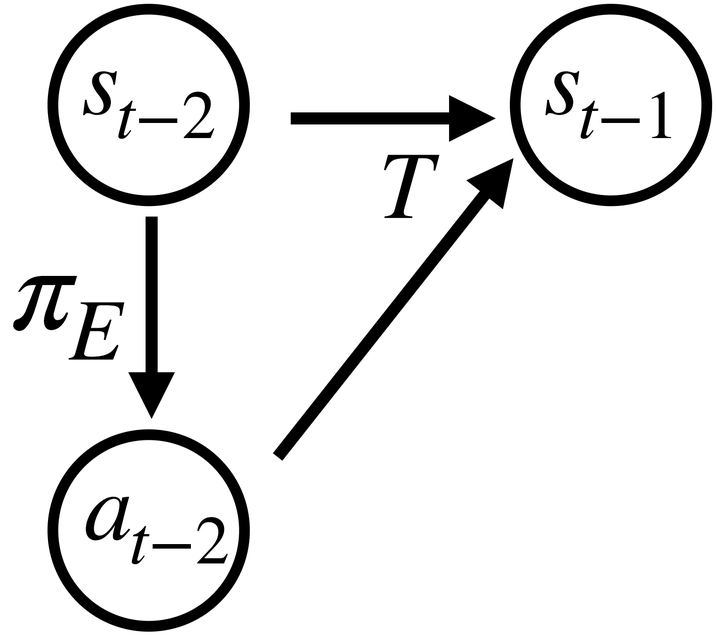

π_E

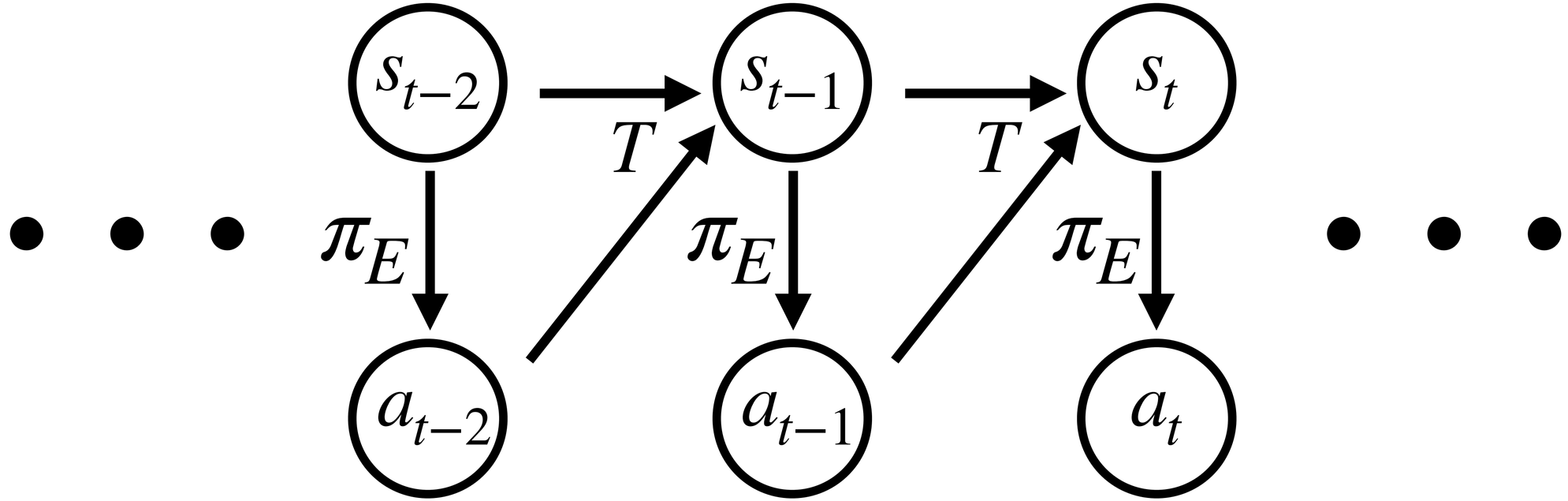

π_E
π_{BC}

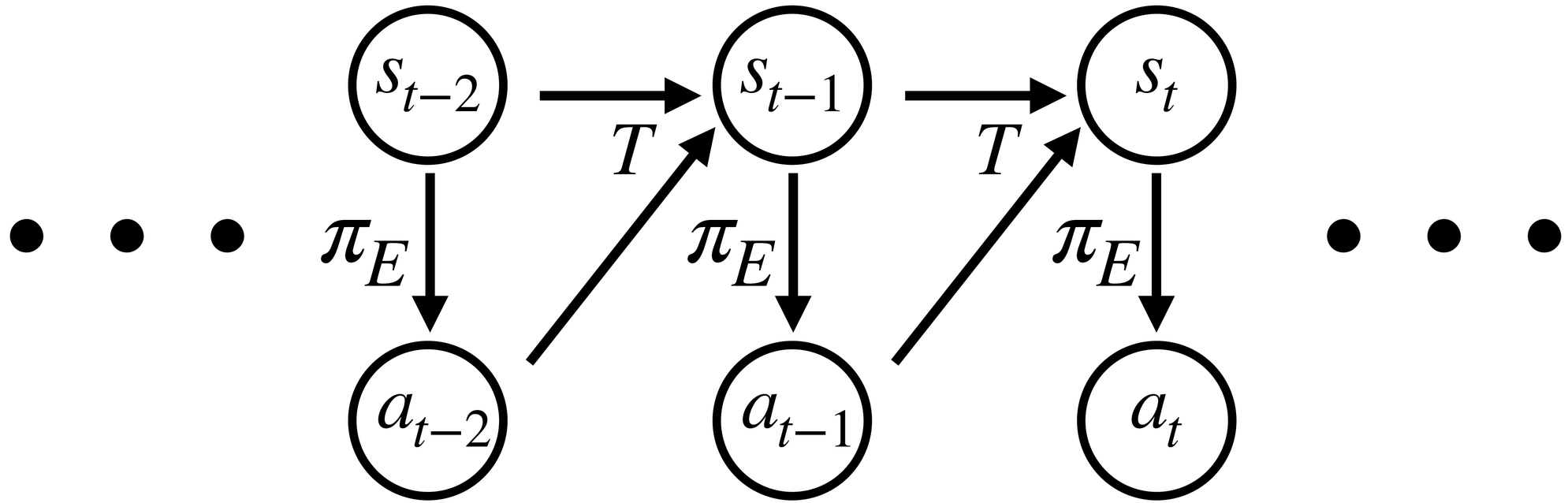

s_{t-2}


s_{t-2}
π_E

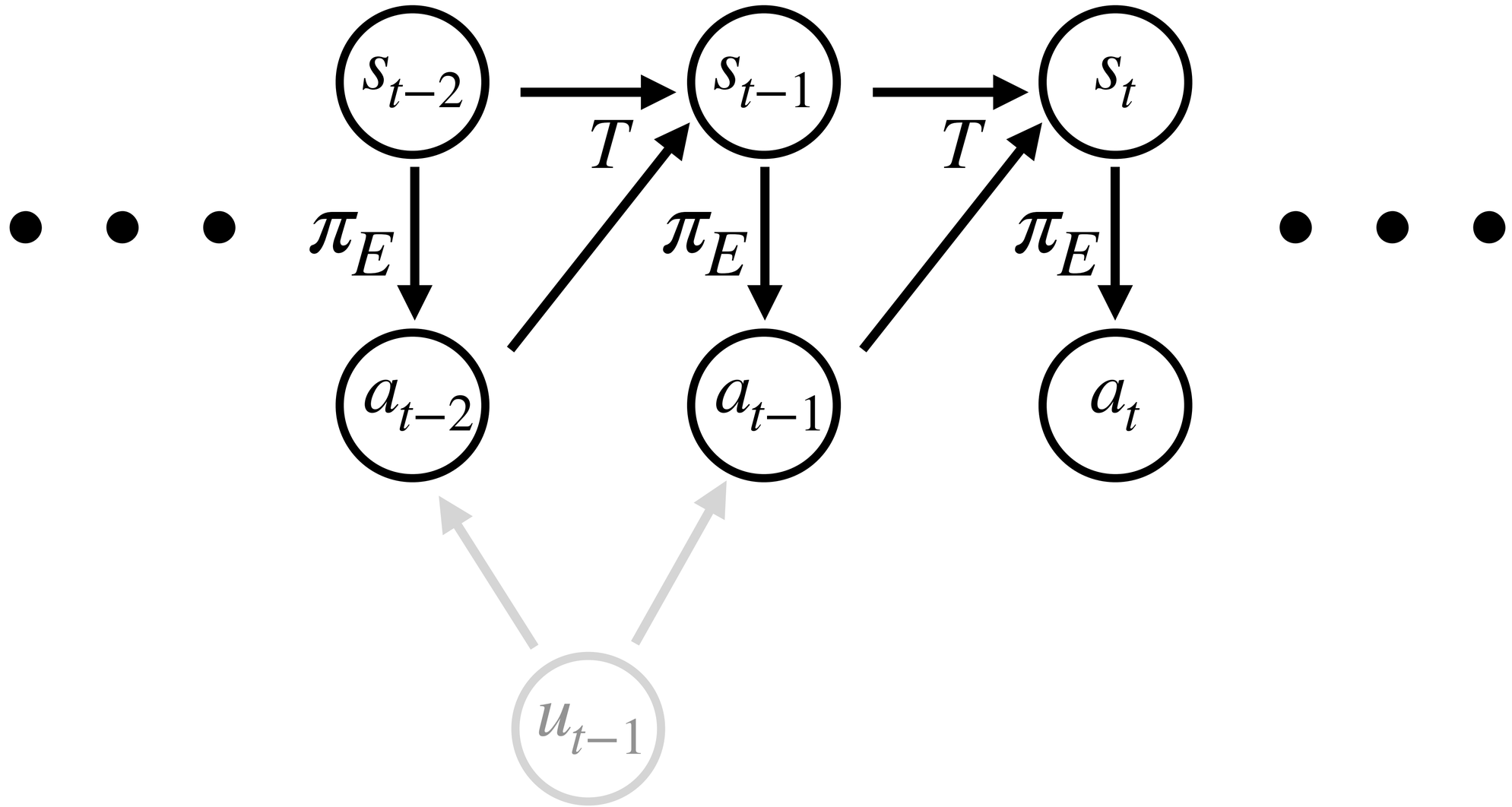

s_{t-2}
π_E
a_{t-2}

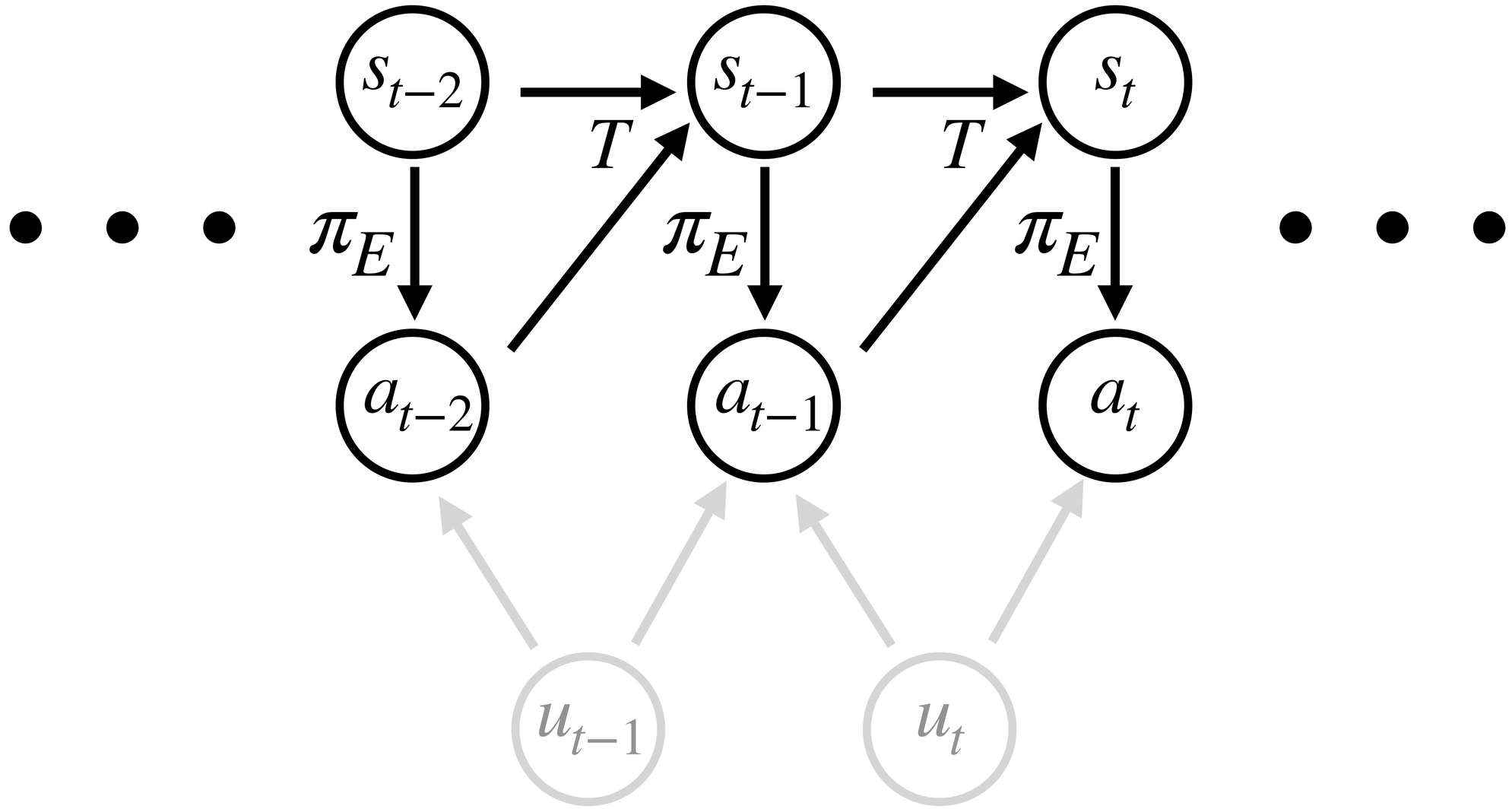

s_{t-2}
s_{t-1}
T
π_E
a_{t-2}

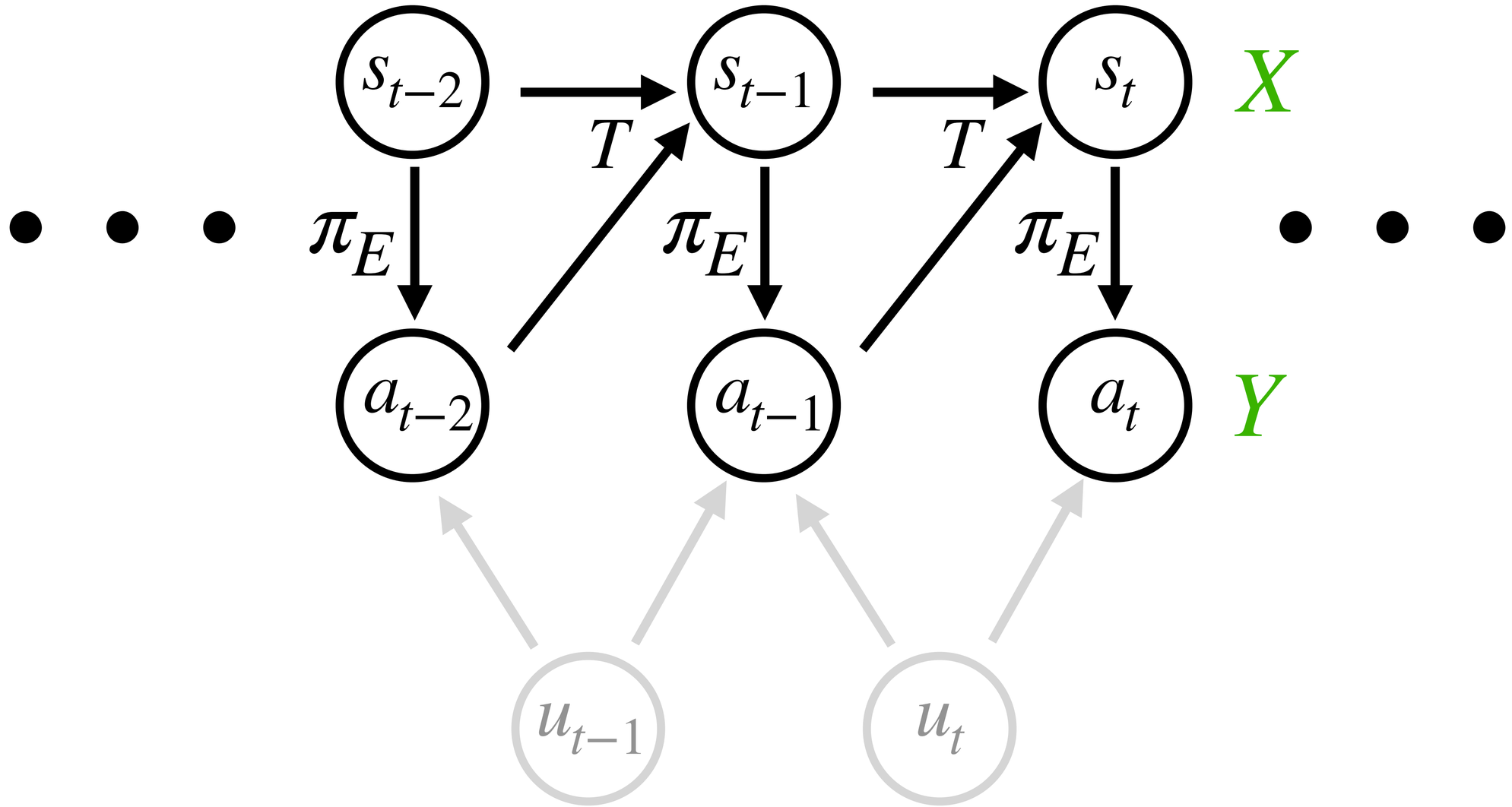

s_{t-2}
s_{t-1}
s_t
T
T
π_E
π_E
π_E
a_{t-2}
a_{t-1}
a_t

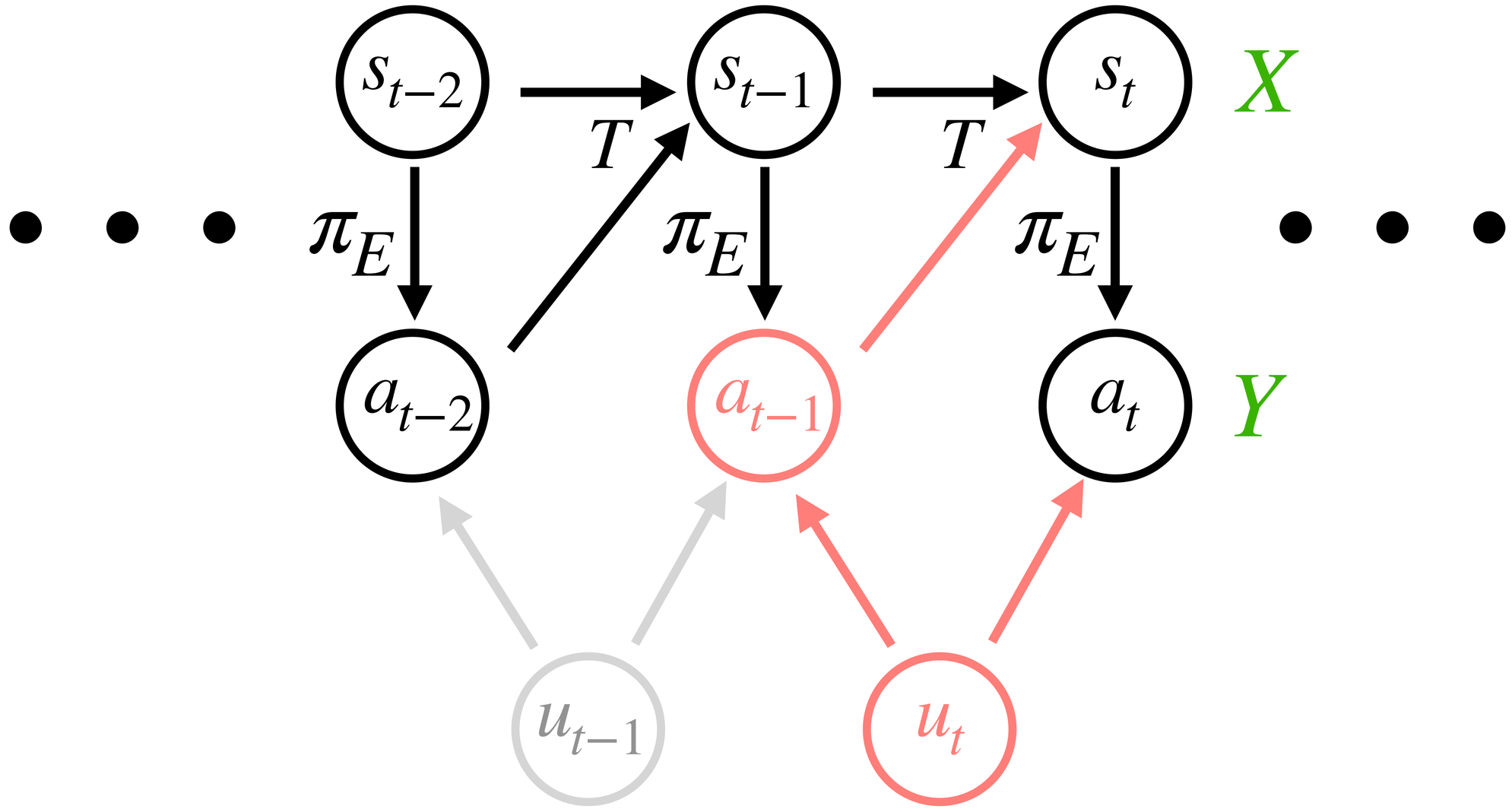


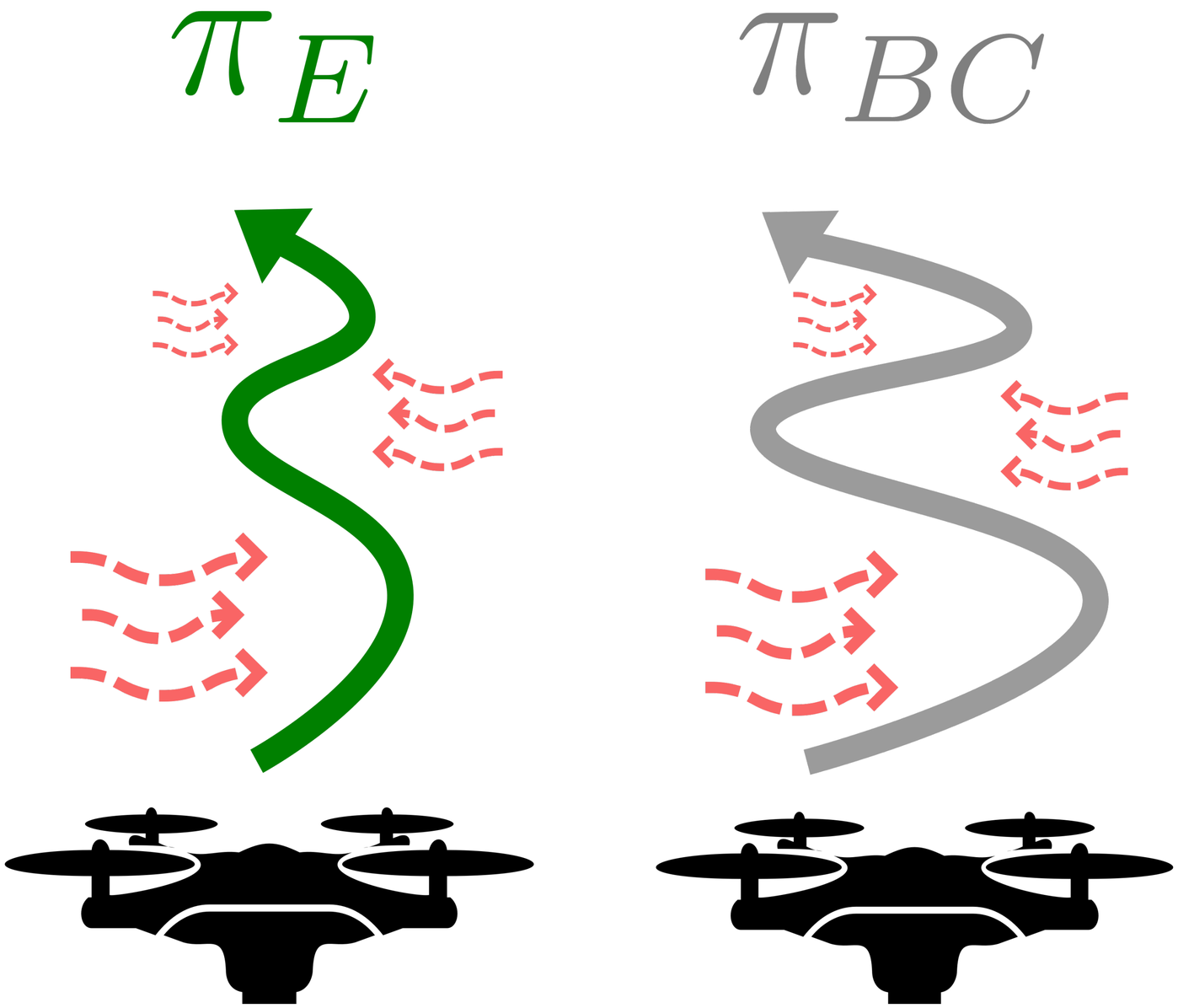
u_{t-1}

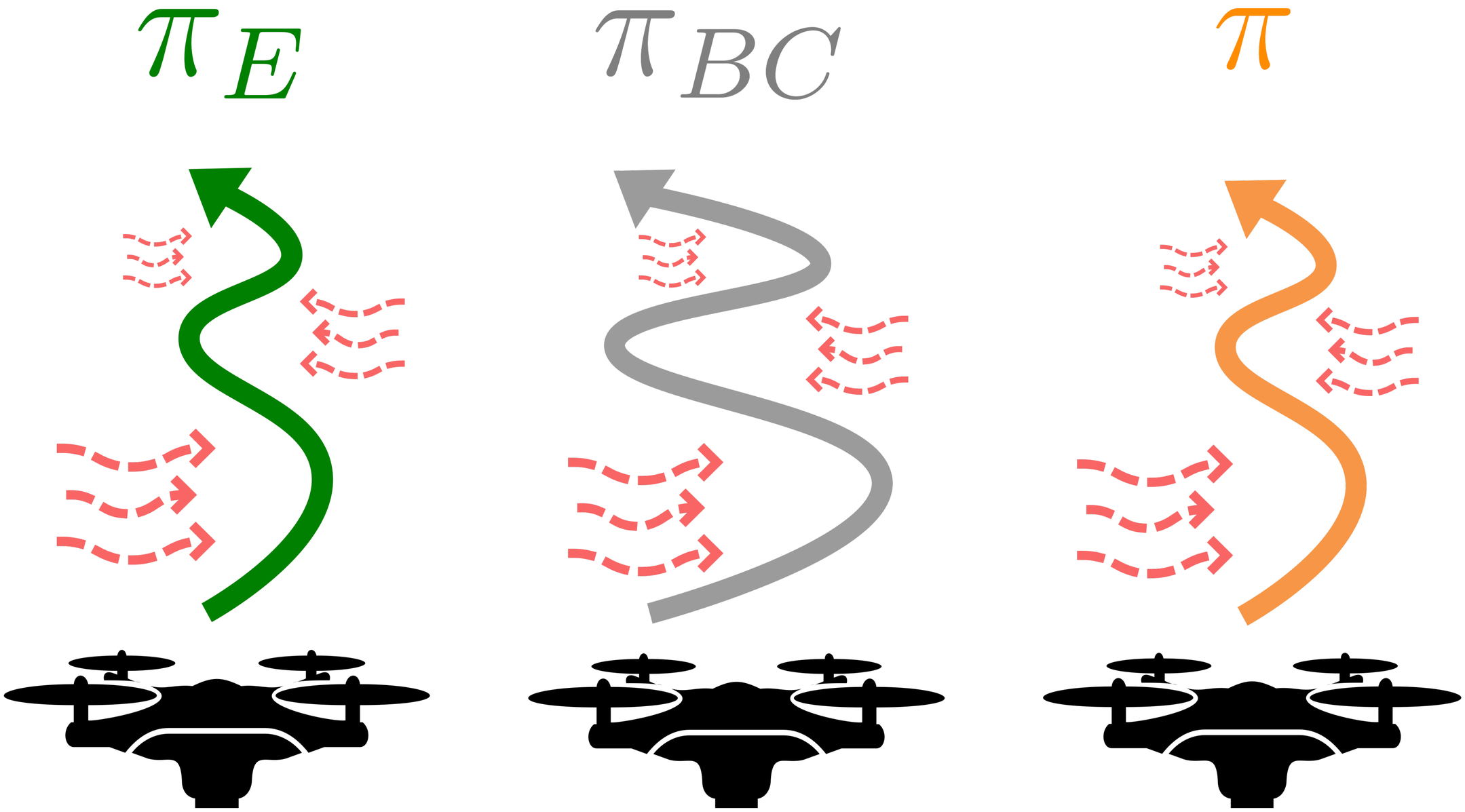

s_{t-2}
s_{t-1}
s_t
T
T
π_E
π_E
π_E
a_{t-2}
a_{t-1}
a_t
u_{t-1}

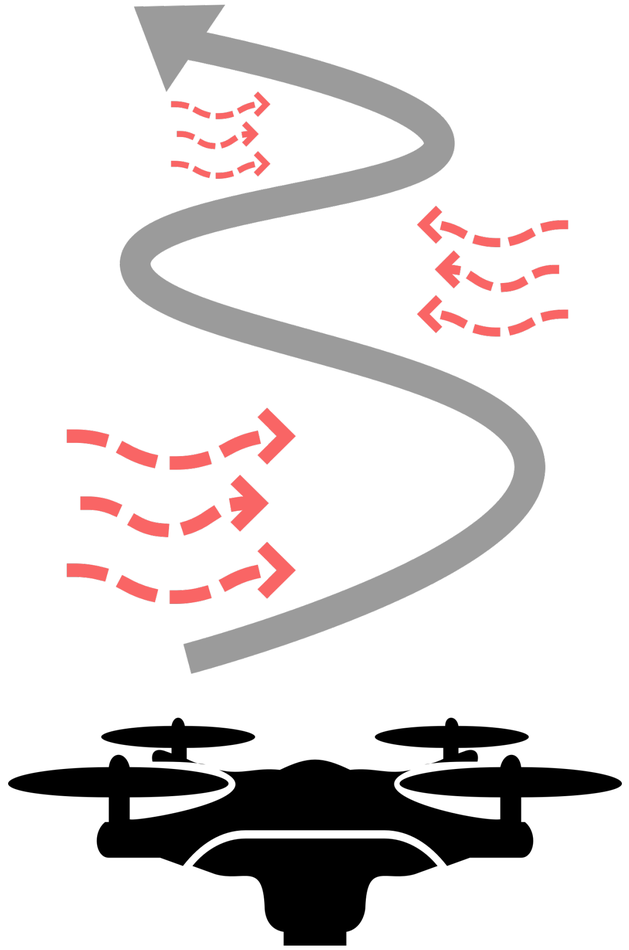

s_{t-2}
s_{t-1}
