Greedy with optimistic initial values and recency-weighted reward ave Per-step reward (averaged over 2000 run(s)) 1.50 Cumulative reward (averaged over 2000 ru 1200 1.25 1000 1.00 800 0.75 600 0.50 400 0.25 200 0.00 epsilon-greedy (e=0.00) 0 Time step 0.9 epsilon-greedy (e=0.00) 0.8 % optimal action selected 0.7 0.6 0.5 0.4 0.3 0.2 0.1 200 0 400 600 800 1000 Time step