Grid World — C-Learning four-input critic logits sampled actor loss no g tau=0.01 — C-Learning four-input critic logits sampled actor loss no g tau=0.01 —— C-Learning four-input critic logits sampled actor loss no g tau=0.01 — C-Learning four-input critic logits sampled actor loss no g tau=0.01 — C-Learning four-input critic logits sampled actor loss no g tau=0.01 — C-Learning four-input critic logits sampled actor loss no g tau=0.01 C-Learning four-input critic logits sampled actor loss no g tau=0.01 —— C-Learning four-input critic logits sampled actor loss no g tau=0.03 - phiT A psi sampled actor loss tau=0.01 lr=3e-5 hs=512 x 4 — phiT\_A\_psi sampled actor loss tau=0.01 lr=3e-5 hs=512 x 4 phiT A psi sampled actor loss tau=0.01 lr=3e-5 hs=512 x 4 phiT A psi sampled actor loss tau=0.01 lr=3e-5 hs=512 x 4 - phiT A psi sampled actor loss tau=0.01 lr=3e-5 hs=512 x 4 phiT A psi sampled actor loss tau=0.01 lr=3e-5 hs=512 x 4 - phiT A psi sampled actor loss tau=0.01 lr=3e-5 hs=512 x 4 phiT A psi sampled actor loss tau=0.01 lr=3e-5 hs=512 x 4 phiT A psi sampled actor loss tau=0.01 lr=3e-5 hs=512 x 4 trial 2 phiT A psi sampled actor loss tau=0.01 lr=3e-5 hs=512 x 4 trial 2 phiT A psi sampled actor loss tau=0.01 lr=3e-5 hs=512 x 4 trial 2 - phiT A psi sampled actor loss tau=0.01 lr=3e-5 hs=512 x 4 trial 2 phiT A psi sampled actor loss tau=0.01 lr=3e-5 hs=512 x 4 trial 2 phiT A psi sampled actor loss tau=0.01 lr=3e-5 hs=512 x 4 trial 2 phiT A psi sampled actor loss tau=0.01 lr=3e-5 hs=512 x 4 trial 2 phiT\_A\_psi sampled actor loss tau=0.01 lr=3e-5 hs=512 x 4 w\_exp phiT A psi sampled actor loss tau=0.01 lr=3e-5 hs=512 x 4 w exp phiT A psi sampled actor loss tau=0.01 lr=3e-5 hs=512 x 4 w exp phiT A psi sampled actor loss tau=0.01 lr=3e-5 hs=512 x 4 w exp - phiT A psi sampled actor loss tau=0.01 lr=3e-5 hs=512 x 4 w exp phiT A psi sampled actor loss tau=0.01 lr=3e-5 hs=512 x 4 w exp phiT A psi sampled actor loss tau=0.01 lr=3e-5 hs=512 x 4 w exp phiT A psi sampled actor loss tau=0.05 lr=3e-5 hs=512 x 4 w exp — phiT A psi sampled actor loss tau=0.05 lr=3e-5 hs=512 x 4 w exp phiT A psi sampled actor loss tau=0.05 lr=3e-5 hs=512 x 4 w exp phiT A psi sampled actor loss tau=0.05 lr=3e-5 hs=512 x 4 w exp - phiT A psi sampled actor loss tau=0.05 lr=3e-5 hs=512 x 4 w exp — phiT A psi sampled actor loss tau=0.05 lr=3e-5 hs=512 x 4 w exp phiT A psi sampled actor loss tau=0.05 lr=3e-5 hs=512 x 4 w exp inner product sampled actor loss tau=0.01 lr=1e-4 hs=512 x 4 inner product sampled actor loss tau=0.01 lr=1e-4 hs=512 x 4 inner product sampled actor loss tau=0.01 lr=1e-4 hs=512 x 4 inner product sampled actor loss tau=0.01 lr=1e-4 hs=512 x 4 inner product sampled actor loss tau=0.01 lr=1e-4 hs=512 x 4 inner product sampled actor loss tau=0.01 lr=1e-4 hs=512 x 4 inner product sampled actor loss tau=0.01 lr=1e-4 hs=512 x 4 inner product sampled actor loss tau=0.05 lr=3e-5 hs=512 x 4 inner product sampled actor loss tau=0.05 lr=3e-5 hs=512 x 4 inner product sampled actor loss tau=0.05 lr=3e-5 hs=512 x 4 inner product sampled actor loss tau=0.05 lr=3e-5 hs=512 x 4 inner product sampled actor loss tau=0.05 lr=3e-5 hs=512 x 4 inner product sampled actor loss tau=0.05 lr=3e-5 hs=512 x 4 inner product sampled actor loss tau=0.05 lr=3e-5 hs=512 x 4 AT phi psi sampled actor loss tau=0.01 lr=3e-5 hs=512 x 4 AT phi psi sampled actor loss tau=0.01 lr=3e-5 hs=512 x 4 - AT phi psi sampled actor loss tau= $0.01 \text{ lr}=3e-5 \text{ hs}=512 \times 4$ AT phi psi sampled actor loss tau=0.01 lr=3e-5 hs=512 x 4 - AT phi psi sampled actor loss tau=0.01 lr=3e-5 hs=512 x 4 AT phi psi sampled actor loss tau=0.01 lr=3e-5 hs=512 x 4 AT phi psi sampled actor loss tau=0.01 lr=3e-5 hs=512 x 4 AT\_phi\_psi sampled actor loss tau=0.05 lr=3e-5 hs=512 x 4 w exp AT phi psi sampled actor loss tau=0.05 lr=3e-5 hs=512 x 4 w exp AT phi psi sampled actor loss tau=0.05 lr=3e-5 hs=512 x 4 w exp - AT\_phi\_psi sampled actor loss tau=0.05 lr=3e-5 hs=512 x 4 w exp AT phi psi sampled actor loss tau=0.05 lr=3e-5 hs=512 x 4 w exp - AT phi psi sampled actor loss tau=0.05 lr=3e-5 hs=512 x 4 w exp AT phi psi sampled actor loss tau=0.05 lr=3e-5 hs=512 x 4 w exp