s_t
T
T
π_E
π_E
π_E
a_{t-2}
a_{t-1}
a_t
u_{t-1}
u_t

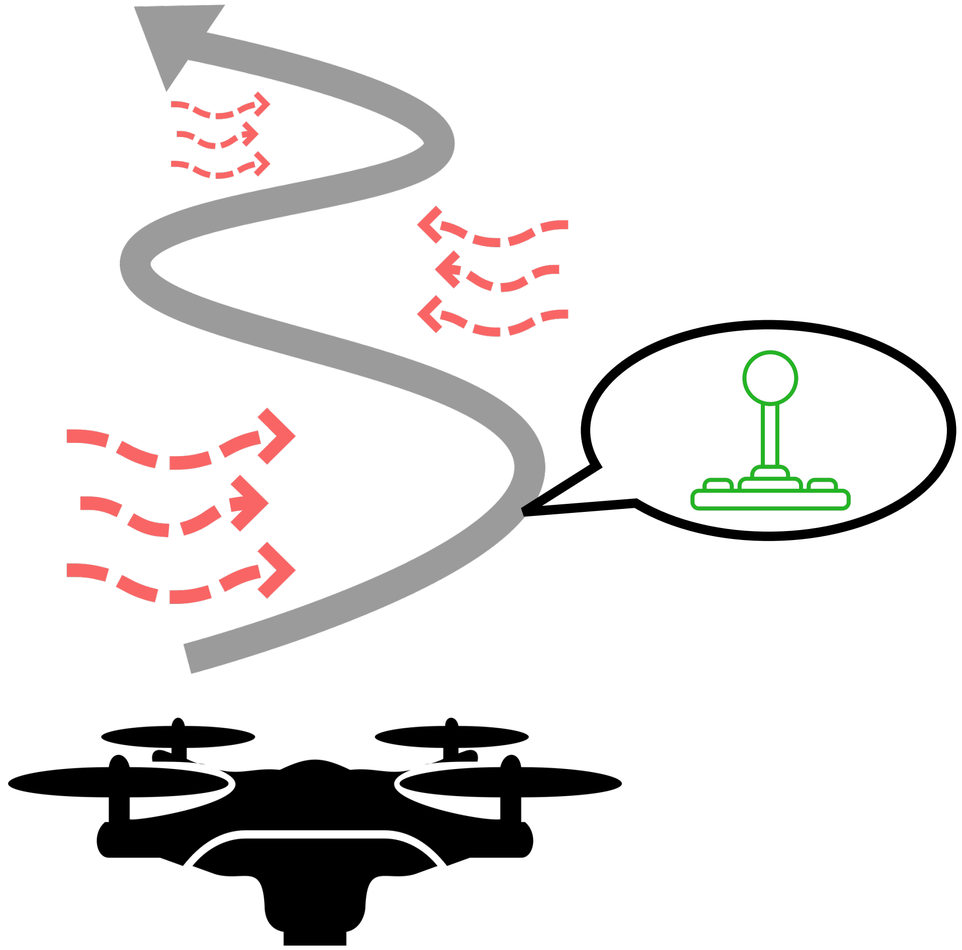

s_{t-2}
s_{t-1}
s_t
X
T
T
π_E
π_E
π_E
a_{t-2}
a_{t-1}
a_t
Y
u_{t-1}
u_t

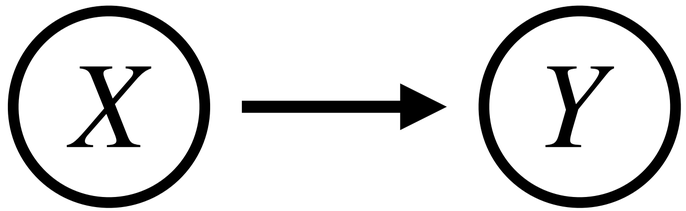

s_{t-2}
s_{t-1}
s_t
X
T
T
π_E
π_E
π_E
a_{t-2}
a_{t-1}
a_t
Y
u_{t-1}
u_t

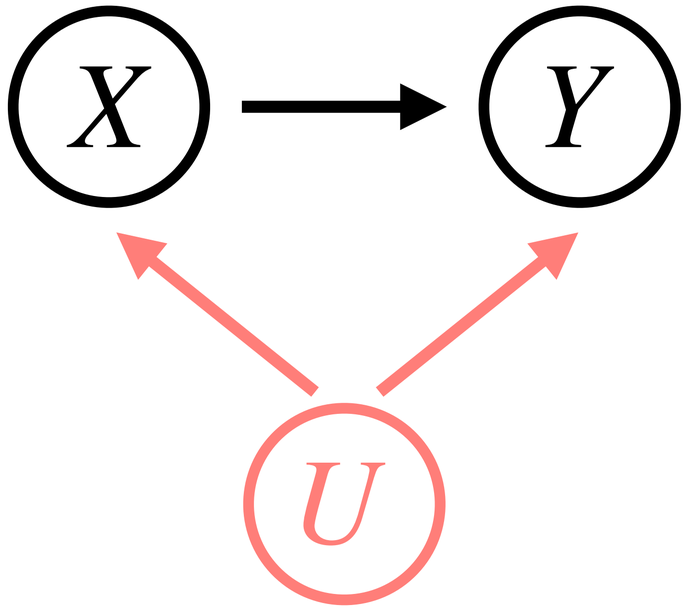

π_E
π_{BC}

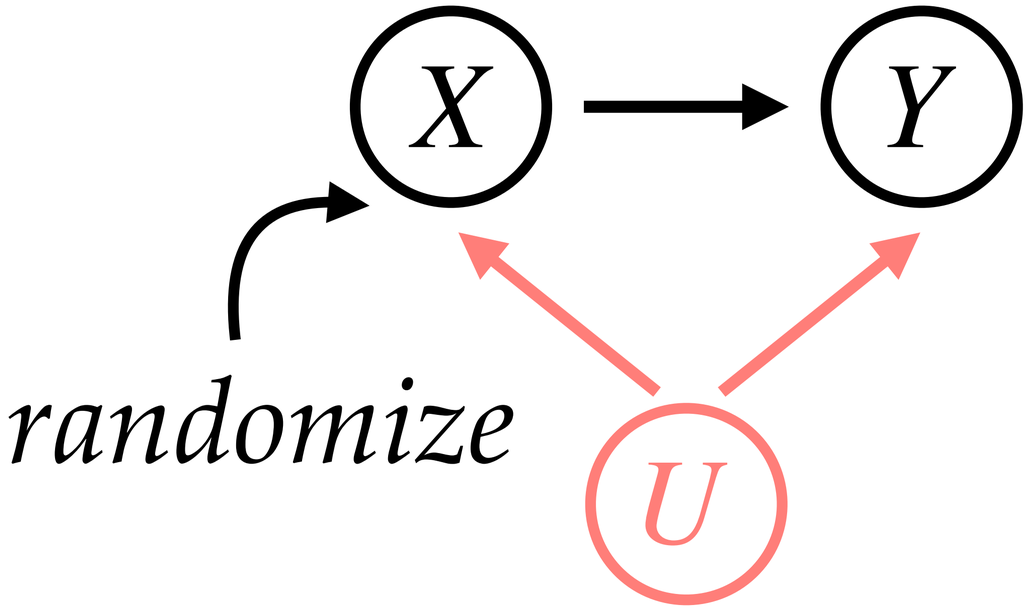

π_E
π_{BC}
π

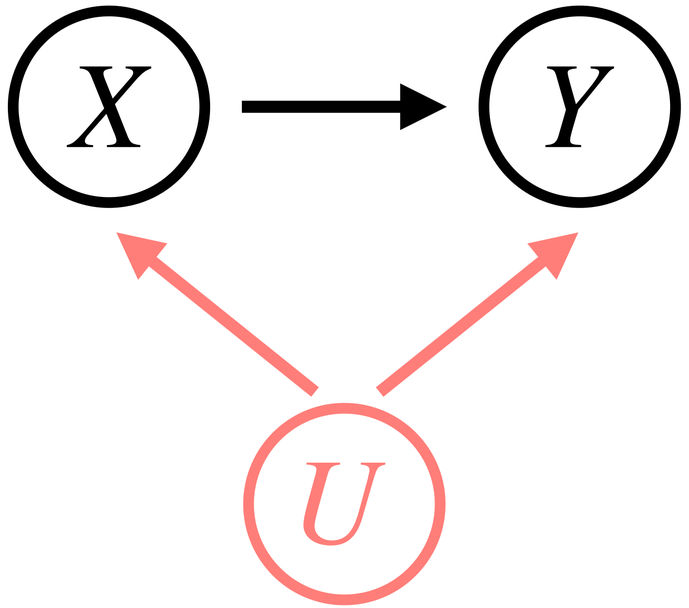


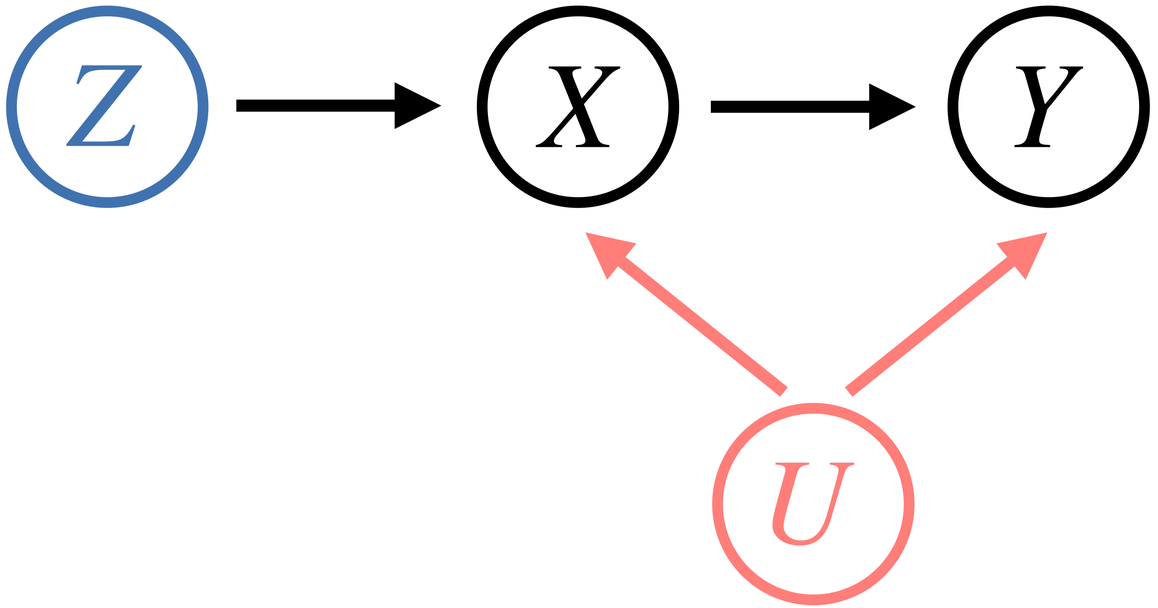


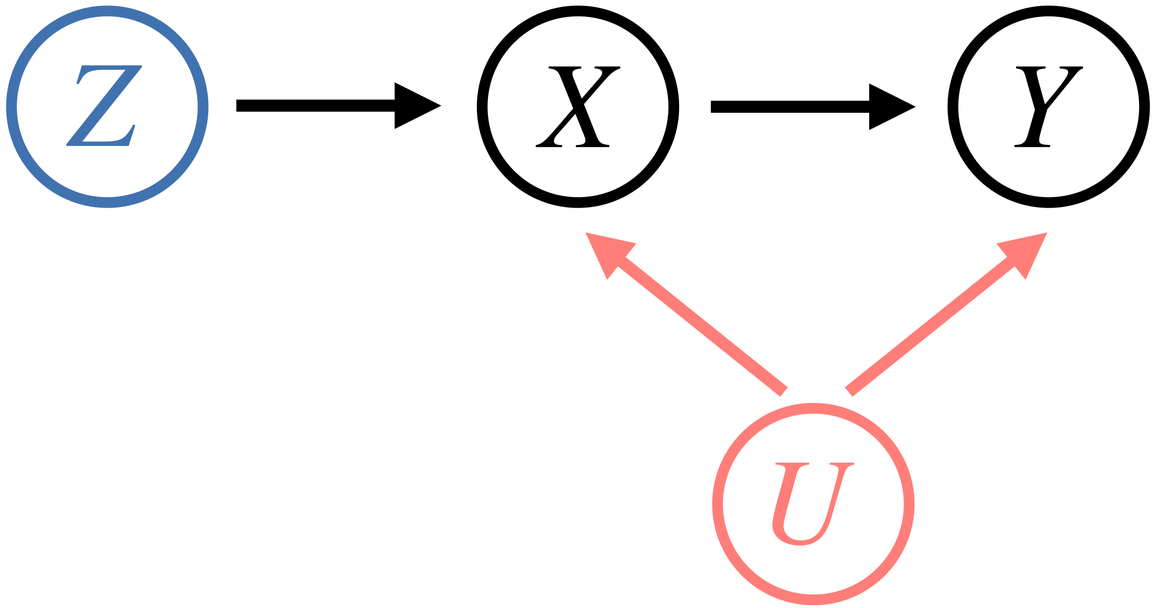

X
Y


X
Y
U

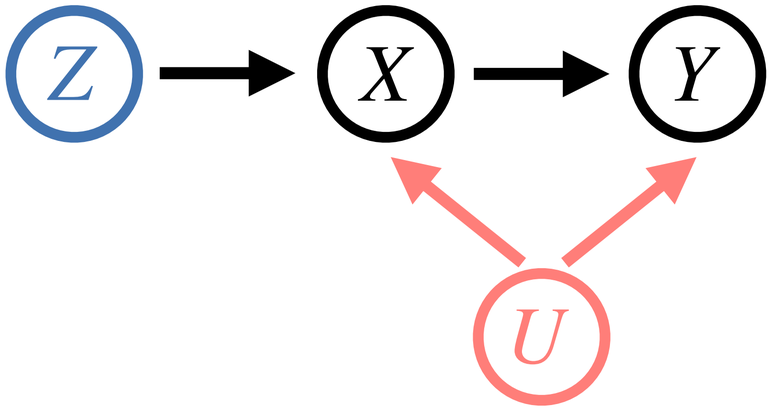

X
Y
randomize
U

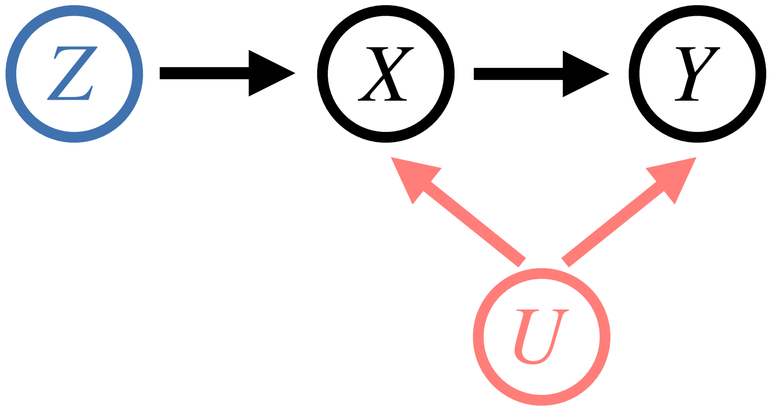

X
Y
U


Z
X
Y
U

