episode 2 steps 500 Total reward explorative Total steps explorative 400 1.04 1.02 300 1.00 200 0.98 100 0.96 $^{0.00}$ To θ a $^{1.00}$ To θ a $^{1.00}$ 0.00 Total steps greedly 1.04 112 1.02 110 1.00 108 106 0.98 104 -0.96 102 -0.050-0.025 0.000 0.025 0.050 -0.050-0.025 0.000 0.025 0.050 **Episode Episode**