Monte Carlo Tree Search

1. Selection:
   Follow tree policy (e.g. $\epsilon$-greedy, UCB, etc.) until reaching a node w/ unexplored actions.

2. Expansion:
   Add node for a random unexplored action(s).

3. Simulation:
   Collect a rollout until termination*
   → we use random policy here.

4. Backup:
   Update values all the way back to root
   → we use average value backups, which is on-policy for random rollout policy.

* ... or can limit horizon and estimate value.
1. Selection:
   Follow tree policy (e.g. ε-greedy, UCB, etc.) until reaching a node w/ unexplored actions

2. Expansion:
   Add node for a random unexplored action(s)

3. Simulation:
   Collect a rollout until termination

4. Backup:
   Update values all the way back to root
Monte Carlo Tree Search

1. Selection:
   Follow tree policy (e.g. E-greedy, UCB, etc) until reaching a node w/ unexplored actions

2. Expansion:
   Add node for a random unexplored action(s)

3. Simulation:
   Collect a rollout until termination

4. Backup
   Update values all the way back to root
Monte Carlo Tree Search

1. Selection:
   Follow tree policy (e.g. E-greedy, UCB, etc) until reaching a node w/ unexplored actions

2. Expansion:
   Add node for a random unexplored action(s)

3. Simulation:
   Collect a rollout until termination

4. Backup
   Update values all the way back to root
Monte Carlo Tree Search

1. **Selection**: Follow tree policy (e.g., E-greedy, UCB, etc) until reaching a node w/ unexplored actions

2. **Expansion**: Add node for a random unexplored action(s)

3. **Simulation**: Collect a rollout until termination

4. **Backup**: Update values all the way back to root
Monte Carlo Tree Search

1. **Selection**:
   Follow tree policy (e.g., E-greedy, UCB, etc.) until reaching a node w/ unexplored actions.

2. **Expansion**:
   Add node for a random unexplored action(s).

3. **Simulation**:
   Collect a rollout until termination.

4. **Backup**
   Update values all the way back to root.
Monte Carlo Tree Search

1. Selection:
   Follow tree policy (e.g. E-greedy, UCB, etc) until reaching a node w/ unexplored actions.

2. Expansion:
   Add node for a random unexplored action(s).

3. Simulation:
   Collect a rollout until termination.

Properties:

✓ Focuses expansion in most promising states.
✓ Caches computations, but minimizes memory use, and ✓ Anytime.
✓ On-the-fly planning for specific decisions.