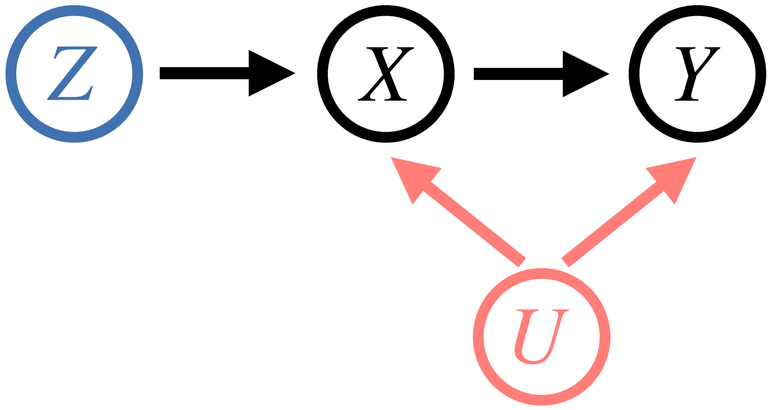


***Key Idea**: We can condition on instrument Z to counter the effect of confounder U on X.*


Z
X
Y
U

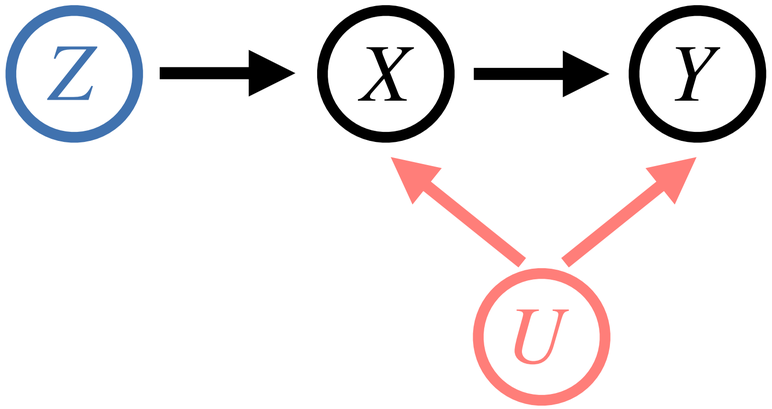


$$X = g(Z, U)$$



$$X = g(Z, U)$$

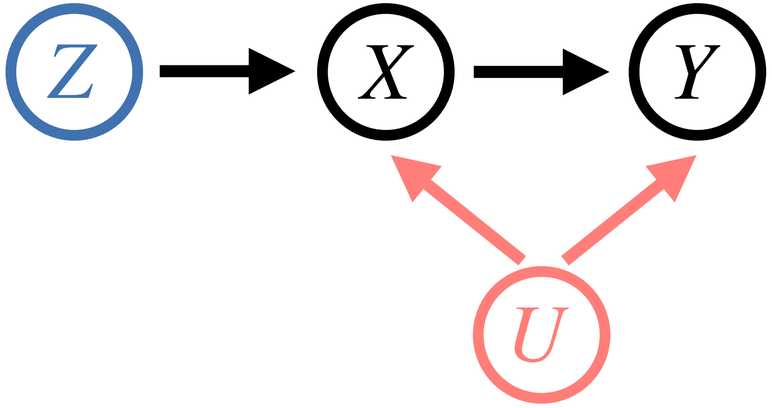
$$Y = h(X) + U$$



$$X = g(Z, U)$$

$$Y = h(X) + U$$

$$\mathbb{E}[U] = 0$$

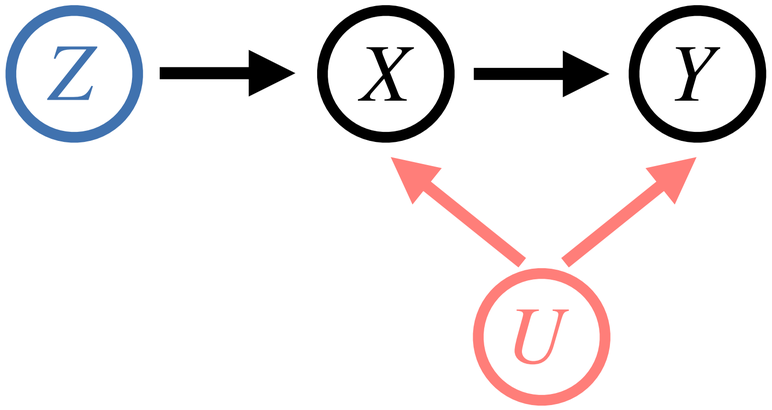


$$X = g(Z, U)$$

$$Y = h(X) + U$$

$$\mathbb{E}[U] = 0$$

$$0 = \mathbb{E}[U] = \mathbb{E}[U \mid z] = \mathbb{E}[Y - h(X) \mid z]$$



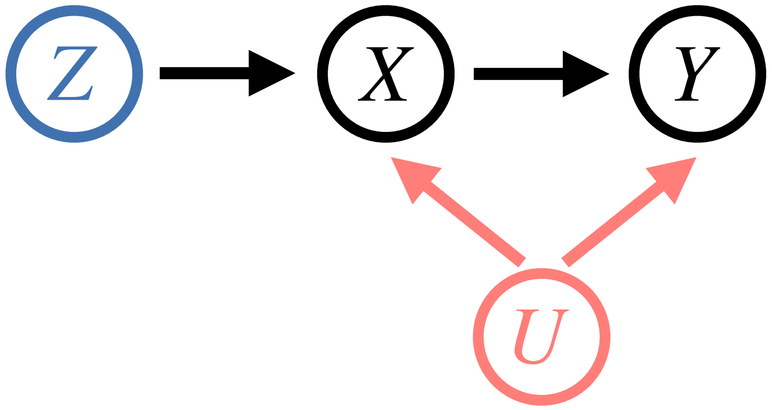
$$X = g(Z, U)$$

$$Y = h(X) + U$$

$$\mathbb{E}[U] = 0$$

$$0 = \mathbb{E}[U] = \mathbb{E}[U \mid z] = \mathbb{E}[Y - h(X) \mid z]$$

