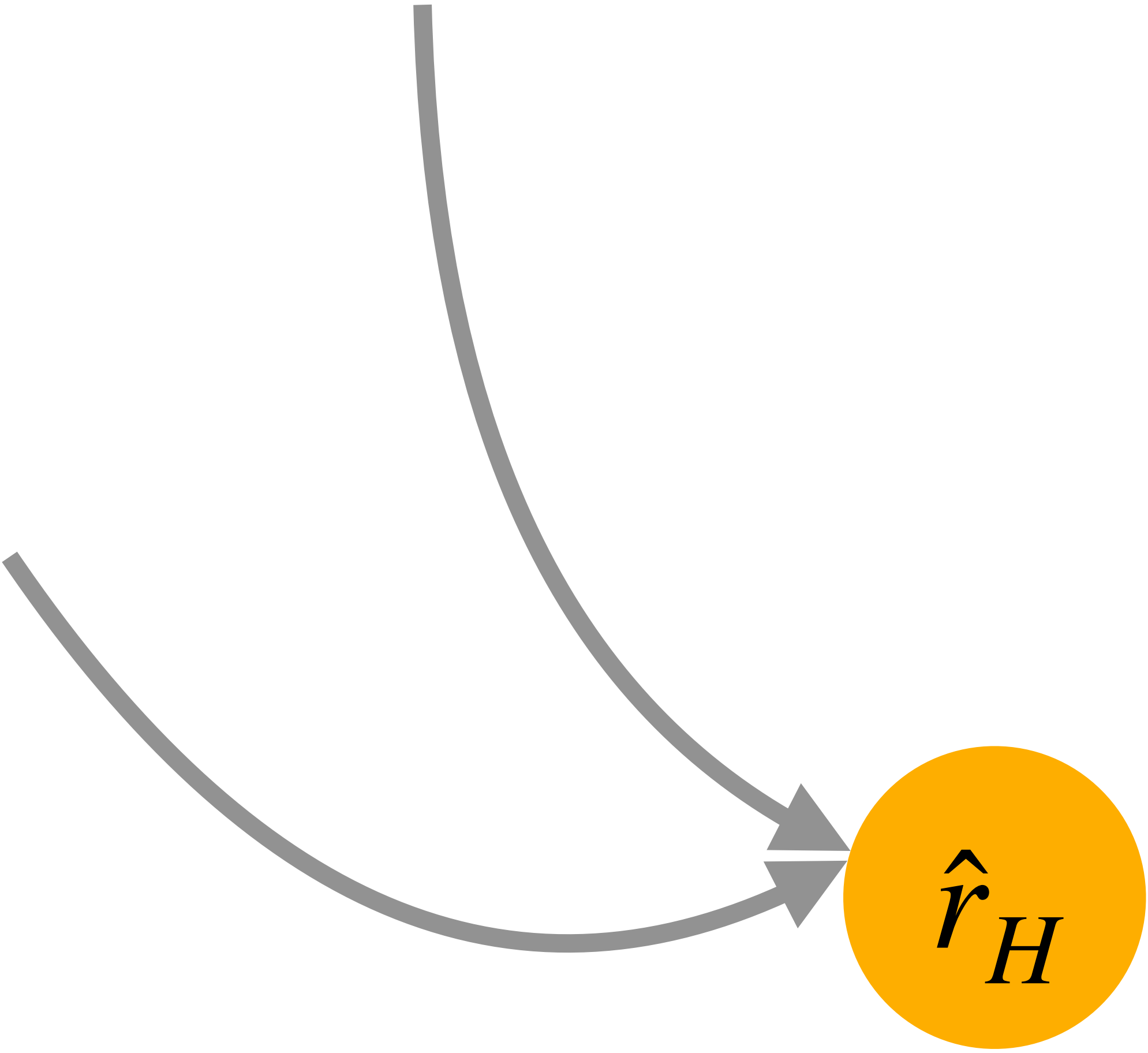
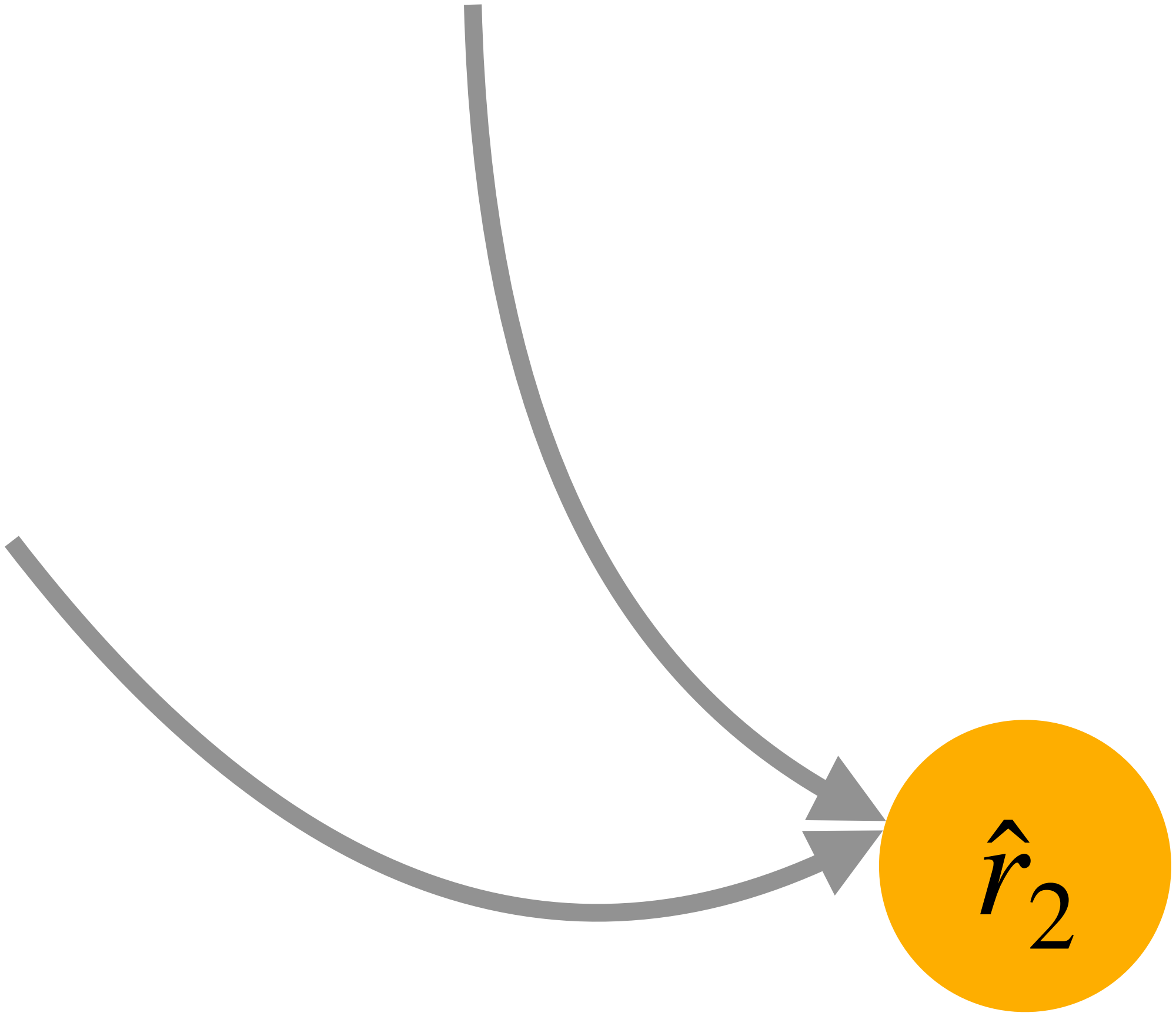
