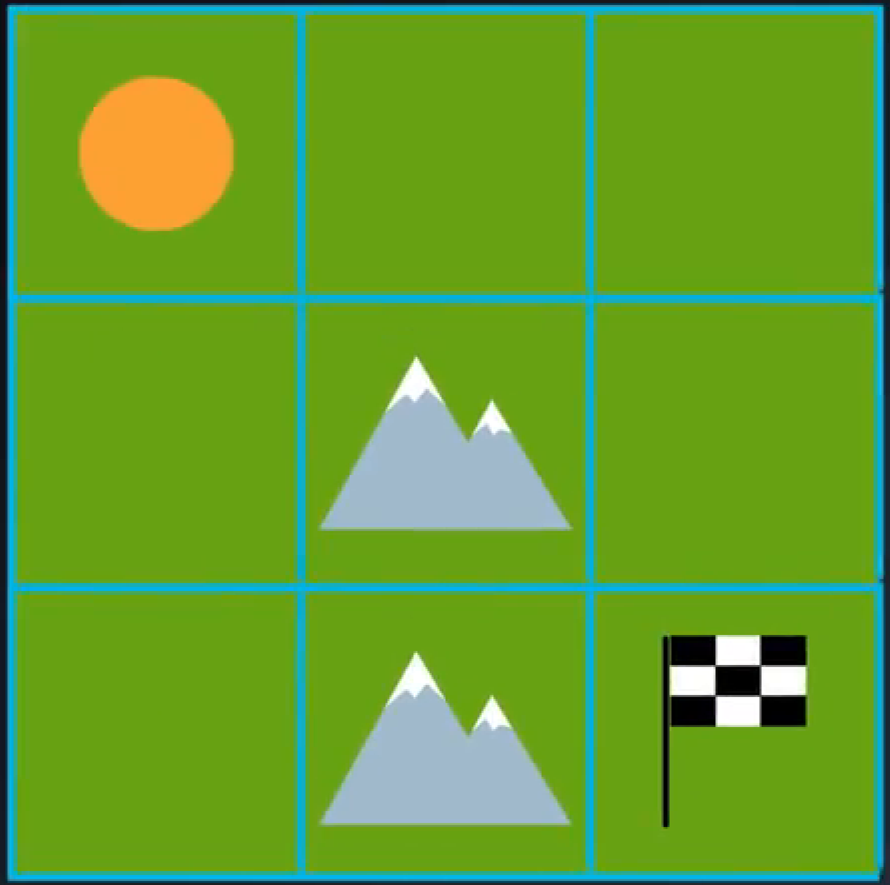
