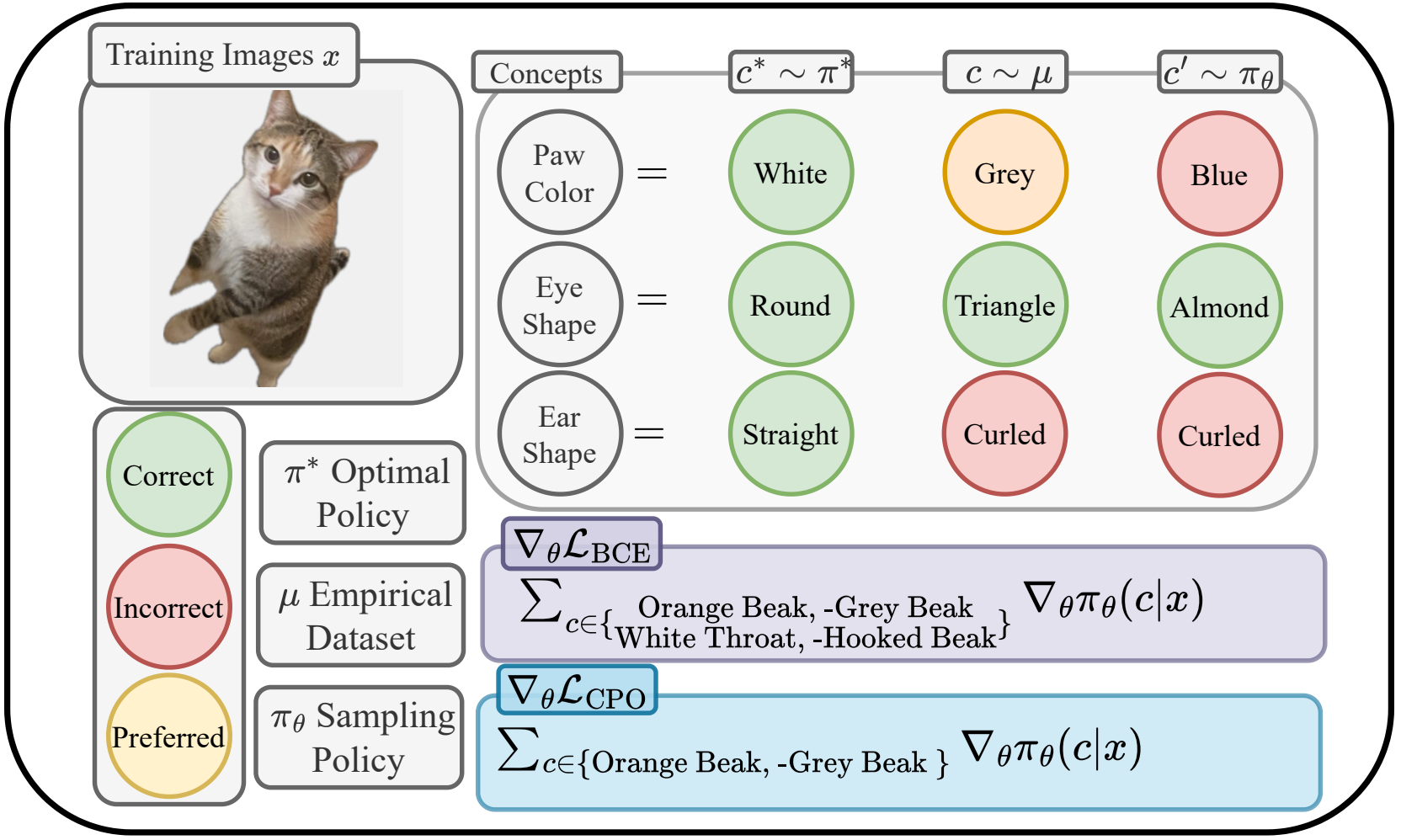
Training Images x
Concepts
$c^* \sim \pi^*$
$c \sim \mu$
$c' \sim \pi_\theta$
Paw Color
=
White
Grey
Blue
Eye Shape
=
Round
Triangle
Almond
Ear Shape
=
Straight
Curled
Curled
Correct
Incorrect
Preferred
π^* Optimal Policy
μ Empirical Dataset
π_θ Sampling Policy
$\nabla_\theta \mathcal{L}_{\text{BCE}}$
$\sum_{c \in \{\text{Orange Beak, -Grey Beak}\atop\text{White Throat, -Hooked Beak}\}} \nabla_\theta \pi_\theta(c|x)$
$\nabla_\theta \mathcal{L}_{\text{CPO}}$
$\sum_{c \in \{\text{Orange Beak, -Grey Beak }\}} \nabla_\theta \pi_\theta(c|x)$