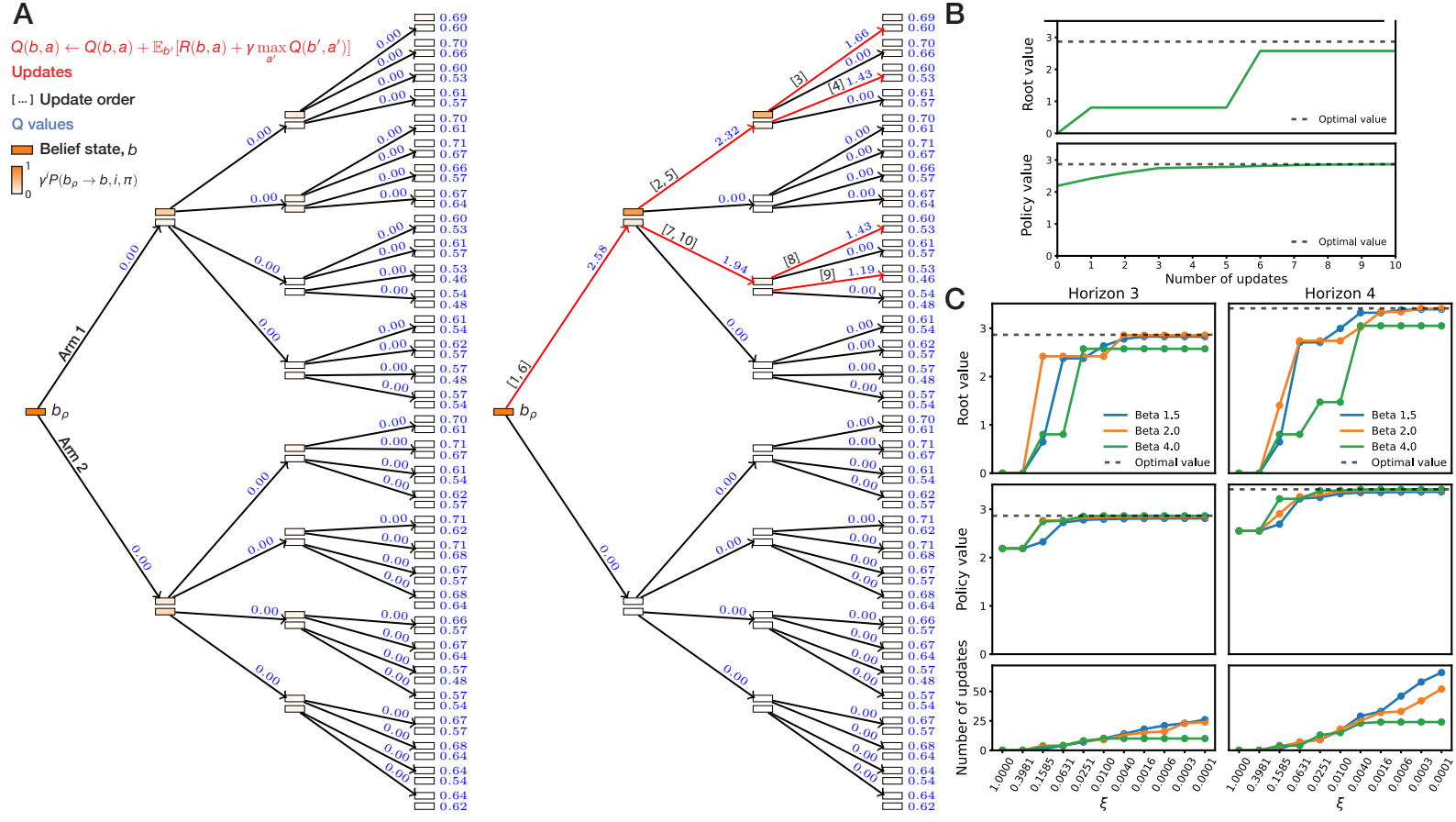
A
$Q(b,a) \leftarrow Q(b,a) + \mathbb{E}_{b'}[R(b,a) + \gamma \max_{a'} Q(b',a')]$
Updates
[...] Update order
Q values
Belief state, b
$\gamma^i P(b_\rho \to b, i, \pi)$
Arm 1
Arm 2
b_ρ
[1, 6]
[2, 5]
[7, 10]
[3]
[4]
[8]
[9]
B
Root value
Policy value
Optimal value
Number of updates
C
Horizon 3
Horizon 4
Beta 1.5
Beta 2.0
Beta 4.0
Optimal value
Root value
Policy value
Number of updates
ξ