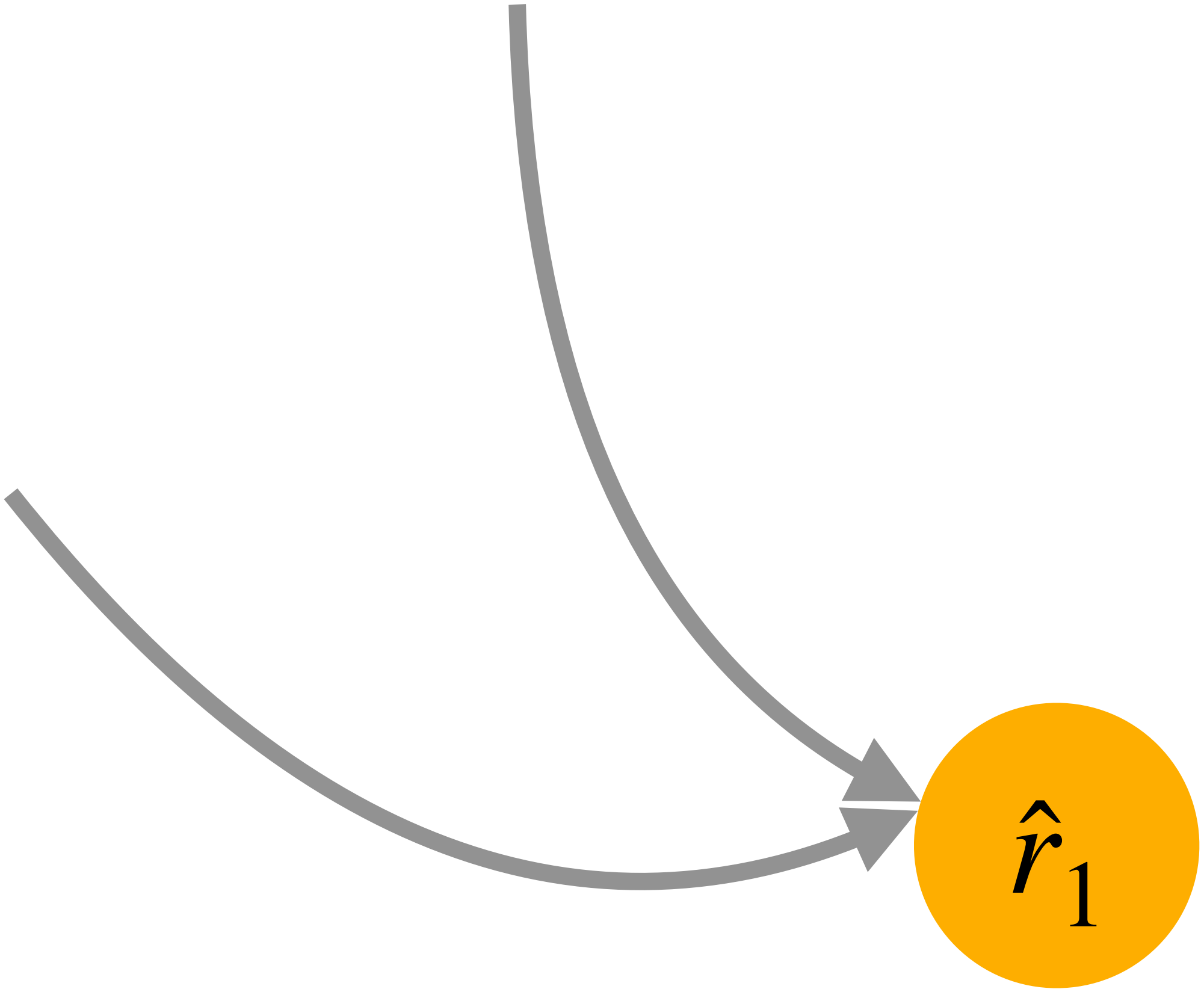