$$\Rightarrow \mathbb{E}[Y \mid z] = \mathbb{E}[h(X) \mid z], \forall z$$



$$X = g(Z, U)$$

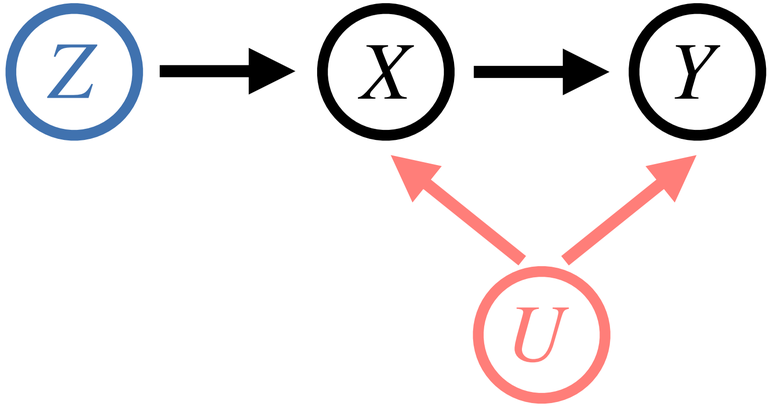
$$Y = h(X) + U$$

$$\mathbb{E}[U] = 0$$

$$0 = \mathbb{E}[U] = \mathbb{E}[U \mid z] = \mathbb{E}[Y - h(X) \mid z]$$

$$\Rightarrow \mathbb{E}[Y \mid z] = \mathbb{E}[h(X) \mid z], \forall z$$

$$\Rightarrow \min_h \mathbb{E}_z[(\mathbb{E}[Y \mid z] - \mathbb{E}[h(X) \mid z])^2]$$



$$X = g(Z, U)$$

$$Y = h(X) + U$$

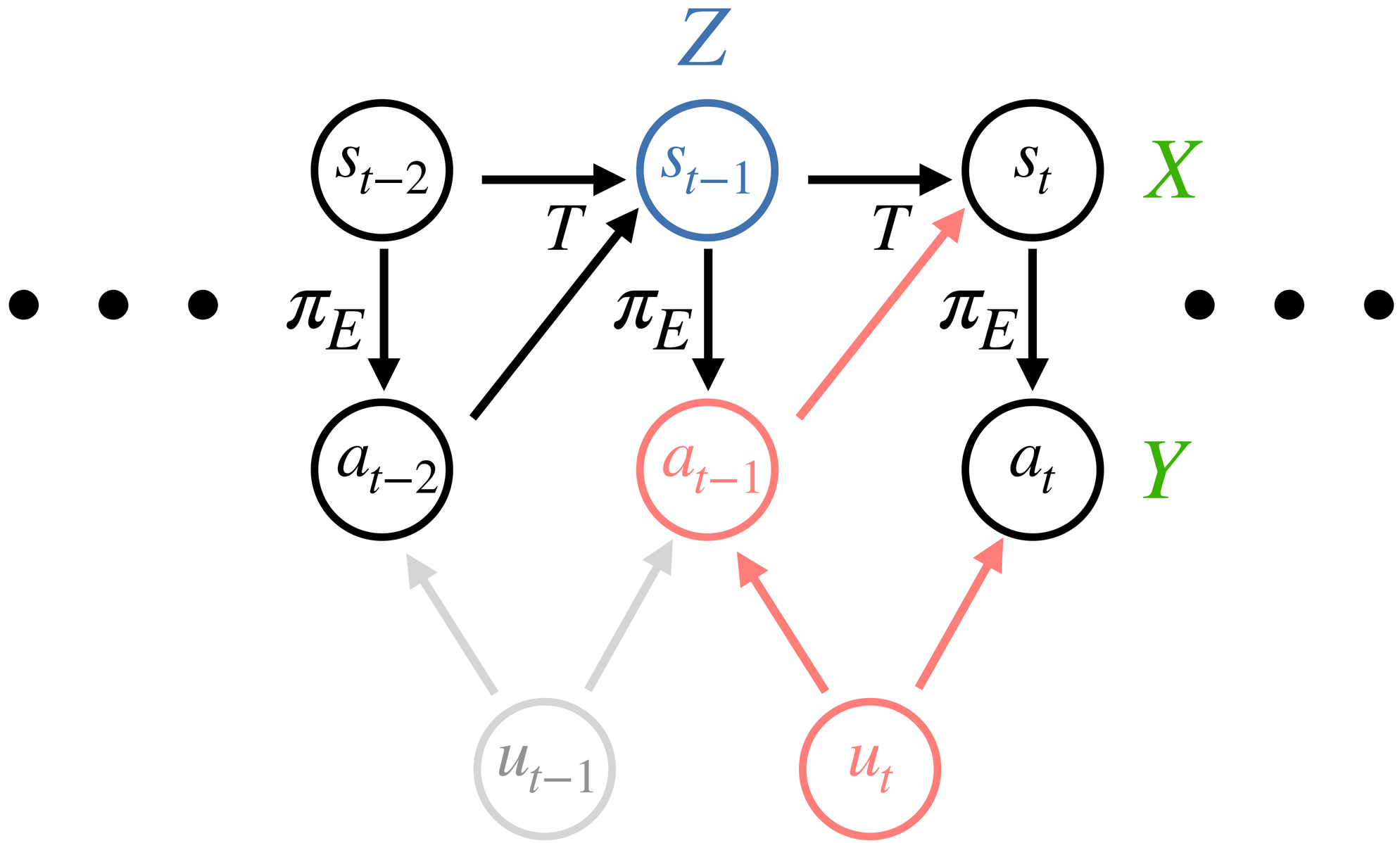
$$\mathbb{E}[U] = 0$$

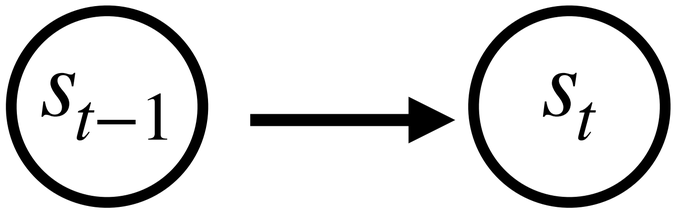
$$0 = \mathbb{E}[U] = \mathbb{E}[U \mid z] = \mathbb{E}[Y - h(X) \mid z]$$

