T-shaped maze with optimal steps of 16

brain
— State-action visited Memory
— Action-state Memory
— Tabular Q-learning with no memory