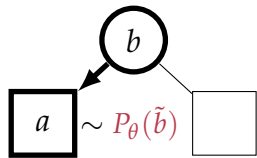
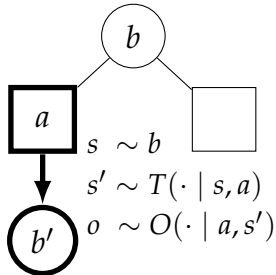


SELECTION



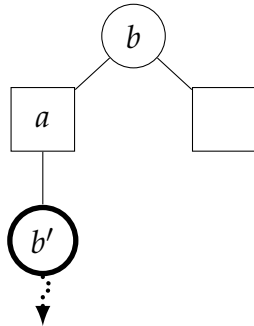
select action that maximizes
 $\bar{Q}(b, a) + c \left(P_{\theta}(\tilde{b}, a) \frac{\sqrt{N(b)}}{1+N(b, a)} \right)$

EXPANSION



$b' \leftarrow \text{UPDATE}(b, a, o)$

SIMULATION



$r + \gamma V_{\theta}(\tilde{b}')$

BACKPROPAGATION