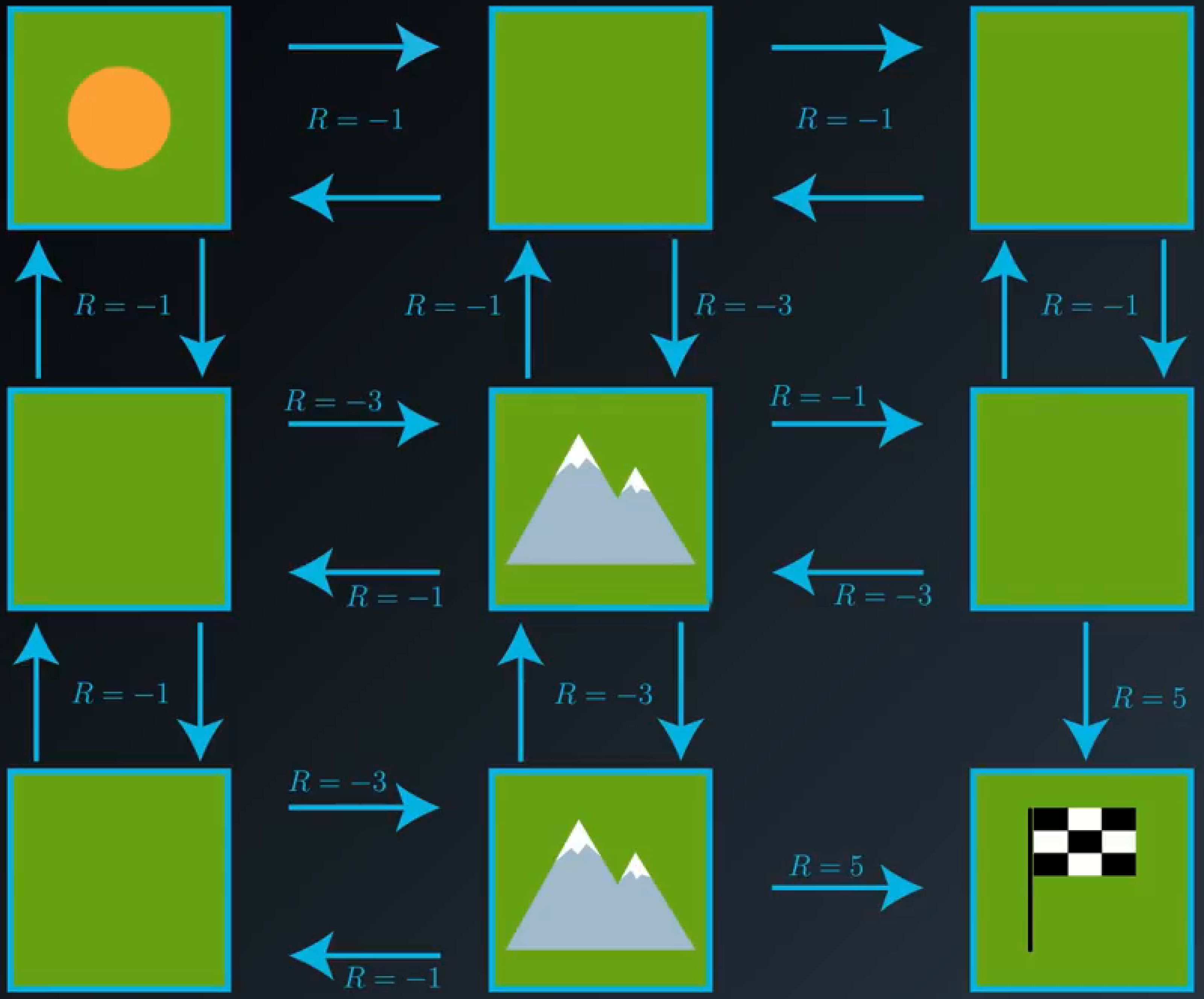
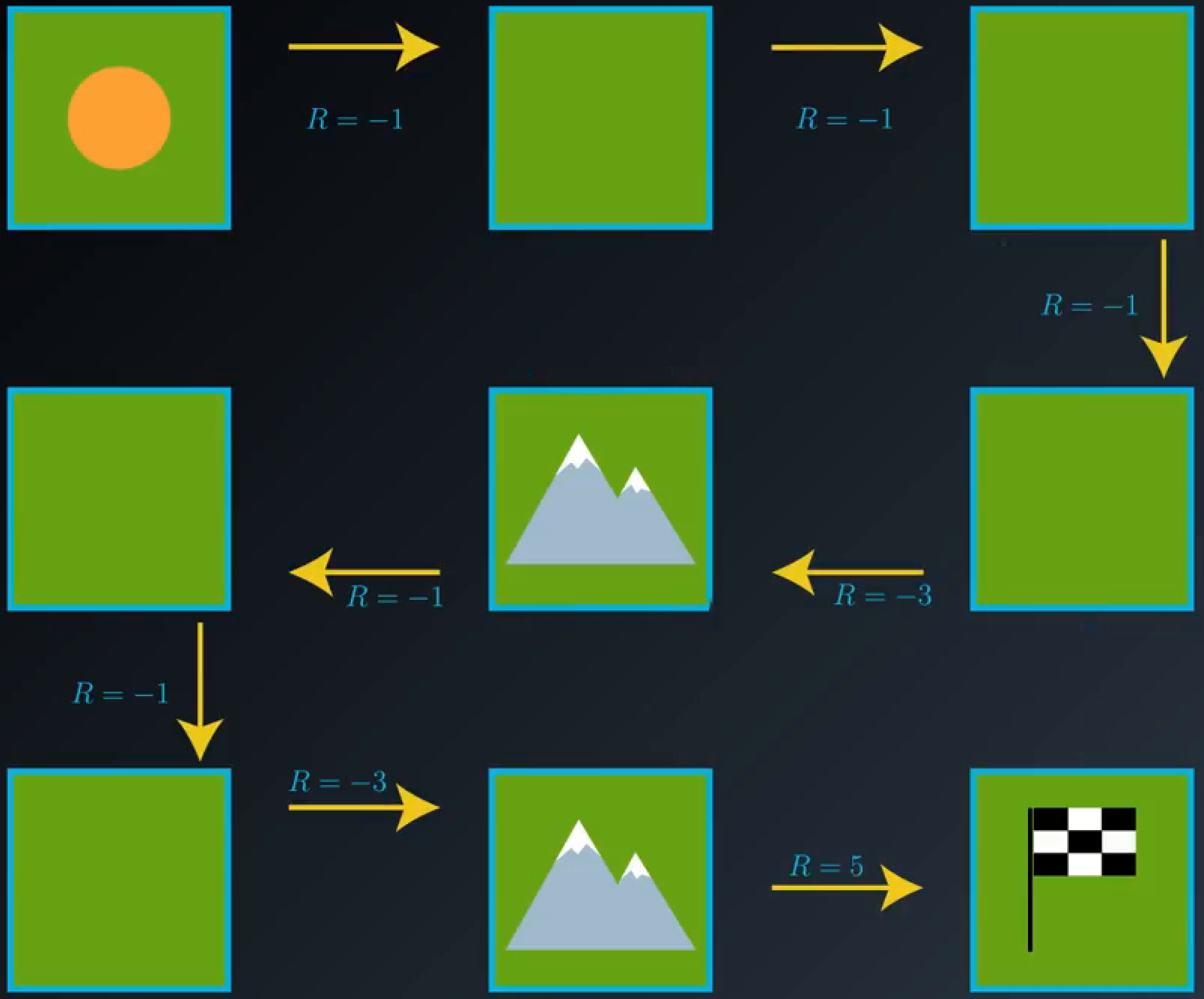
