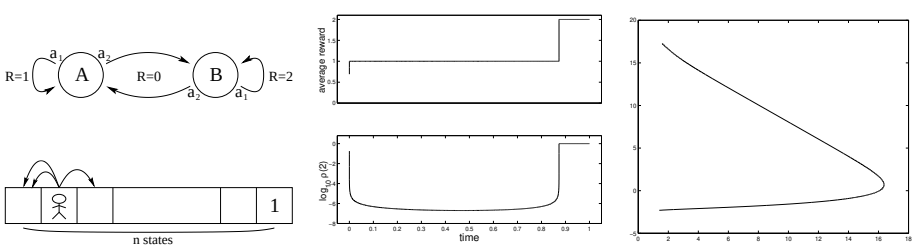
R=1
A
a1
a2
R=0
B
R=2
a2
a1
1
n states
average reward
log10 p(2)
time