```
Evaluation  Generate an episode $S_0, A_0, R_1, \ldots, S_T$ using $\pi$.
            For $t \leftarrow 0$ to $T-1$:
                If $(S_t, A_t)$ is a first visit (with return $G_t$):
                    $Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha(G_t - Q(S_t, A_t))$
```

Before moving on to the next concept, use the above coding environment to verify the following facts about about how to set the value of α when implementing constant-α MC control.

- You should always set the value for α to a number greater than zero and less than (or equal to) one.
  - If $\alpha = 0$, then the action-value function estimate is never updated by the agent.
  - If $\alpha = 1$, then the final value estimate for each state-action pair is always equal to the last return that was experienced by the agent (after visiting the pair).
- Smaller values for α encourage the agent to consider a longer history of returns when calculating the action-value function estimate. Increasing the value of α ensures that the agent focuses more on the most recently sampled returns.

Note that it is also possible to verify the above facts by slightly rewriting the update step as follows:

$$Q(S_t, A_t) \leftarrow (1-\alpha)Q(S_t, A_t) + \alpha G_t$$

where it is now more obvious that α controls how much the agent trusts the most recent return G_t over the estimate $Q(S_t, A_t)$ constructed by considering all past returns.

IMPORTANT NOTE: It is important to mention that when implementing constant-α MC control, you must be careful to not set the value of α too close to 1. This is because very large values can keep the algorithm from converging to the optimal policy π_*. However, you must also be careful to not set the value of α too low, as this can result in an agent who learns too slowly. The best value of α for your implementation will