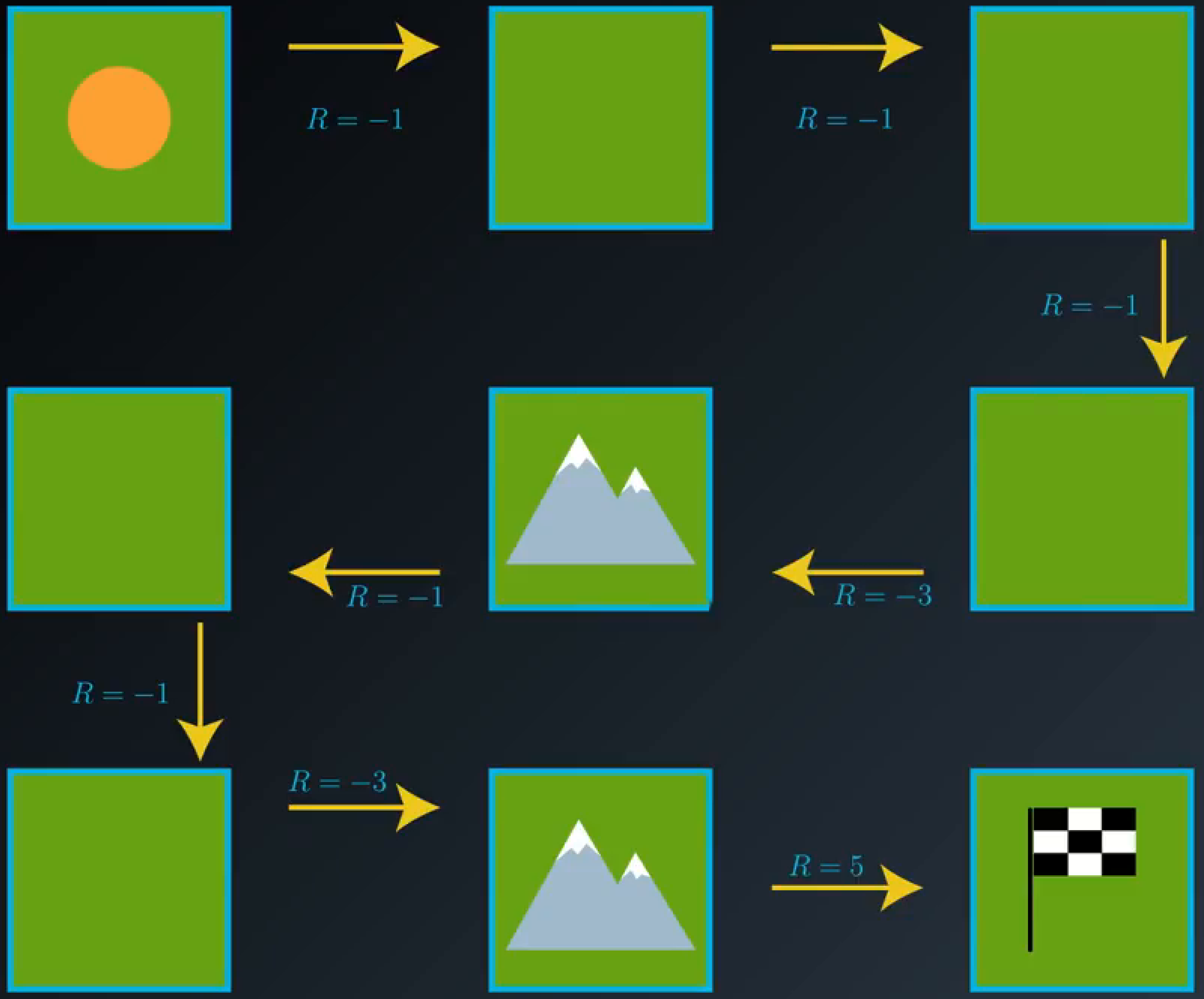