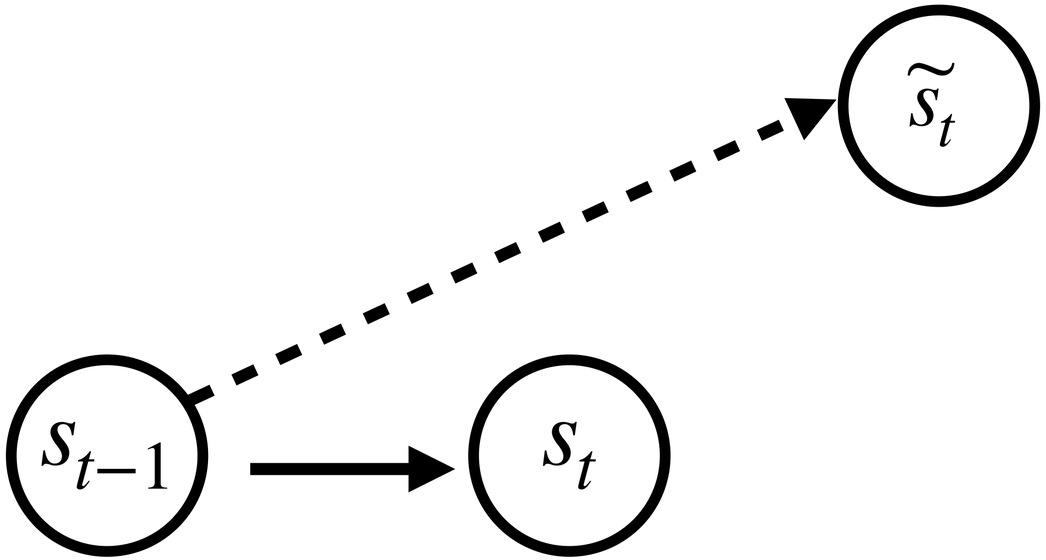
$$\Rightarrow \mathbb{E}[Y \mid z] = \mathbb{E}[h(X) \mid z], \forall z$$

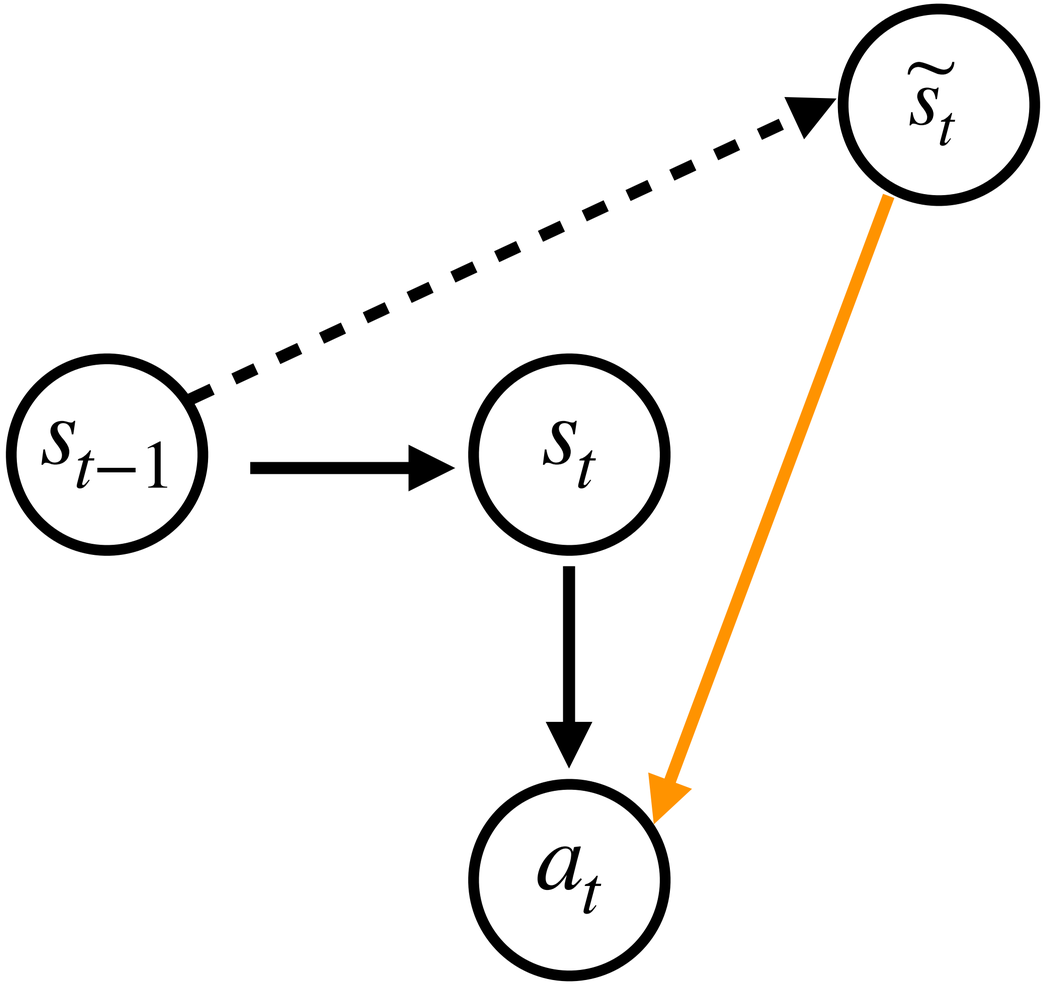
$$\Rightarrow \min_h \mathbb{E}_z[(\mathbb{E}[Y \mid z] - \mathbb{E}[h(X) \mid z])^2]$$

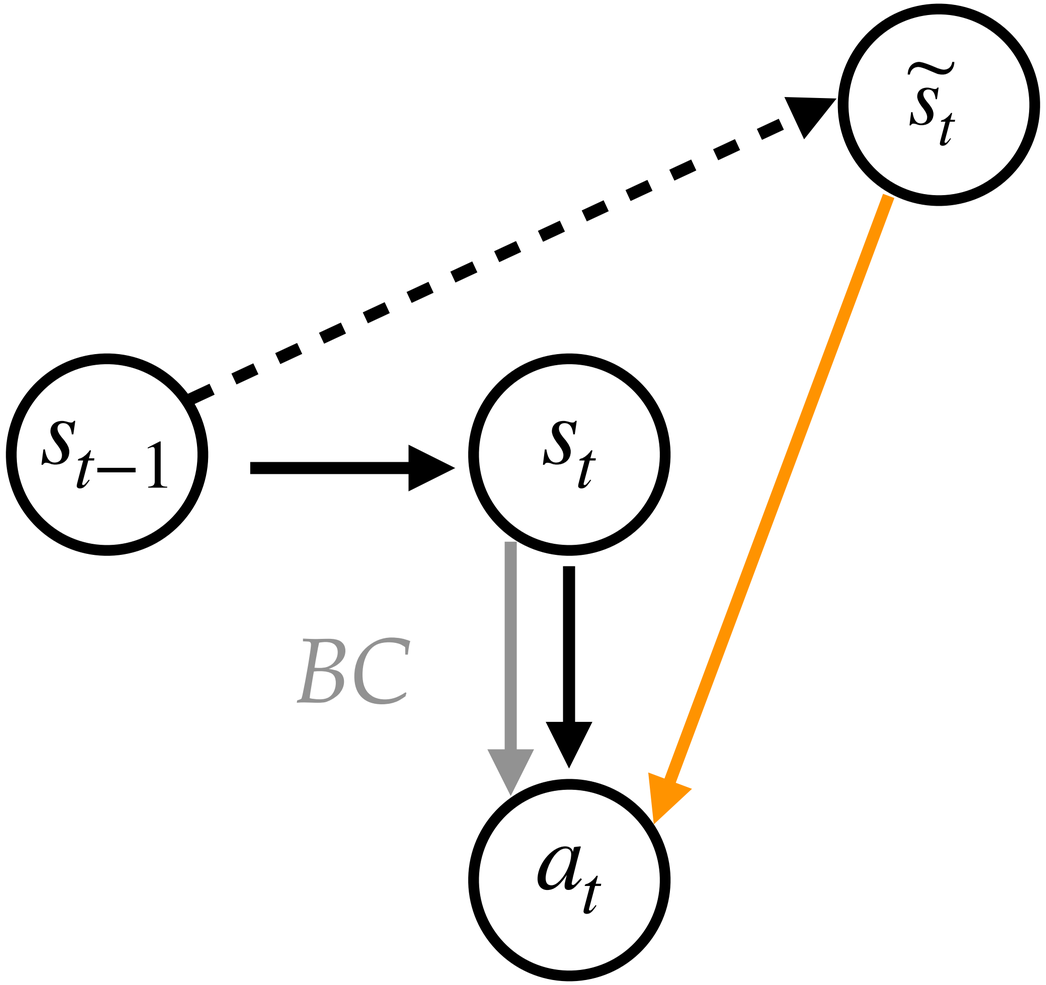
$$\Leftrightarrow \min_h \max_f \mathbb{E}_z[2(Y - h(X))f(Z) - f^2(Z)]$$

