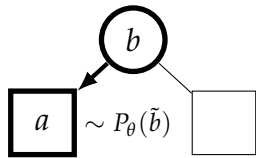
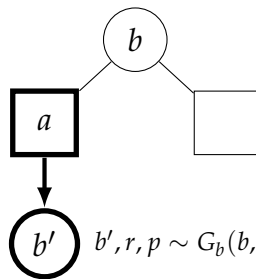


SELECTION



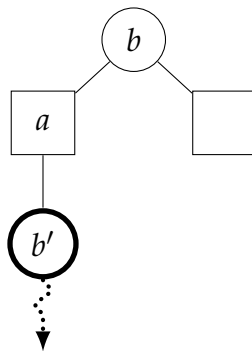
select action that maximizes
 $\bar{Q}(b, a) + c \left(P_{\theta}(\tilde{b}, a) \frac{\sqrt{N(b)}}{1 + N(b, a)} \right)$
 s.t. $F(b, a) \leq \Delta'(b)$

EXPANSION



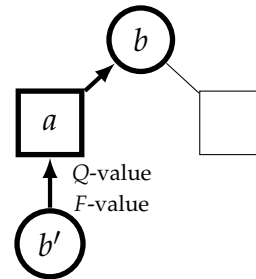
$b', r, p \sim G_b(b, a)$

SIMULATION

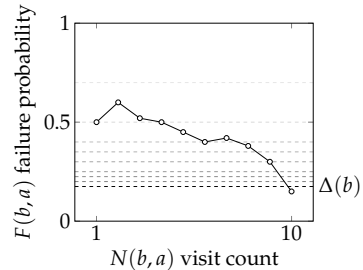


$r + \gamma V_{\theta}(\tilde{b}')$
 $p + \delta(1 - p)F_{\theta}(\tilde{b}')$

BACKPROPAGATION



ADAPTATION



$$\begin{aligned} \text{err} &= \mathbb{1}\{F(b, a) > \Delta(b)\} \\ \Delta(b) &= \Delta(b) + \eta(\text{err} - \Delta_0) \\ \Delta'(b) &= \max\{\Delta_0, \Delta(b)\} \end{aligned}$$