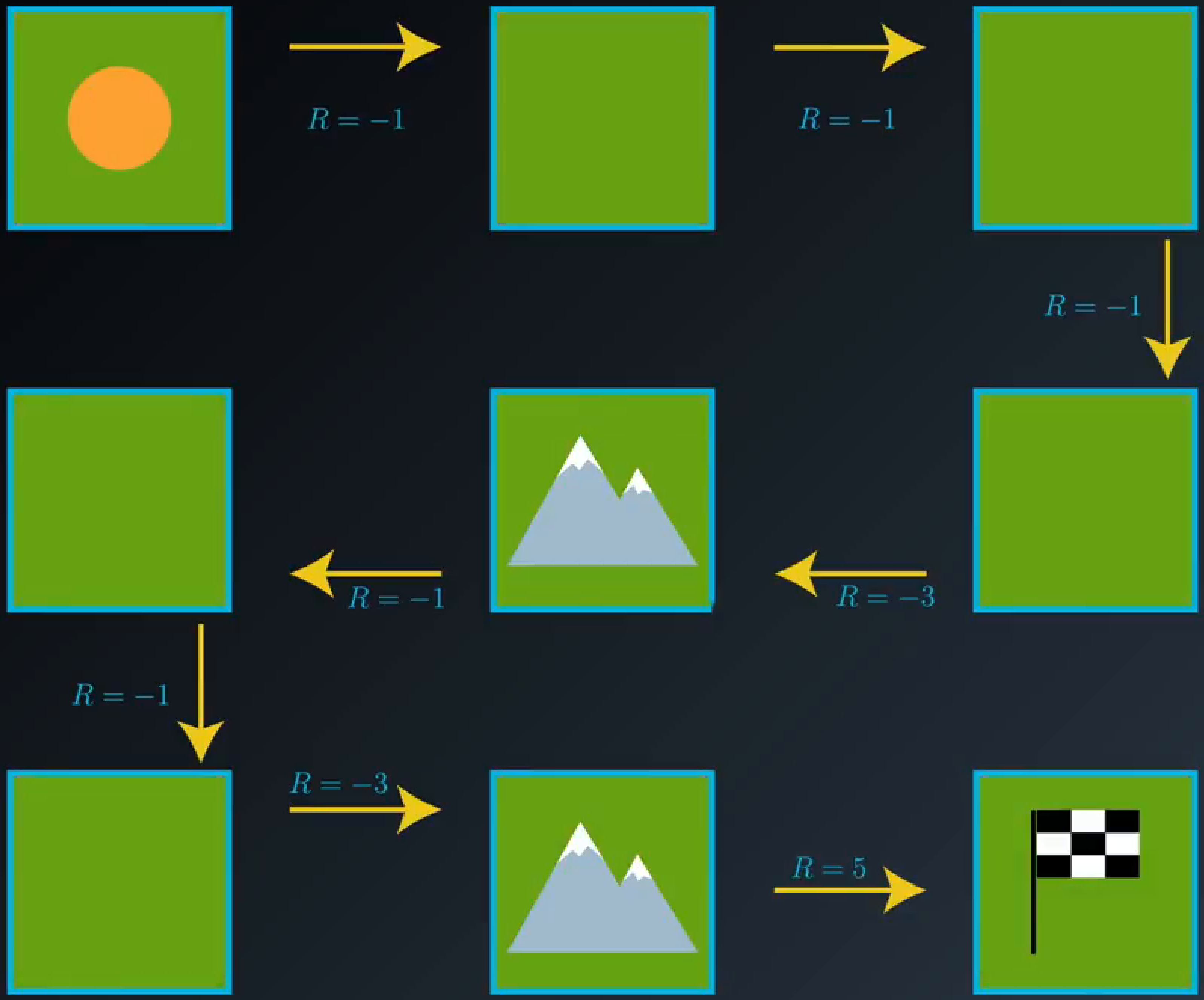