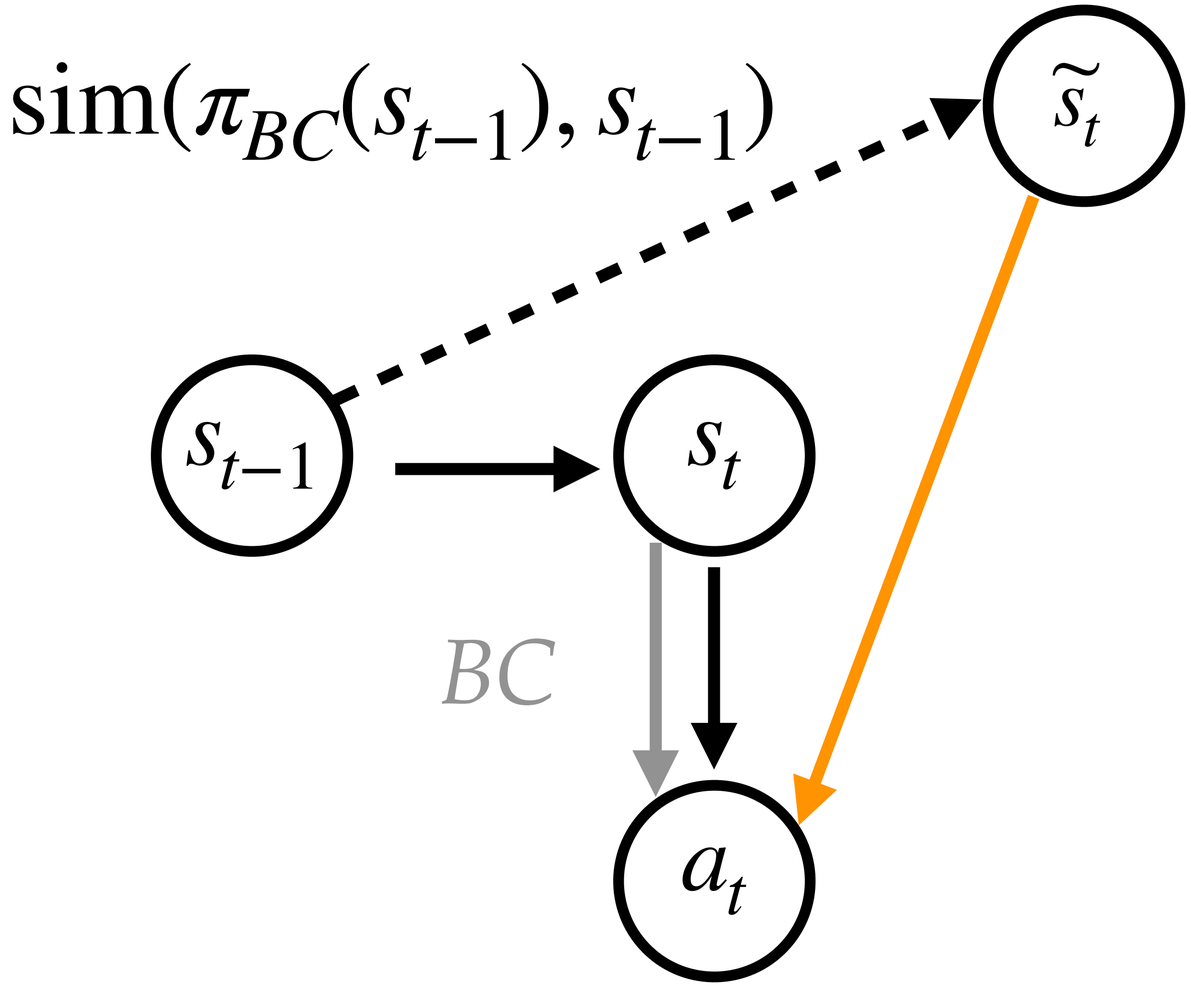

Z
s_{t-2}
s_{t-1}
s_t
X
T
T
π_E
π_E
π_E
a_{t-2}
a_{t-1}
a_t
Y
u_{t-1}
u_t

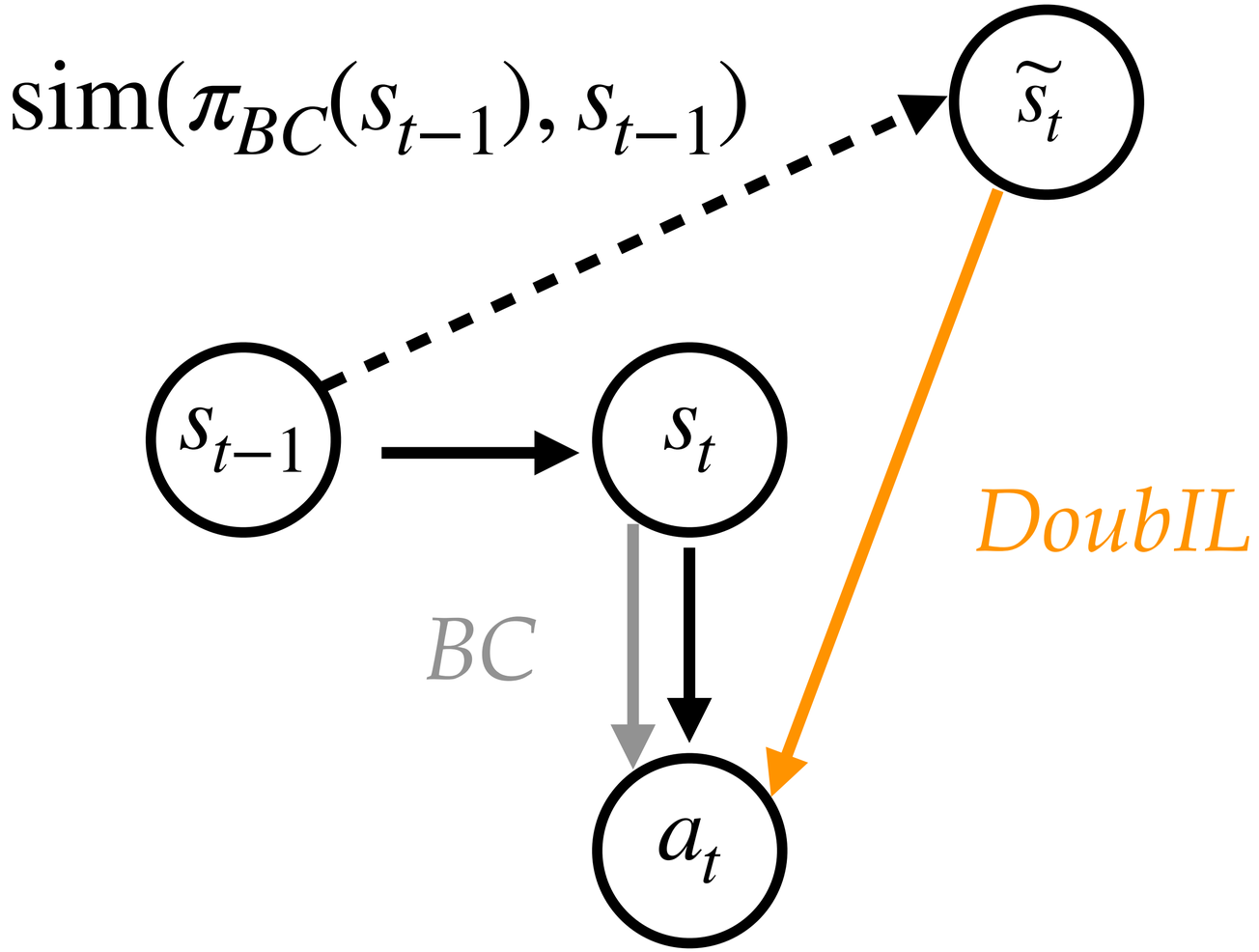

s_{t-1}
s_t

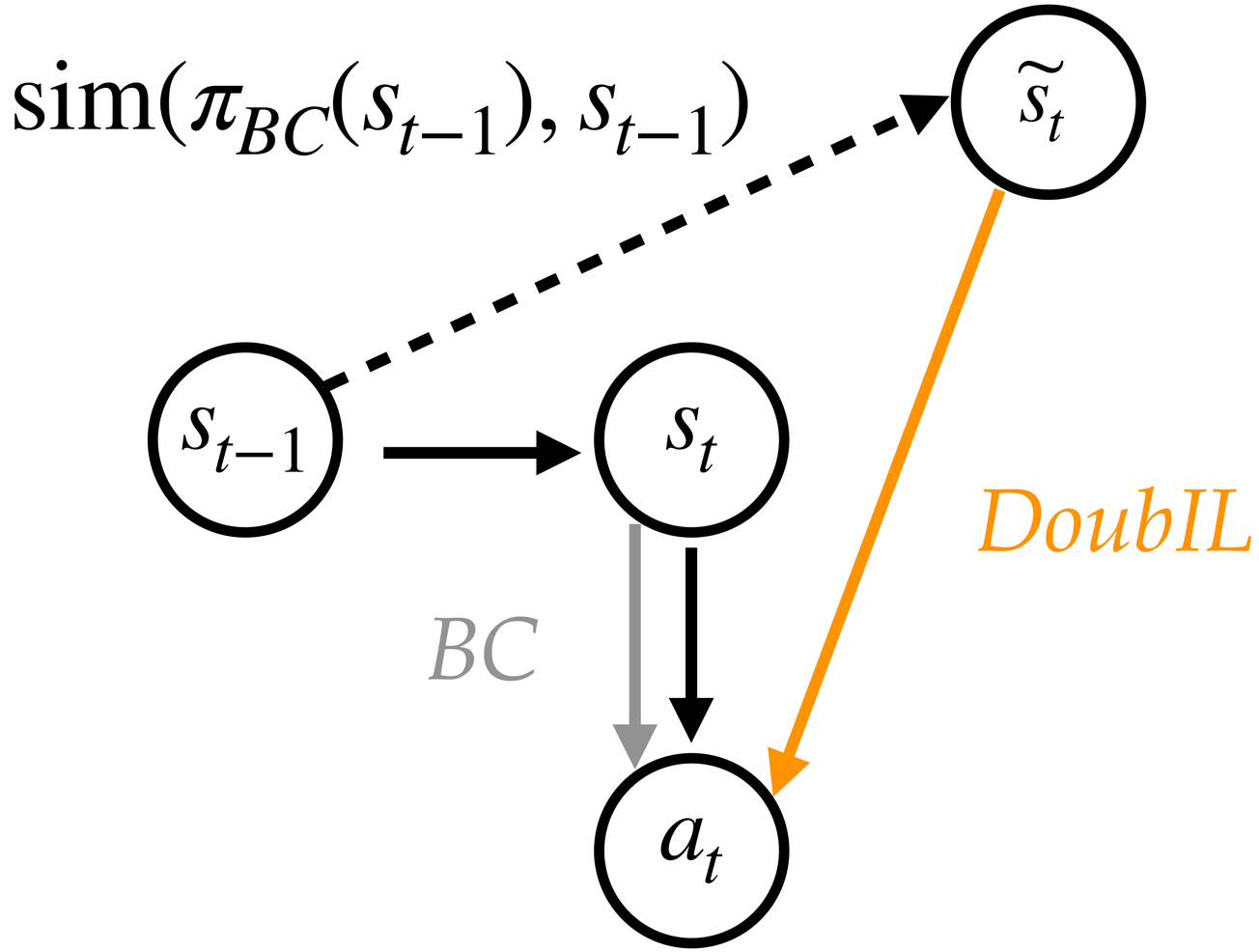

$\widetilde{s}_t$
s_{t-1}
s_t


$\tilde{s}_t$
s_{t-1}
s_t
a_t


$\tilde{s}_t$
s_{t-1}
s_t
BC
a_t


$\mathrm{sim}(\pi_{BC}(s_{t-1}), s_{t-1})$
$\tilde{s}_t$
s_{t-1}
s_t
BC
a_t


$\mathrm{sim}(\pi_{BC}(s_{t-1}), s_{t-1})$
$\tilde{s}_t$
s_{t-1}
s_t
a_t
BC
DoubIL



$$J(\pi_E) - J(\pi) \leq c(\sqrt{\epsilon} + \sqrt{\delta})\kappa(\Pi)T^2$$

$$\min_{\pi} \max_{f} \mathbb{E}[2(a_t - \pi(s_t))f(s_{t-1}) - f(s_{t-1})^2]$$

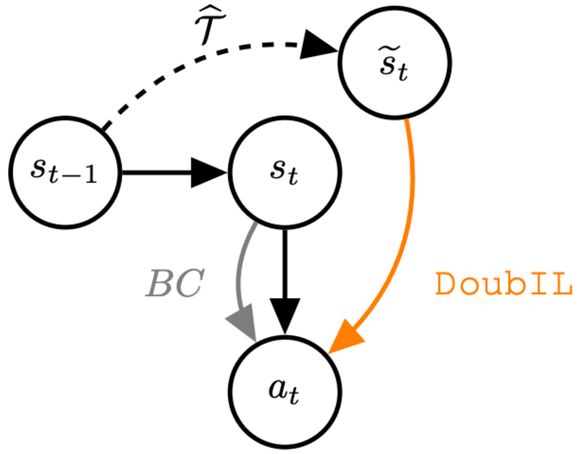
$$\min_{\pi} \max_{f} \mathbb{E}[2(a_t - \pi(s_t))f(s_{t-1}) - f(s_{t-1})^2]$$

$$J(\pi_E) - J(\pi) \leq c\sqrt{\epsilon}\kappa(\Pi)T^2$$

Instrumental Variable Imitation Learning

generative modeling

DoubIL



$$J(\pi_E) - J(\pi) \leq c(\sqrt{\epsilon} + \sqrt{\delta})\kappa(\Pi)T^2$$

game-theoretic

ResiduIL

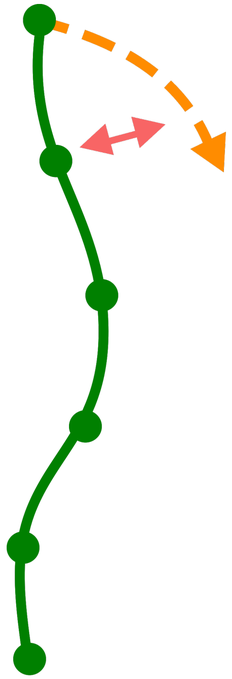
$$\min_{\pi} \max_{f} \mathbb{E}[2(a_t - \pi(s_t))f(s_{t-1}) - f(s_{t-1})^2]$$

$$J(\pi_E) - J(\pi) \leq c\sqrt{\epsilon}\kappa(\Pi)T^2$$

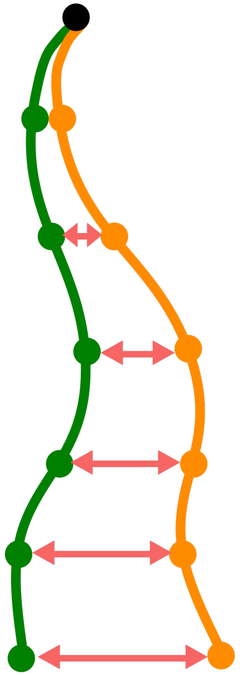
$\pi_E \overset{f}{\longleftrightarrow} \pi$

[Swamy et al, 2021]

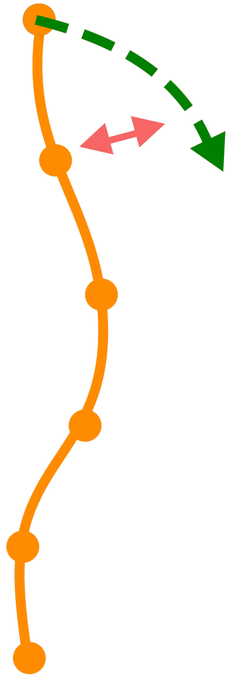
Offline



Online



Interactive



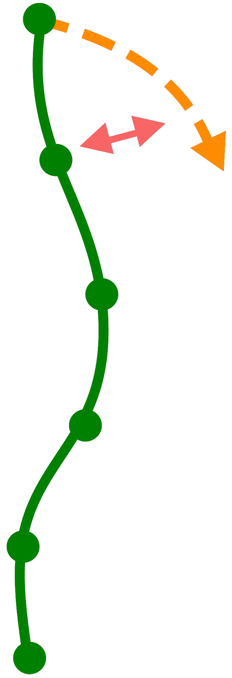
Inconsistent,
IVR Consistent

Consistent

$\pi_E \overset{f}{\leftrightarrow} \pi$

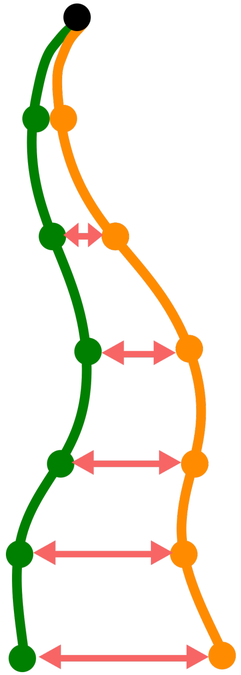
[Swamy et al, 2021]

Offline



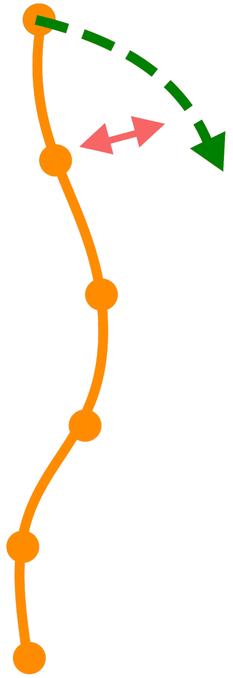
Inconsistent,
IVR Consistent

Online

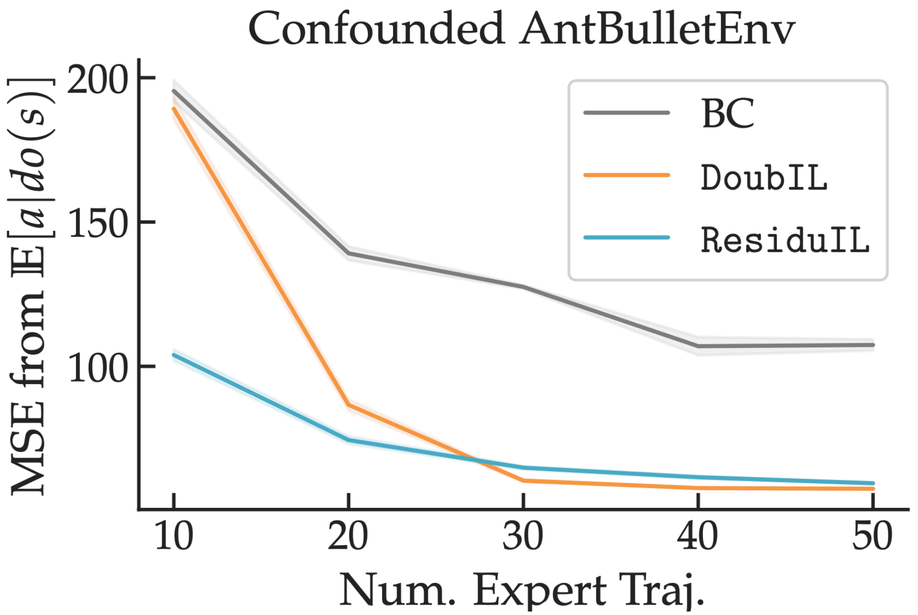


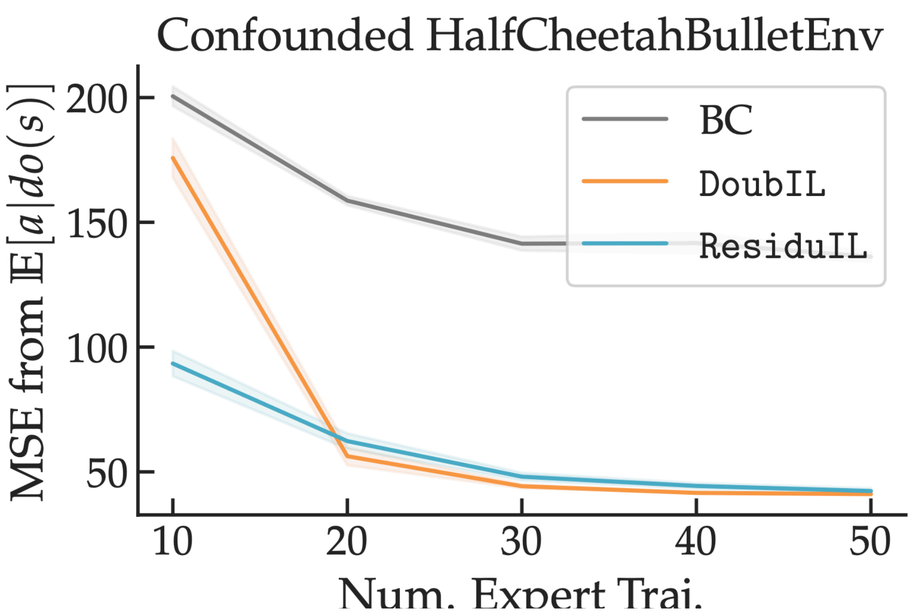
Inconsistent,
Hybrid?

