Left action probability for Q-Learning and Double Q-Learning