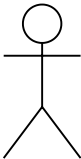


$$a \sim \pi(\cdot | s)$$



Agent

s', r

a

Environment

$$s' \sim p(\cdot | s, a)$$

$$r = r(s, a)$$