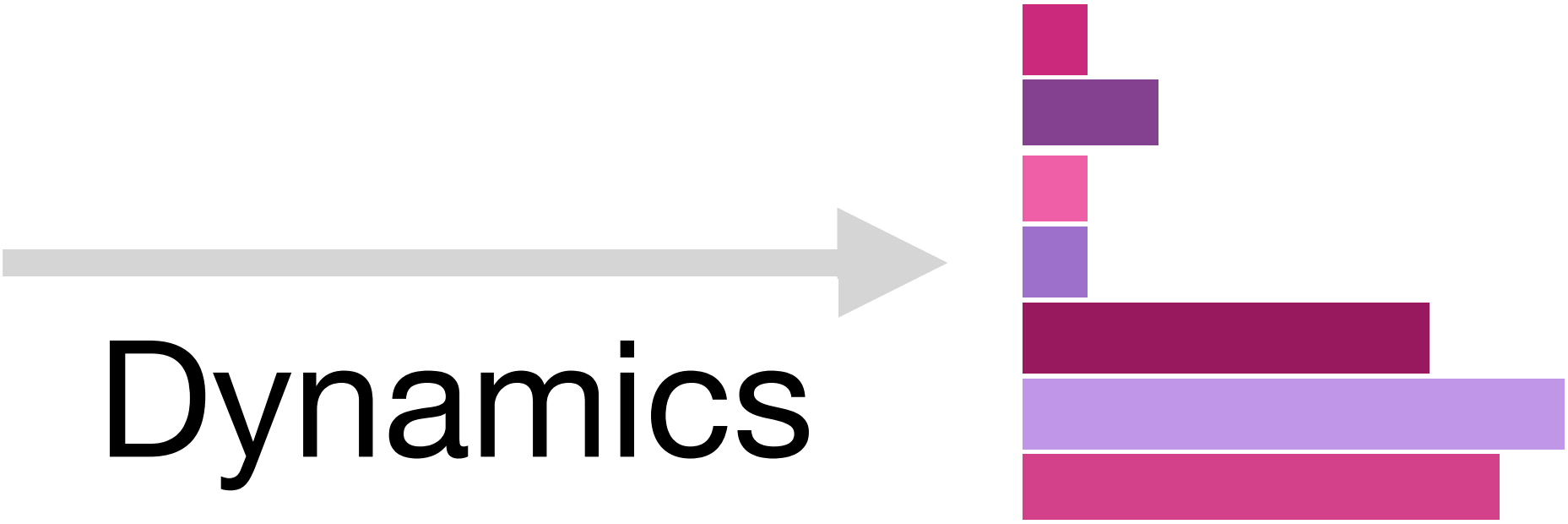
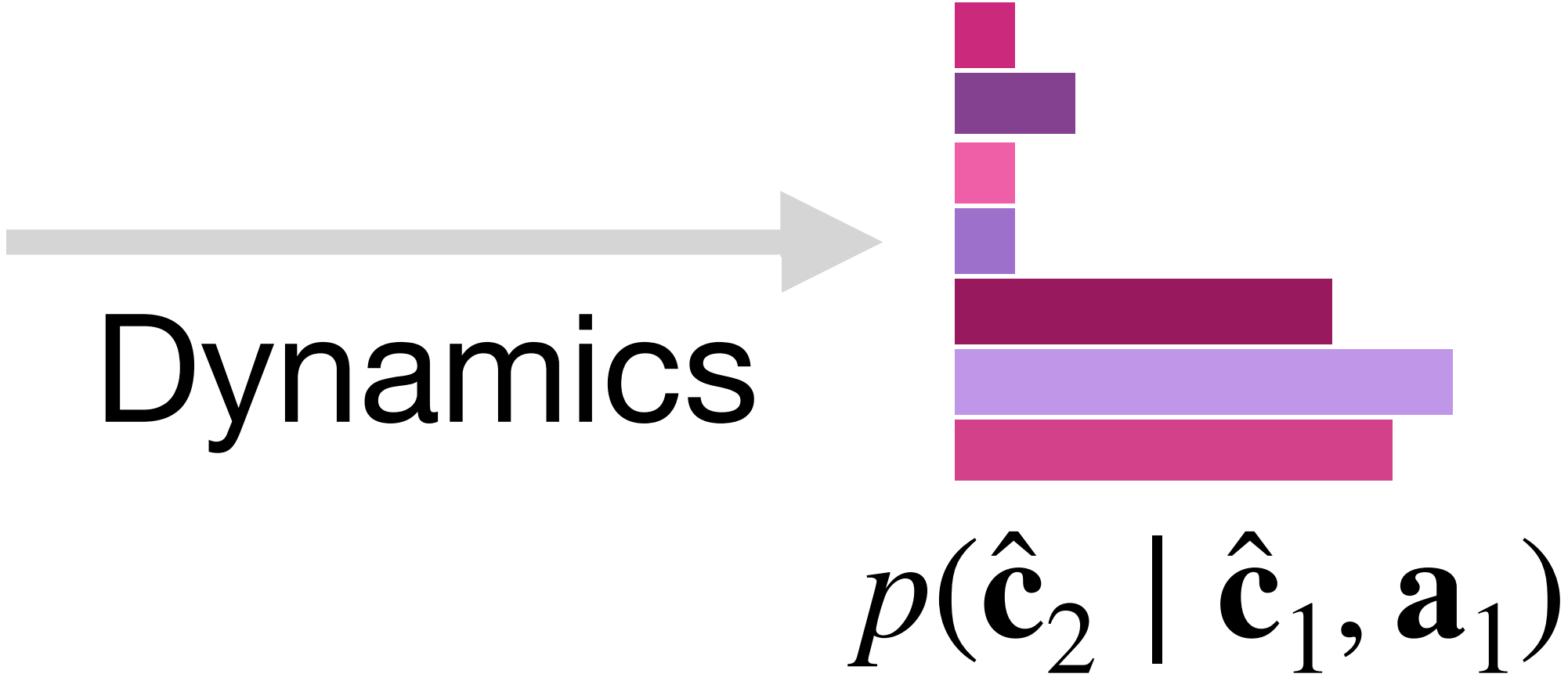

