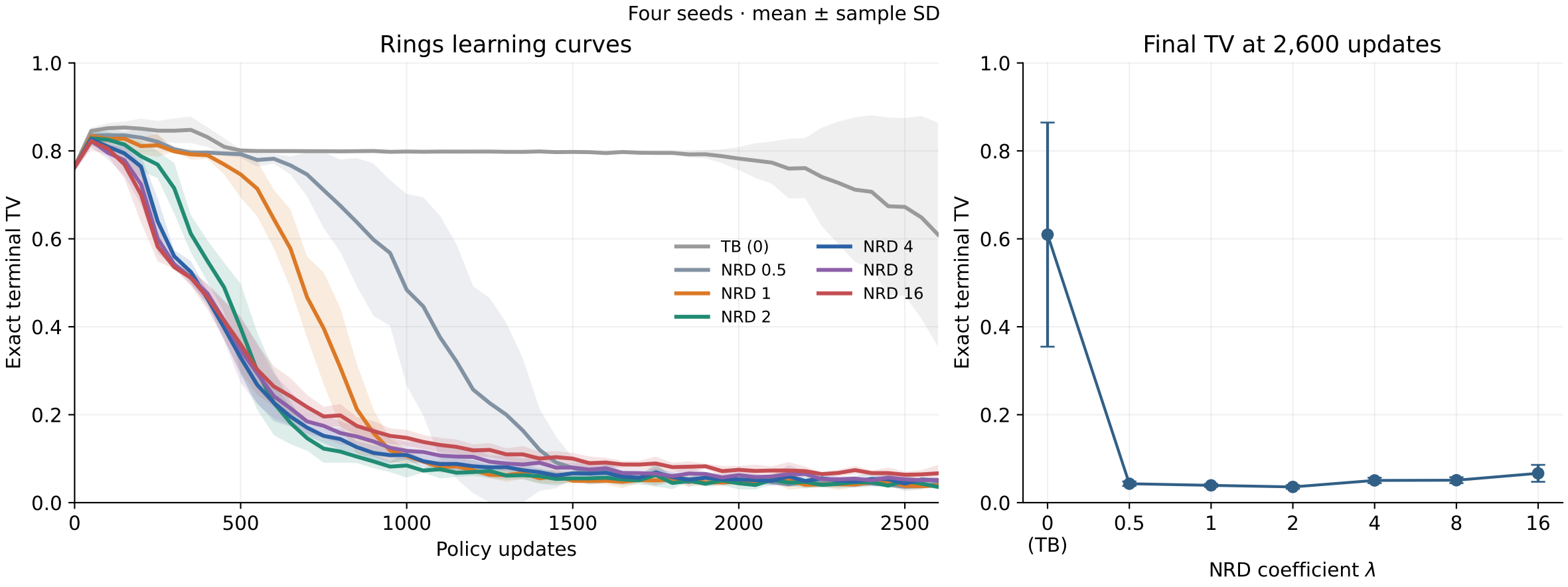

Four seeds · mean ± sample SD
Rings learning curves
Exact terminal TV
Policy updates
TB (0)
NRD 0.5
NRD 1
NRD 2
NRD 4
NRD 8
NRD 16
Final TV at 2,600 updates
NRD coefficient λ
(TB)