Q-Learning — Policy Heatmap (b=7)