$\hat{r}_H$

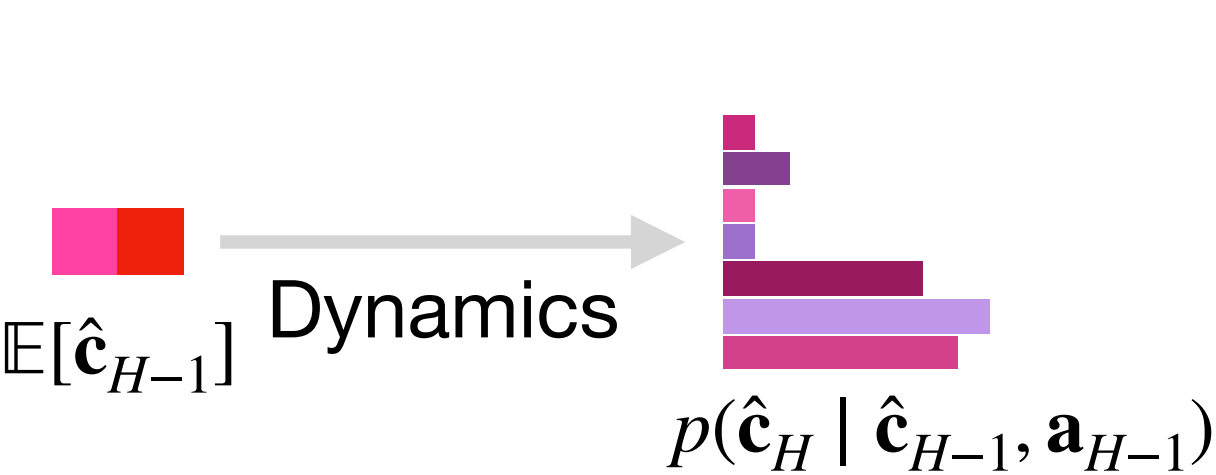

$\hat{r}_2$

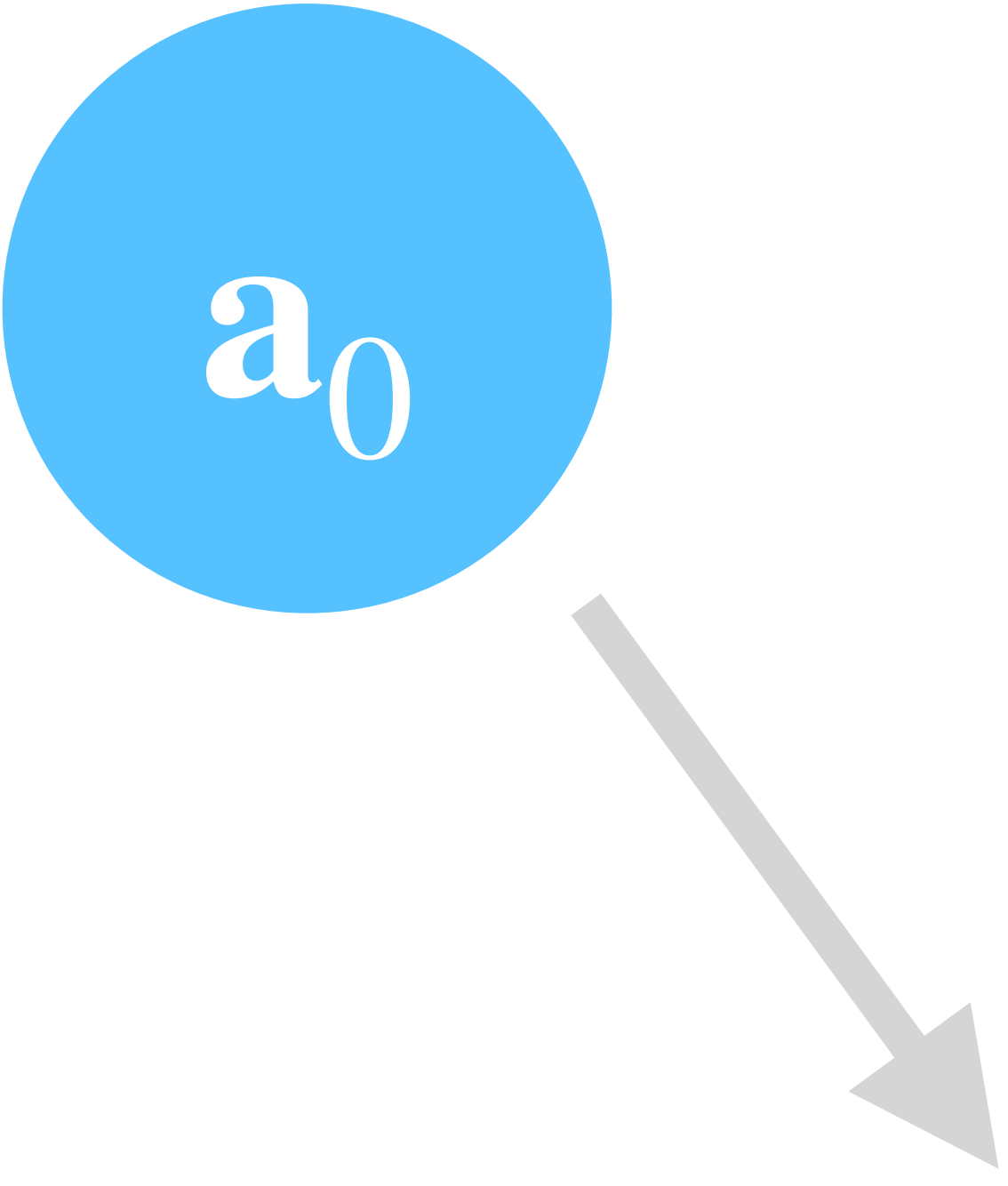

$\hat{r}_1$