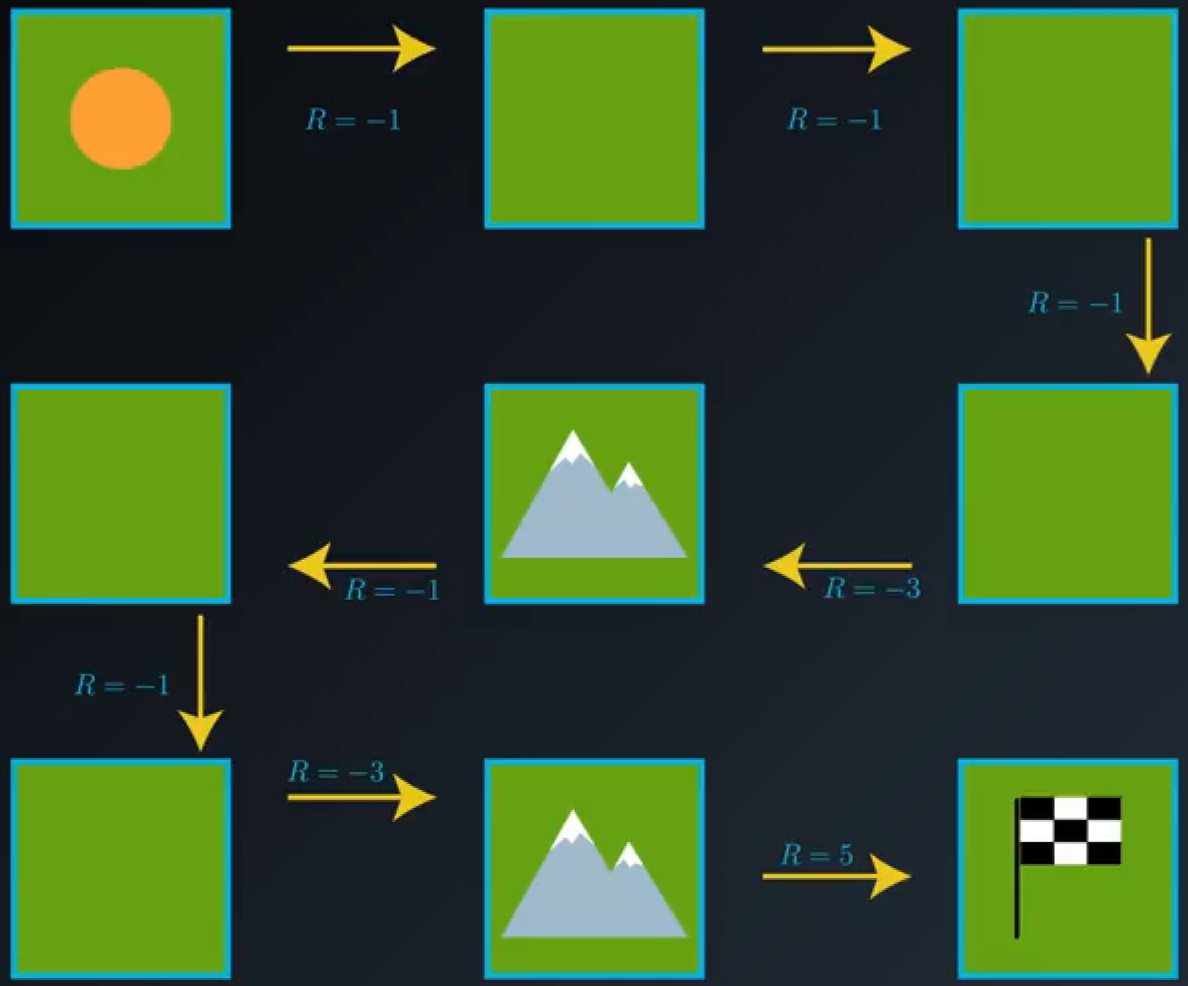

$R=-1$
$R=-1$
$R=-1$
$R=-1$
$R=-3$
$R=-1$
$R=-3$
$R=-1$
$R=-1$
$R=-3$
$R=-1$
$R=-3$
$R=5$
$R=-3$
$R=5$
$R=-1$


$R=-1$
$R=-1$