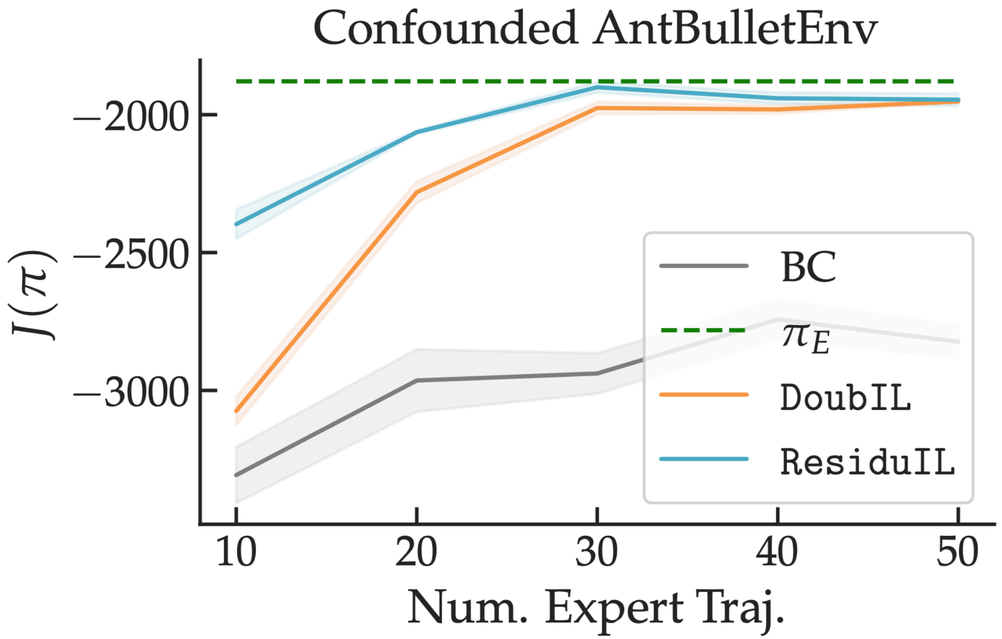
Interactive

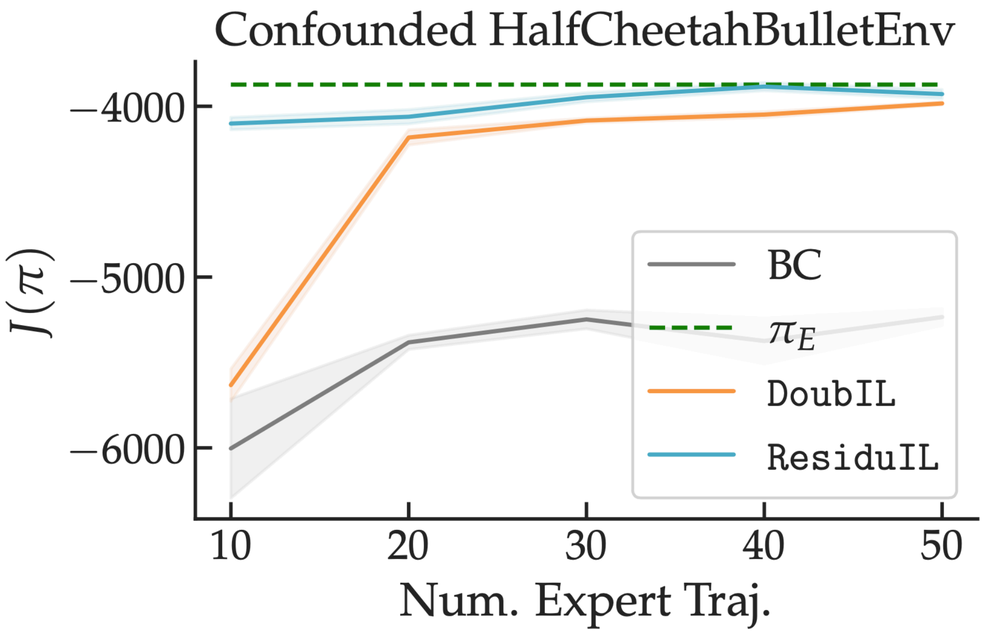


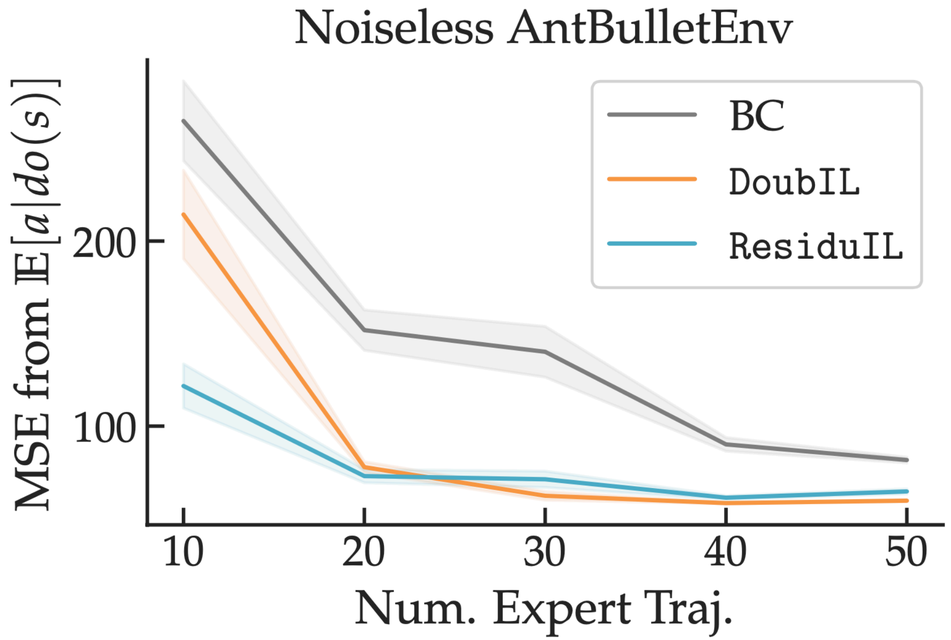
Consistent

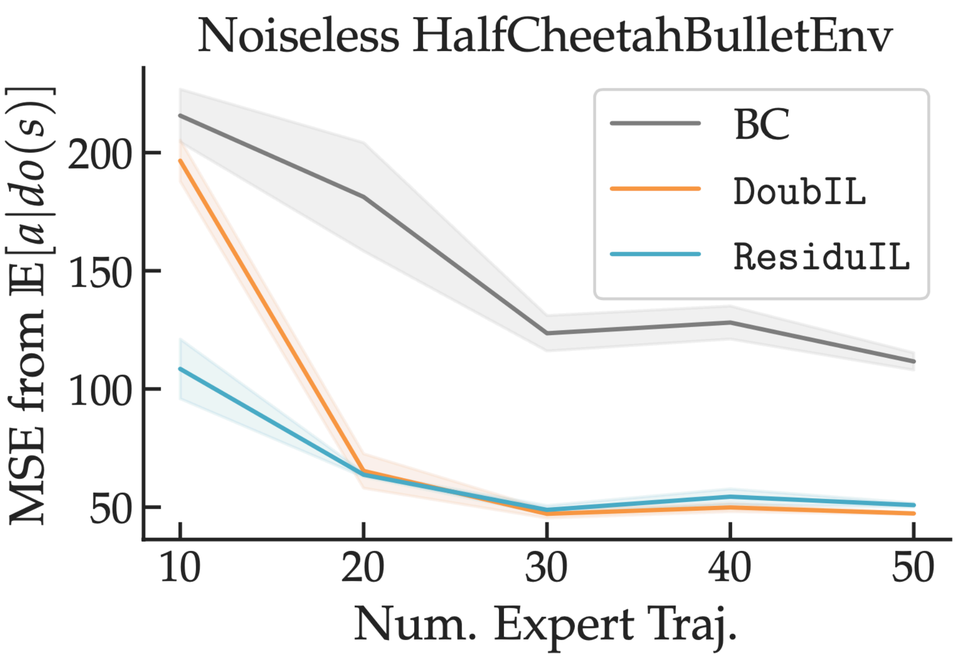

Confounded AntBulletEnv
MSE from $\mathbb{E}[a|do(s)]$
200
150
100
10
20
30
40
50
Num. Expert Traj.
BC
DoubIL
ResiduIL

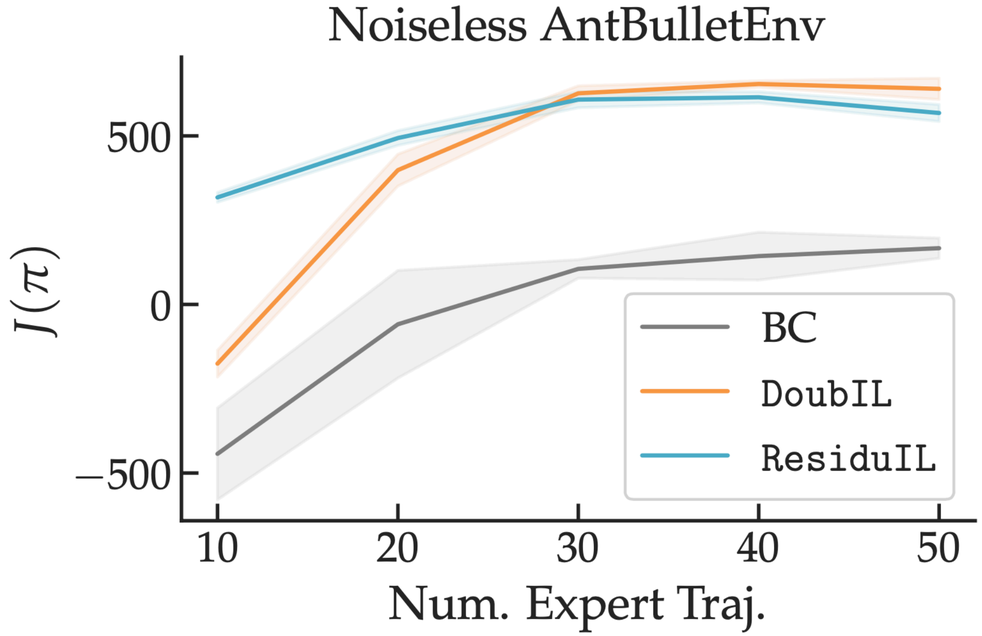

Confounded HalfCheetahBulletEnv
MSE from $\mathbb{E}[a|do(s)]$
200
150
100
50
10
20
30
40
50
Num. Expert Traj.
BC
DoubIL
ResiduIL

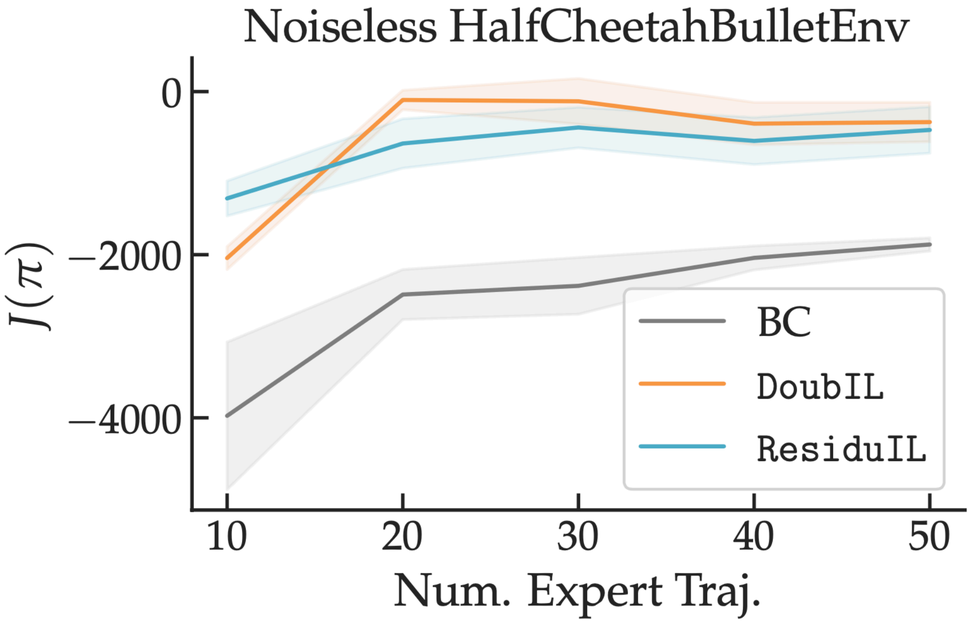

Confounded AntBulletEnv
$J(\pi)$
−2000
−2500
−3000
10
20
30
40
50
Num. Expert Traj.
BC
π_E
DoubIL
ResiduIL