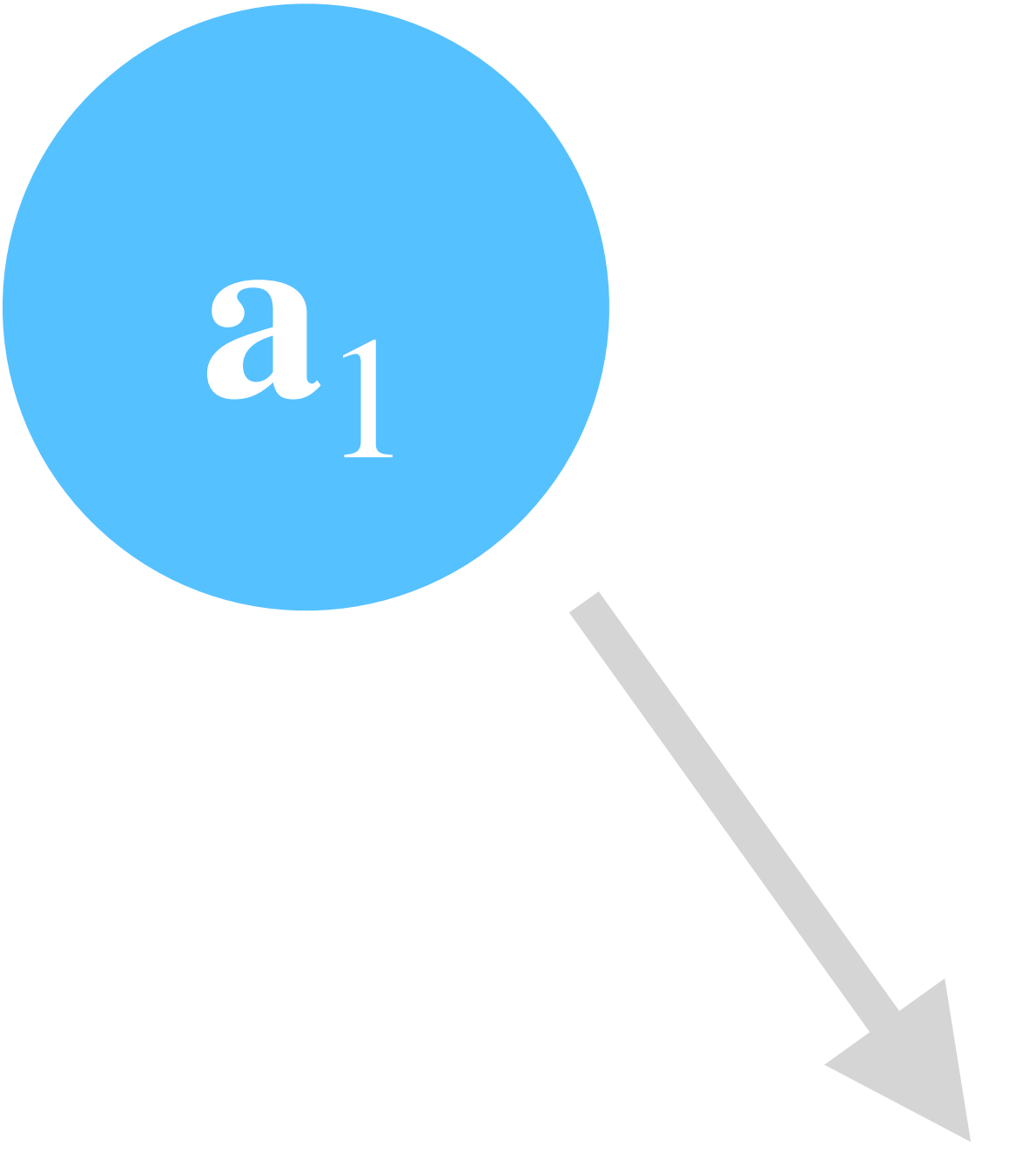

Dynamics
$p(\mathbf{c}_1 \mid \mathbf{c}_0, \mathbf{a}_0)$

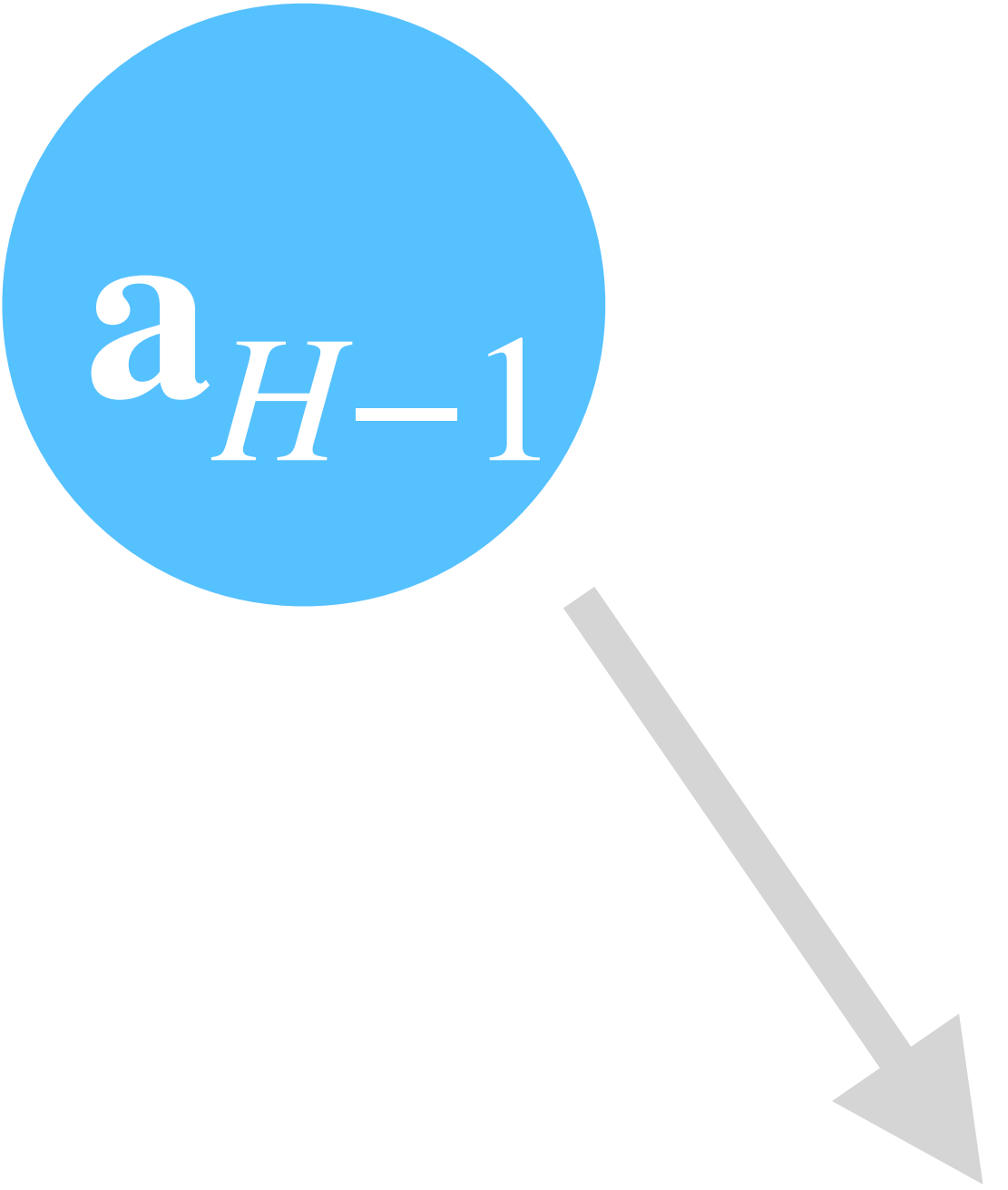

Dynamics
