Action probabilities after 5000 episodes