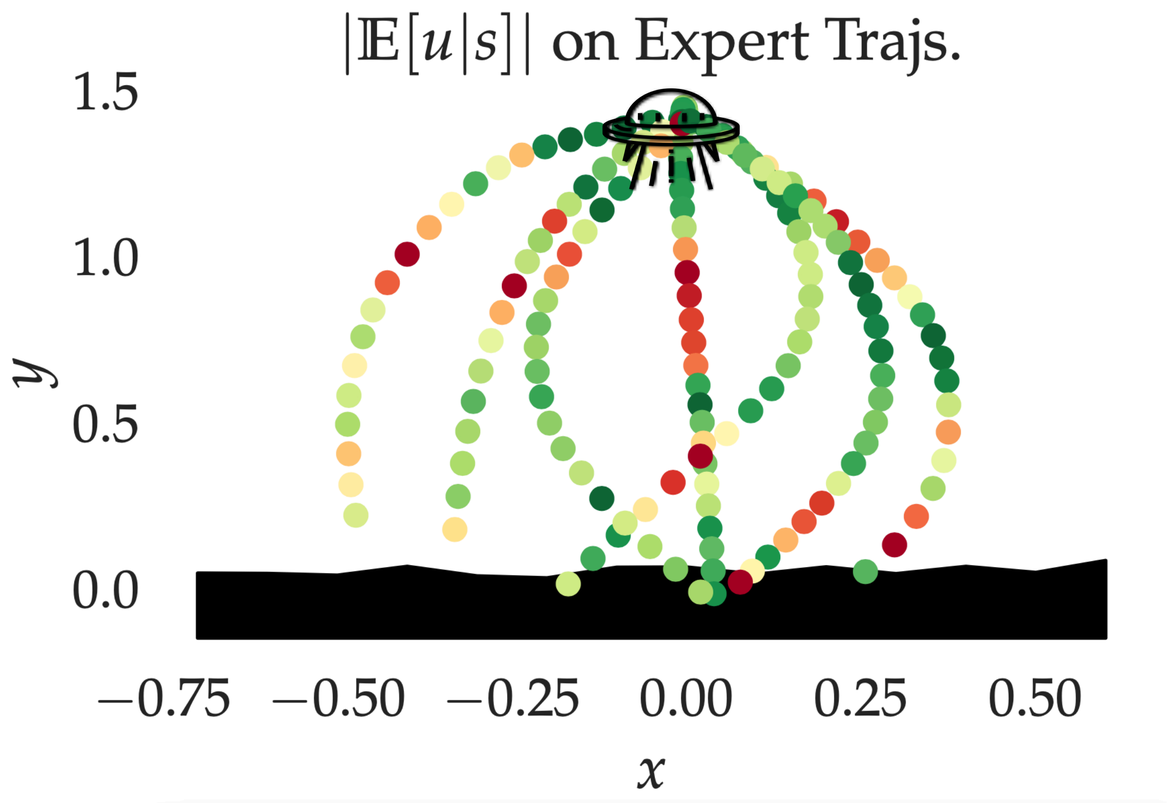

Confounded HalfCheetahBulletEnv
$J(\pi)$
−4000
−5000
−6000
10
20
30
40
50
Num. Expert Traj.
BC
π_E
DoubIL
ResiduIL

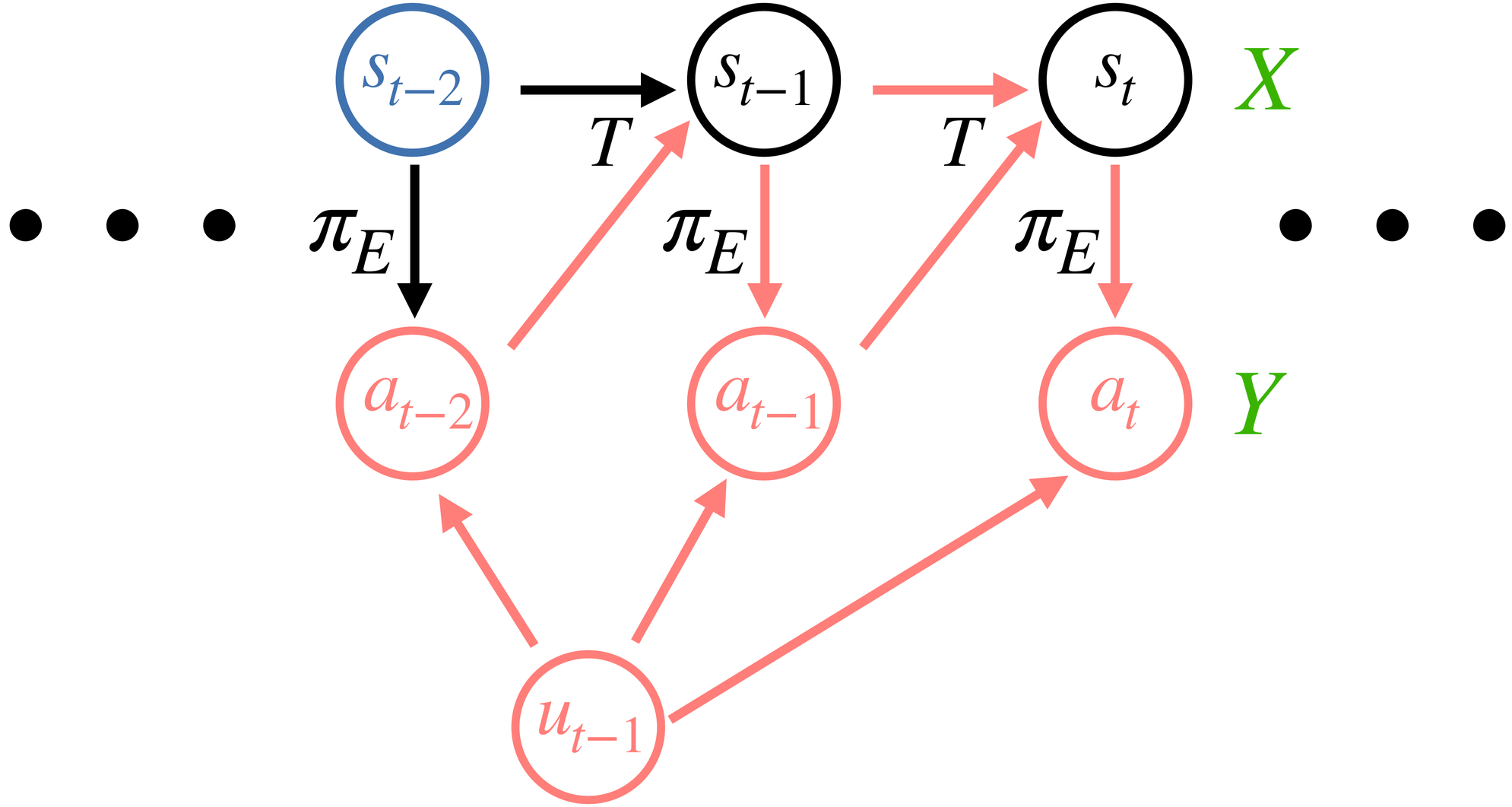

Noiseless AntBulletEnv
MSE from $\mathbb{E}[a|do(s)]$
200
100
10
20
30
40
50
Num. Expert Traj.
BC
DoubIL
ResiduIL

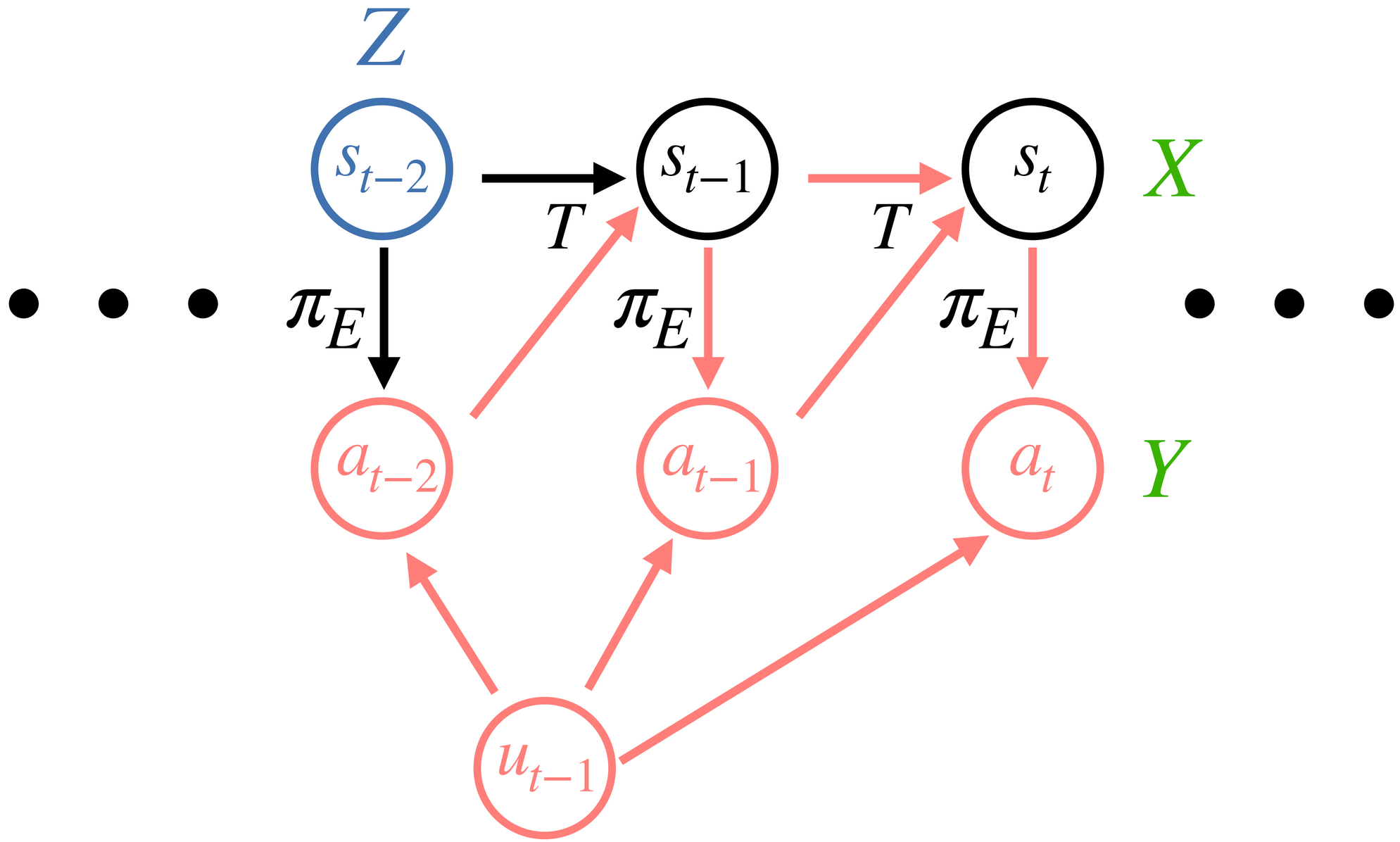

Noiseless HalfCheetahBulletEnv
MSE from $\mathbb{E}[a|do(s)]$
200
150
100
50
10
20
30
40
50
Num. Expert Traj.
BC
DoubIL
ResiduIL

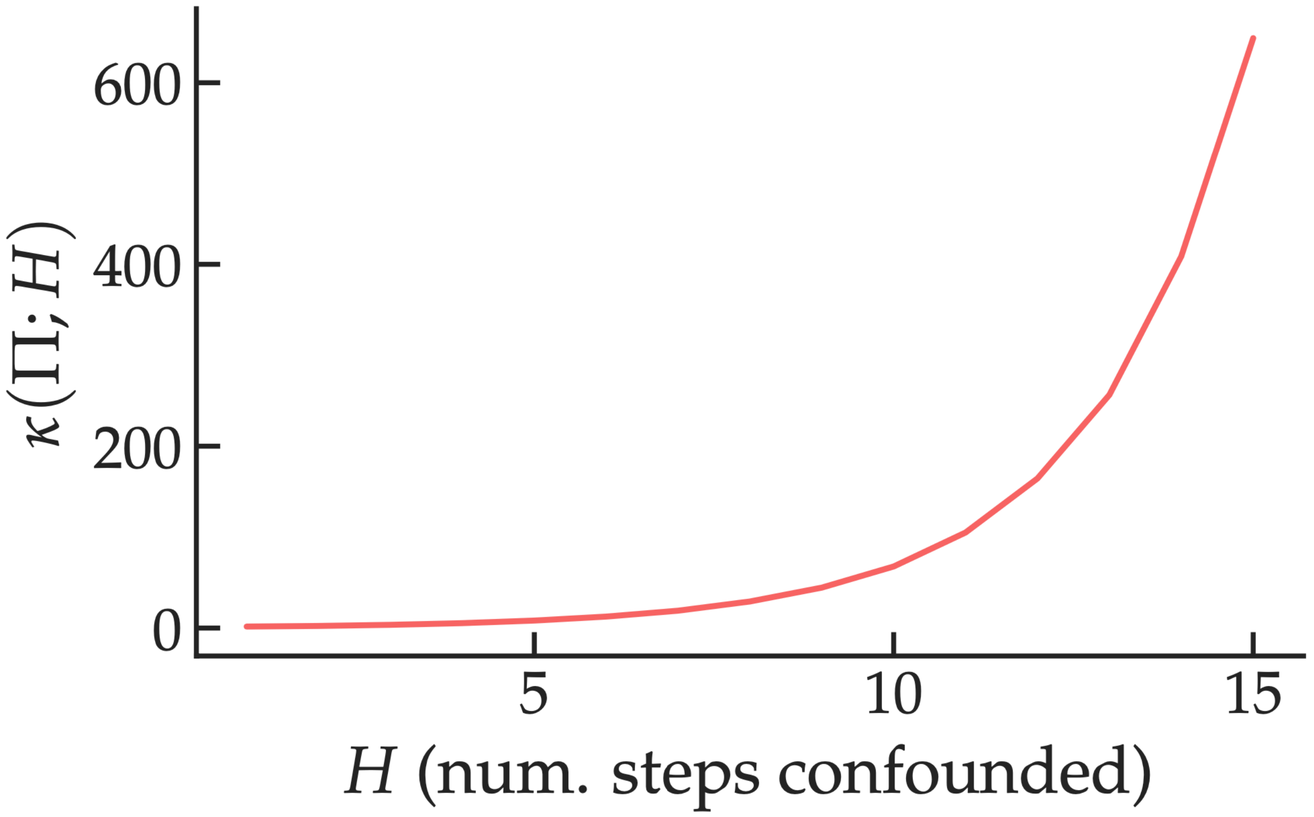

Noiseless AntBulletEnv
$J(\pi)$
500
0
−500
10
20
30
40
50
Num. Expert Traj.
BC
DoubIL
ResiduIL


Noiseless HalfCheetahBulletEnv
$J(\pi)$
0
−2000
−4000
10
20
30
40
50
Num. Expert Traj.
BC
DoubIL
ResiduIL


$|\mathbb{E}[u|s]|$ on Expert Trajs.
1.5
1.0
0.5
0.0
y
−0.75
−0.50
−0.25
0.00
0.25
0.50
x


s_{t-2}
s_{t-1}
s_t
X
T
T
π_E
π_E
π_E
a_{t-2}
a_{t-1}
a_t
Y
u_{t-1}


Z
s_{t-2}
s_{t-1}
s_t
X
T
T
π_E
π_E
π_E
a_{t-2}
a_{t-1}
a_t
Y
u_{t-1}


600
400
200
0
$\kappa(\Pi; H)$
5
10
15
H (num. steps confounded)

Thanks!

https://gokul.dev/causil/

gswamy@cmu.edu