$p(\hat{\mathbf{c}}_2 \mid \hat{\mathbf{c}}_1, \mathbf{a}_1)$

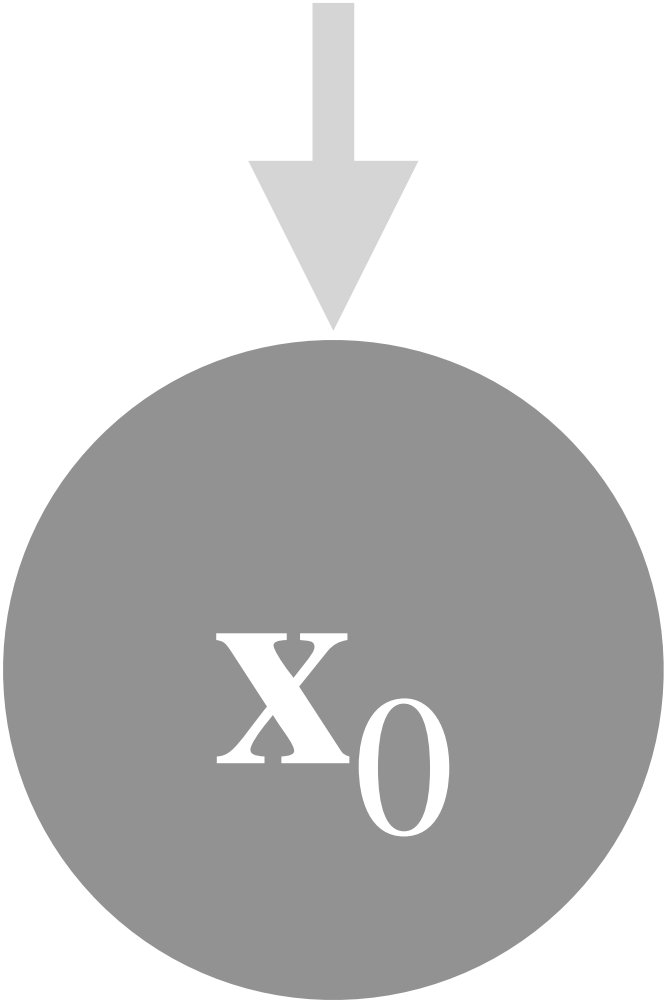
$\mathbf{c}_0$