$R=-1$
$R=-3$
$R=-1$
$R=-1$
$R=-3$
$R=5$

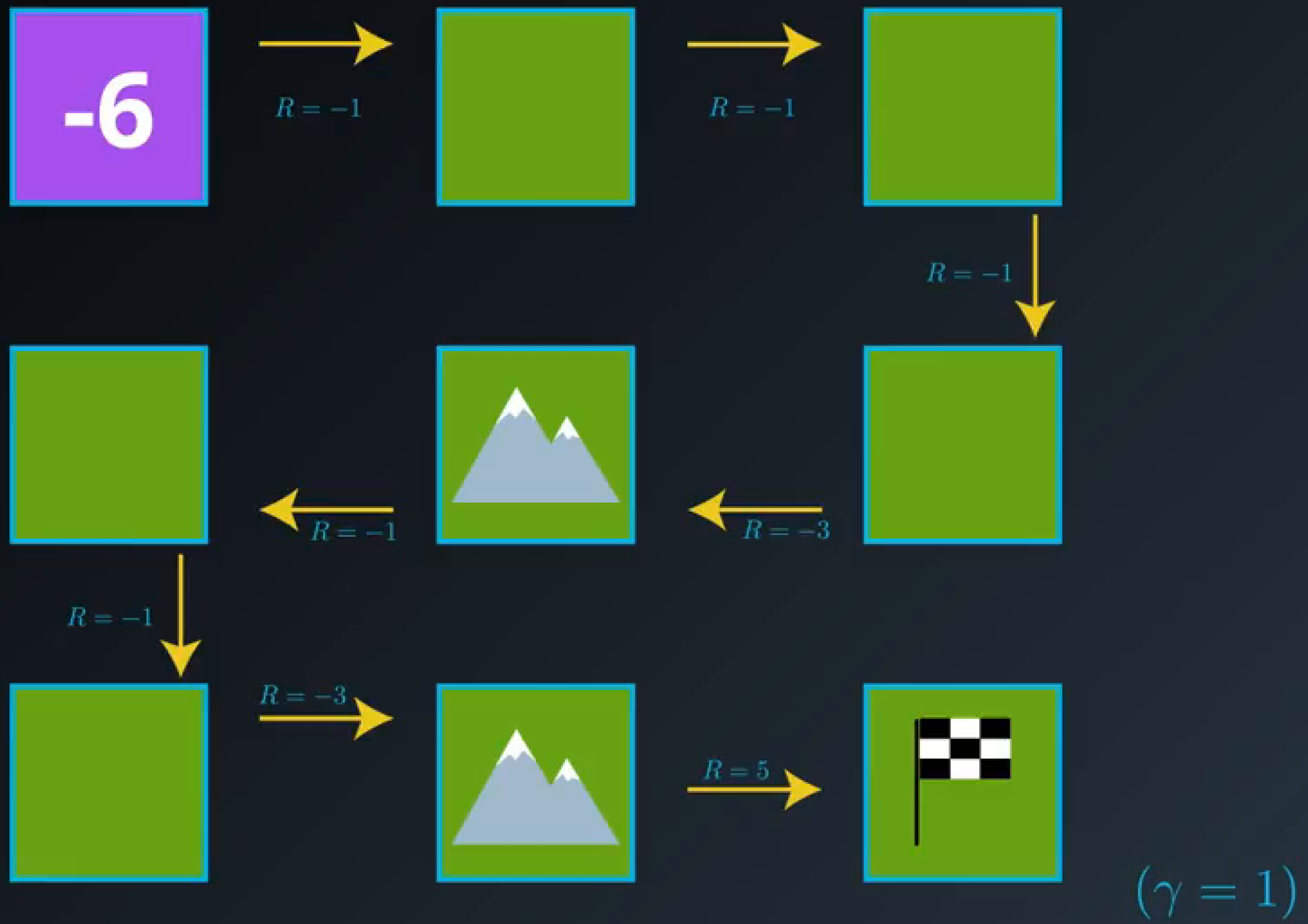


$R=-1$
$R=-1$
$R=-1$
$R=-1$
$R=-3$
$R=-1$
$R=-3$
$R=5$

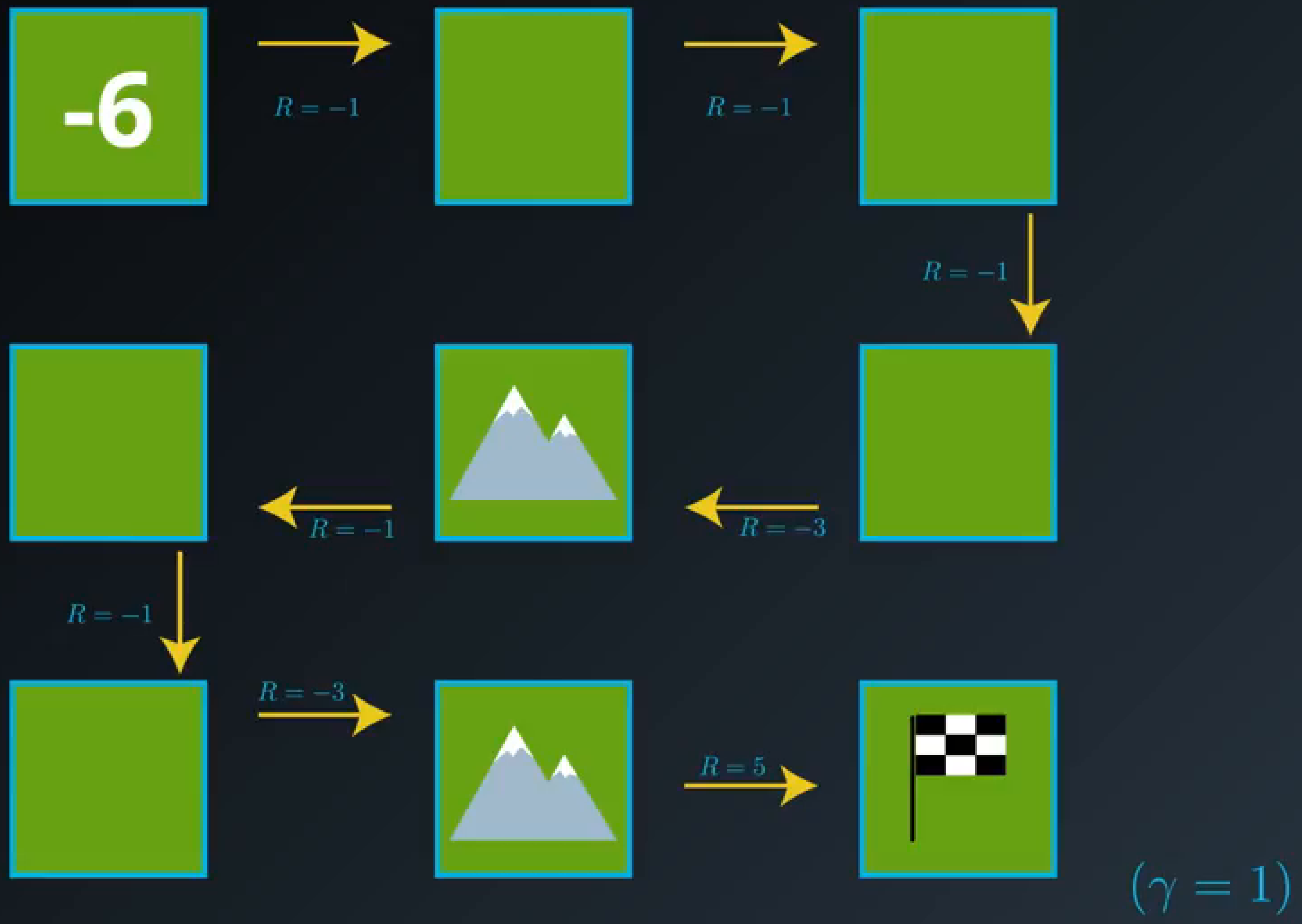

