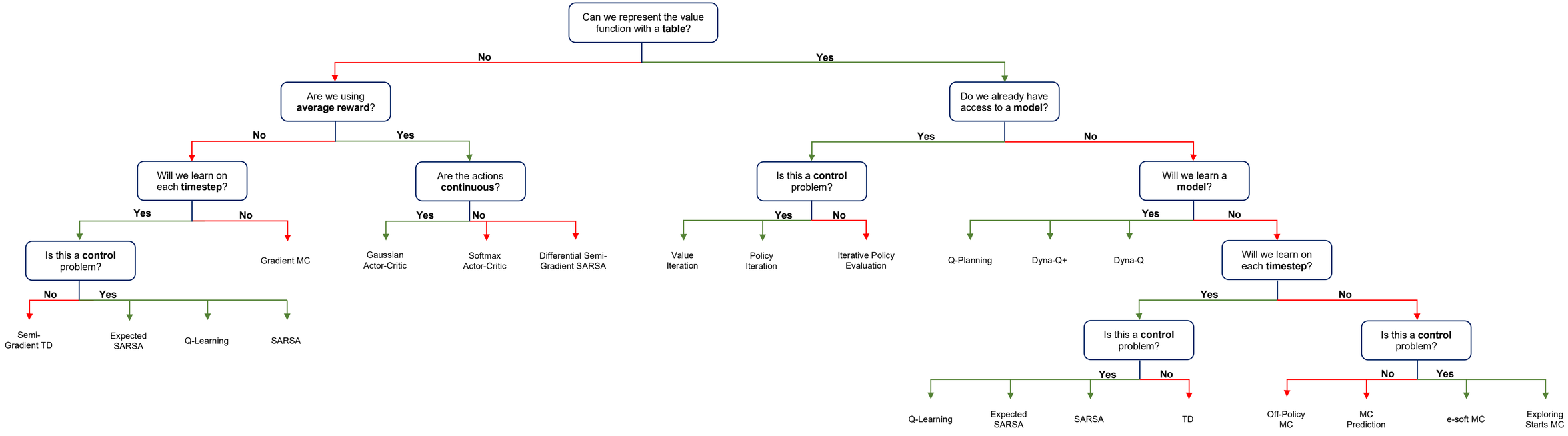
Can we represent the value function with a **table**?
No
Yes
Are we using **average reward**?
Do we already have access to a **model**?
No
Yes
Yes
No
Will we learn on each **timestep**?
Are the actions **continuous**?
Is this a **control** problem?
Will we learn a **model**?
Yes
No
Yes
No
Yes
No
Is this a **control** problem?
Gradient MC
Gaussian Actor-Critic
Softmax Actor-Critic
Differential Semi-Gradient SARSA
Value Iteration
Policy Iteration
Iterative Policy Evaluation
Q-Planning
Dyna-Q+
Dyna-Q
Will we learn on each **timestep**?
No
Yes
Yes
No
Semi-Gradient TD
Expected SARSA
Q-Learning
SARSA
Is this a **control** problem?
Is this a **control** problem?
Yes
No
No
Yes
Q-Learning
Expected SARSA
SARSA
TD
Off-Policy MC
MC Prediction
e-soft MC
Exploring Starts MC