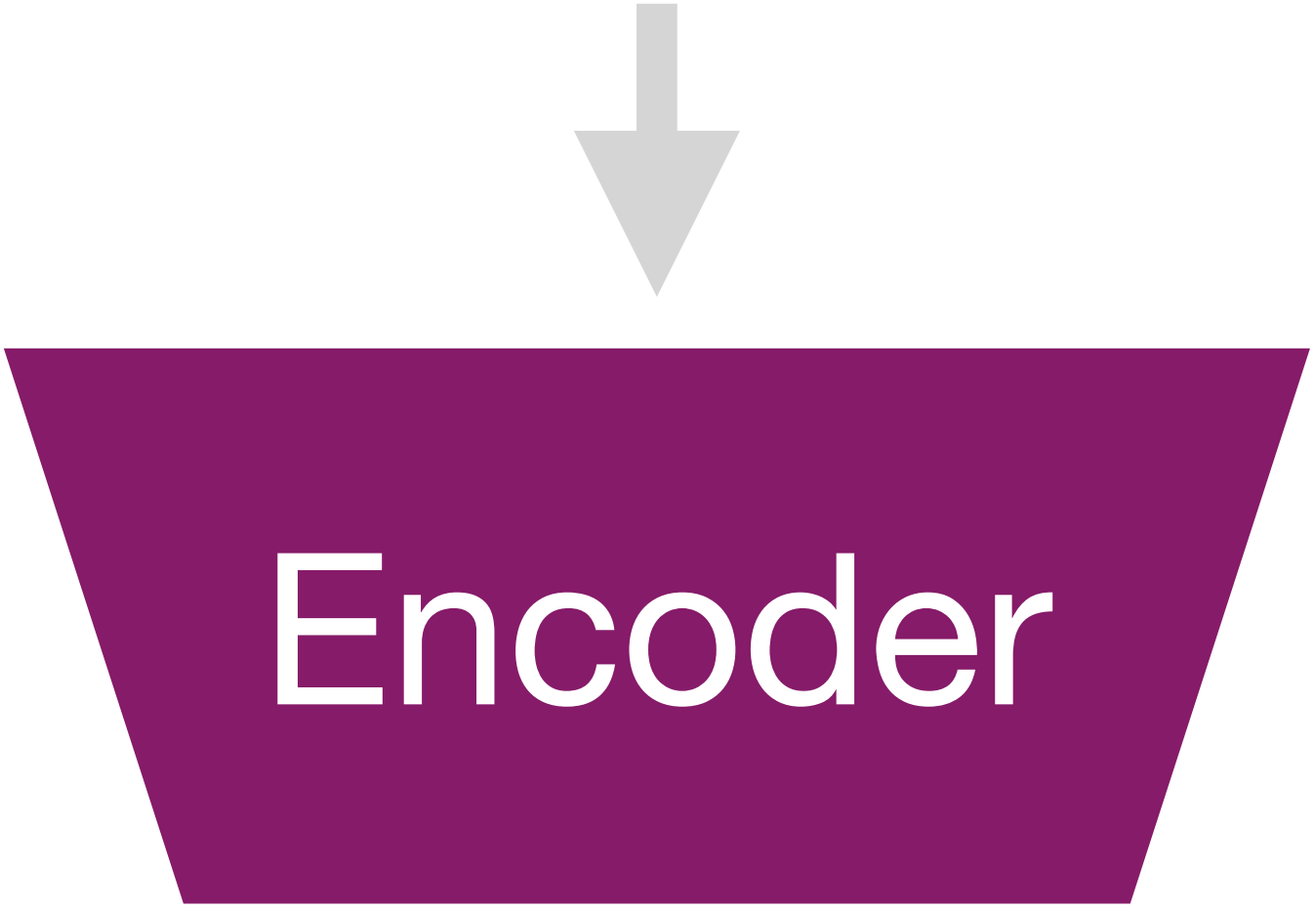

Dynamics
$\mathbb{E}[\hat{\mathbf{c}}_{H-1}]$
$p(\hat{\mathbf{c}}_H \mid \hat{\mathbf{c}}_{H-1}, \mathbf{a}_{H-1})$