$R=-1$
$R=-1$
$R=-1$
$R=-3$
$R=-1$
$R=-1$
$R=-3$
$R=5$



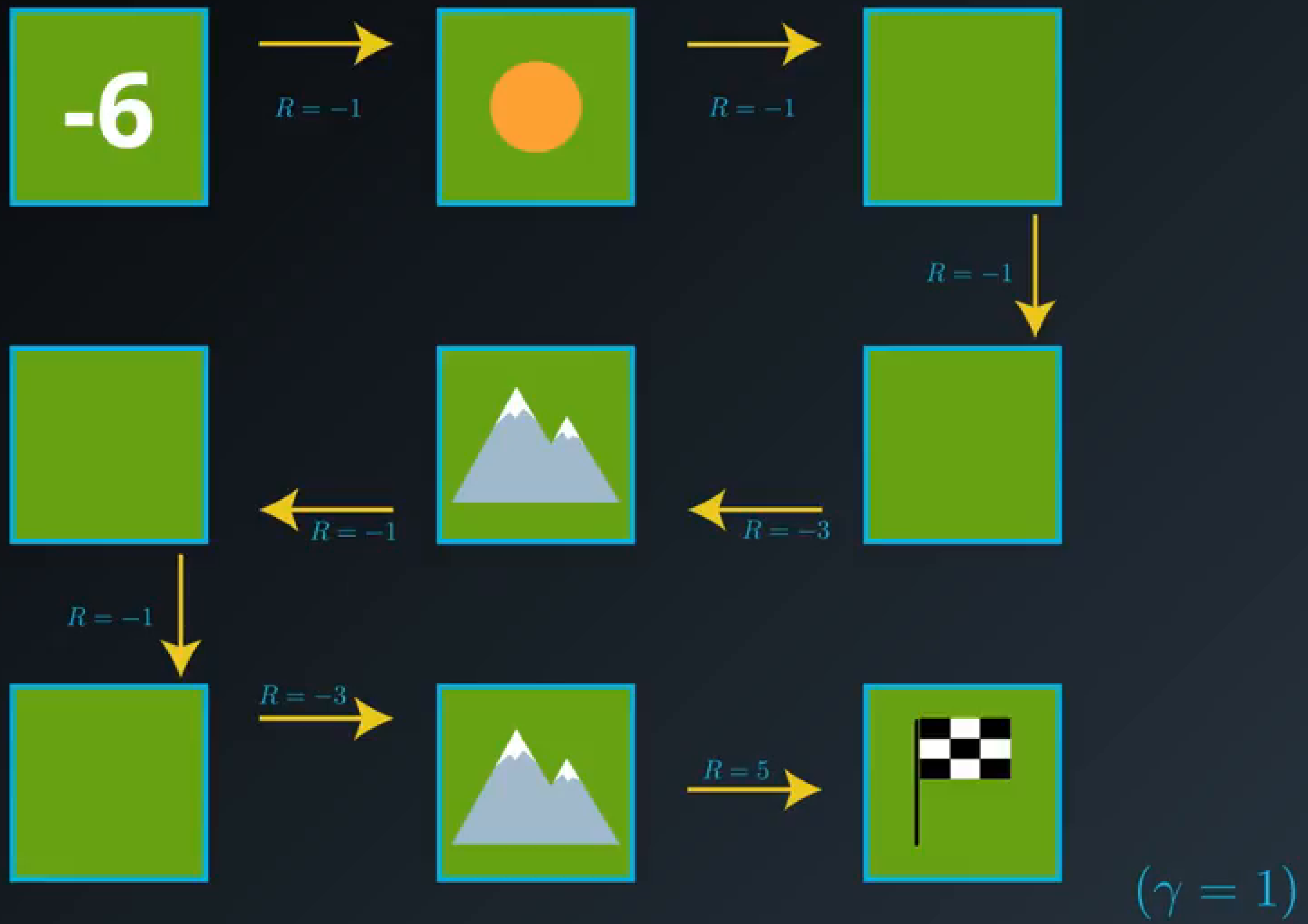
$$(-1) + (-1) + (-1) + (-3) + (-1) + (-1) + (-3) + 5 = -6$$



