Greedy with optimistic initial values and recency-weighted reward ave Per-step reward (averaged over 2000 runs) 1.50 1200 1000 800 800 400 400 pakendari dayi etti mili 0 1.25 1.00 0.75 0.50 0.25 - 200 0.00 epsilon-greedy (e=0.00) Time step 0.9 epsilon-greedy (e=0.00) 8.0 % 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