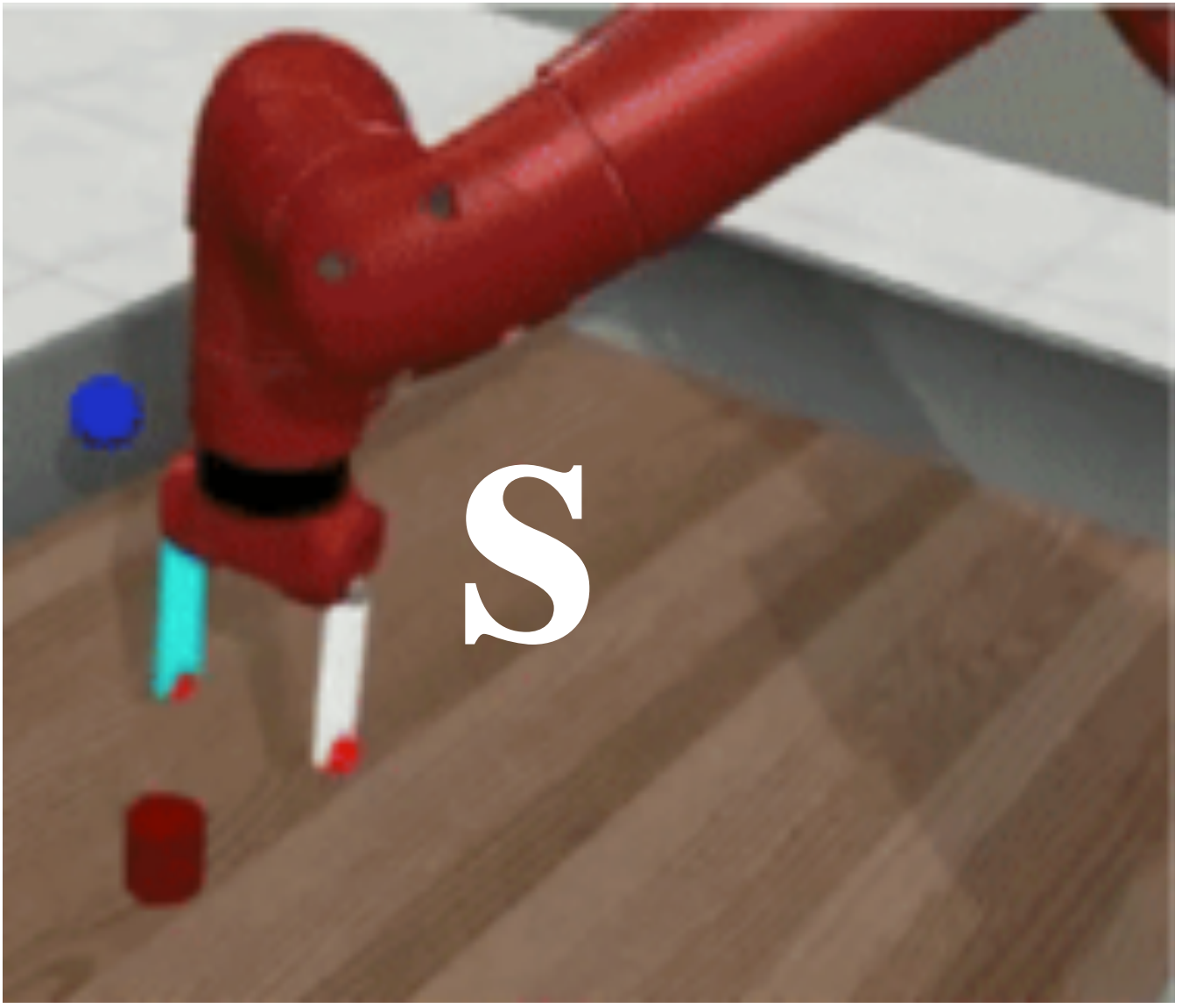

a_0


a_1


$\mathbf{a}_{H-1}$


$\mathbf{x}_0$


Encoder


S

DCWM: Decision-time Planning

$$J(\mathbf{a}_{0:H}, \mathbf{s}) = \gamma^H Q_\psi(\hat{\mathbf{c}}_H, \mathbf{a}_H) + \sum_{h=0}^{H-1} \gamma^h R_\xi(\hat{\mathbf{c}}_h, \mathbf{a}_h)$$

$$\mathbb{E}[\hat{\mathbf{c}}_{h+1}] = \sum_{i=0}^{|\mathscr{C}|} \overbrace{\Pr(\hat{\mathbf{c}}_{h+1} = \mathbf{c}^{(i)} \mid \hat{\mathbf{c}}_h, \mathbf{a}_h)}^{\text{prob. of code } i} \underbrace{\mathbf{c}^{(i)}}_{\text{code}}$$

Reward func.

Bootstrap with action-value

$$\mathbb{E}[\hat{\mathbf{c}}_1]$$

DCWM: Decision-time Planning

Bootstrap with action-value

Reward func.

$$J(\mathbf{a}_{0:H}, \mathbf{s}) = \gamma^H Q_\psi(\hat{\mathbf{c}}_H, \mathbf{a}_H) + \sum_{h=0}^{H-1} \gamma^h R_\xi(\hat{\mathbf{c}}_h, \mathbf{a}_h)$$

$$\mathbb{E}[\hat{\mathbf{c}}_{h+1}] = \sum_{i=0}^{|\mathscr{C}|} \underbrace{\Pr(\hat{\mathbf{c}}_{h+1} = \mathbf{c}^{(i)} \mid \hat{\mathbf{c}}_h, \mathbf{a}_h)}_{\text{prob. of code } i} \underbrace{\mathbf{c}^{(i)}}_{\text{code}}$$

DCWM: Decision-time Planning

Model Predictive Path Integral Control (MPPI)

Iteration 1