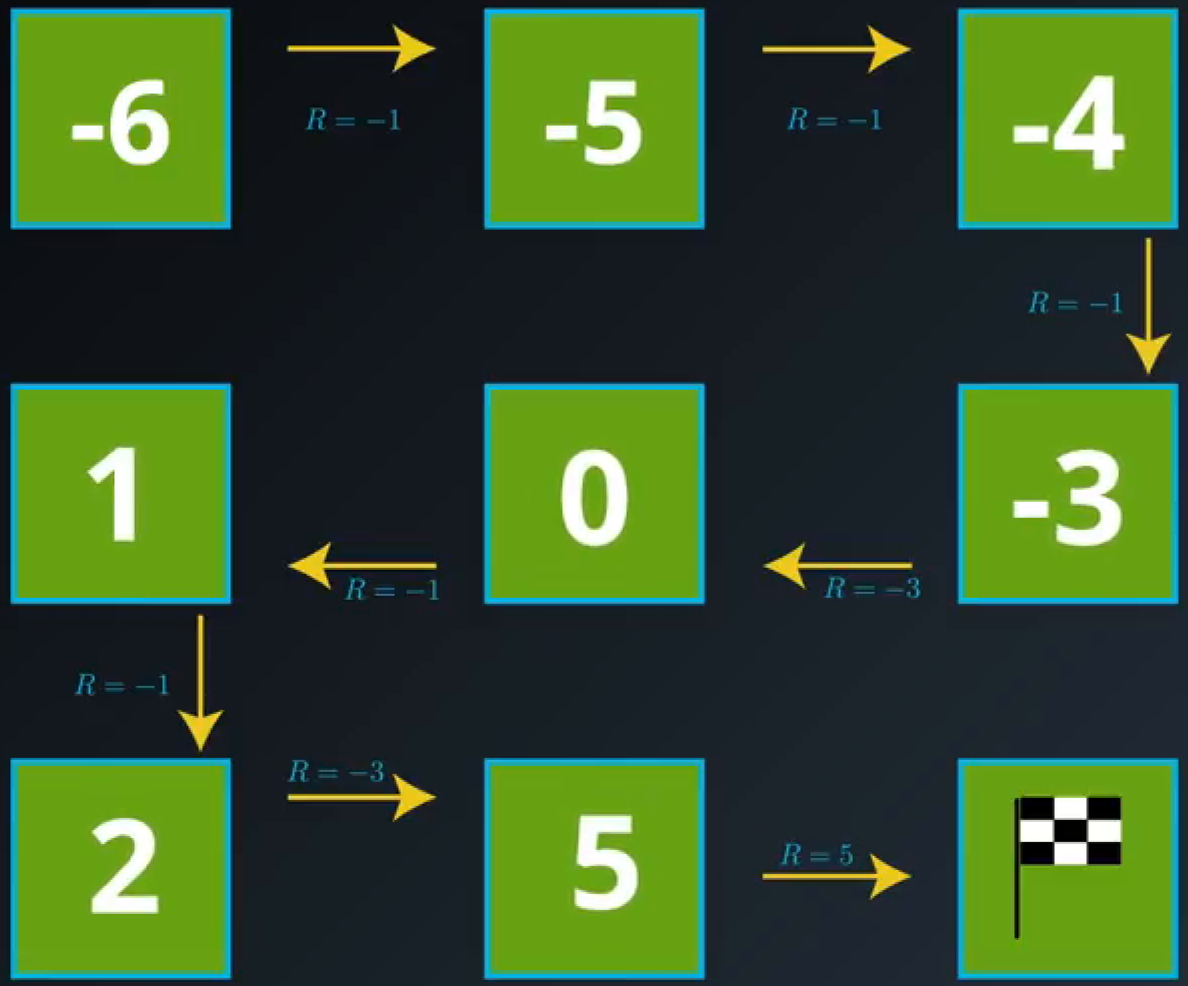
(-1) + (-1) + (-1) + (-3) + (-1) + (-1) + (-3) + 5 = -6



-6
R = −1
R = −1
R = −1
R = −1
R = −3
R = −1
R = −3
R = 5
(γ = 1)



(-1) + (-1) + (-3) + (-1) + (-1) + (-3) + 5 = -5


-6
$R=-1$
-5
$R=-1$
-4
$R=-1$
-3
$R=-3$
0
$R=-1$
1
$R=-1$
2
$R=-3$
5
$R=5$

-6
$R=-1$
-5
$R=-1$
-4
$R=-1$
1
$R=-1$
0
$R=-3$
-3
$R=-1$
2
$R=-3$
5
$R=5$
0

-6
R = −1
-5
R = −1
-4
R = −1
1
0
-3
R = −1
R = −3
R = −1
2
R = −3
5
R = 5
0

-6
$R = -1$
-5
$R = -1$
-4
$R = -1$
1
$R = -1$
0
$R = -3$
-3
$R = -1$
2
$R = -3$
5
$R = 5$
0

-6	-5	-4
1	0	-3
2	5	0

For each state, the state-value function yields the expected return, if the agent started in that state, and then followed the policy for all time steps.

-6	-5	-4
1	0	-3
2	5	0

For each state, the state-value function yields the expected return, if the agent started in that state, and then followed the policy for all time steps.

-6	-5	-4
1	0	-3
2	5	0

Definition

We call v_π the state-value function for policy π

-6	-5	-4
1	0	-3
2	5	0

-6	-5	-4
1	0	-3
2	5	0

Definition

We call v_π the <u>state-value function</u> for policy π

The value of state s under a policy π is

$$v_\pi(s) = \mathbb{E}_\pi[G_t | S_t = s]$$

For each **state s**

it yields the **expected return**

-6	-5	-4
1	0	-3
2	5	0

Definition

We call v_π the <u>state-value function</u> for policy π

The value of state s under a policy π is

$$v_\pi(s) = \mathbb{E}_\pi[G_t | S_t = s]$$

For each **state s**
it yields the **expected** **return**
if the agent **starts in state s**
and then uses **the policy**
to choose its actions for all time steps

-6	-5	-4
1	0	-3
2	5	0

Definition

We call v_π the state-value function for policy π

The value of state s under a policy π is

$$v_\pi(s) = \mathbb{E}_\pi[G_t | S_t = s]$$

For each **state s**

it yields the **expected return**

if the agent **starts in state s**

and then uses **the policy**

to choose its actions for all time steps

Note #1: The notation $\mathbb{E}\pi[\cdot]$ *is borrowed from the suggested textbook*. $\mathbb{E}$\pi[\cdot] is defined as the expected value of a random variable, given that the agent follows policy π.

Note #2: In this course, we will use "return" and "discounted return" interchangably. For an arbitrary time step t, both terms refer to $G_t \doteq R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \ldots = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}$, where $\gamma \in [0, 1]$. In particular, when we refer to "return", it is not necessarily the case that $\gamma = 1$, and when we refer to "discounted return", it is not necessarily true that $\gamma < 1$. (*This also holds for the readings in the recommended textbook*.)