it has to continue to **explore** the range of possibilities by visiting every state-action pair). That said, the agent should always act *somewhat greedily*, towards its goal of maximizing return *as quickly as possible*. This motivated the idea of an ϵ-greedy policy.

We refer to the need to balance these two competing requirements as the **Exploration-Exploitation Dilemma**. One potential solution to this dilemma is implemented by gradually modifying the value of ϵ when constructing ϵ-greedy policies.

Setting the Value of ϵ, in Theory

It makes sense for the agent to begin its interaction with the environment by favoring **exploration** over **exploitation**. After all, when the agent knows relatively little about the environment's dynamics, it should distrust its limited knowledge and **explore**, or try out various strategies for maximizing return. With this in mind, the best starting policy is the equiprobable random policy, as it is equally likely to explore all possible actions from each state. You discovered in the previous quiz that setting $\epsilon = 1$ yields an ϵ-greedy policy that is equivalent to the equiprobable random policy.

At later time steps, it makes sense to favor **exploitation** over **exploration**, where the policy gradually becomes more greedy with respect to the action-value function estimate. After all, the more the agent interacts with the environment, the more it can trust its estimated action-value function. You discovered in the previous quiz that setting $\epsilon = 0$ yields the greedy policy (or, the policy that most favors exploitation over exploration).

Thankfully, this strategy (of initially favoring exploration over exploitation, and then gradually preferring exploitation over exploration) can be demonstrated to be optimal.

Greedy in the Limit with Infinite Exploration (GLIE)

In order to guarantee that MC control converges to the optimal policy π_*, we need to ensure that two conditions are met. We refer to these conditions as **Greedy in the Limit with Infinite Exploration (GLIE)**. In particular, if:

- every state-action pair s, a (for all $s \in \mathcal{S}$ and $a \in \mathcal{A}(s)$) is visited infinitely many