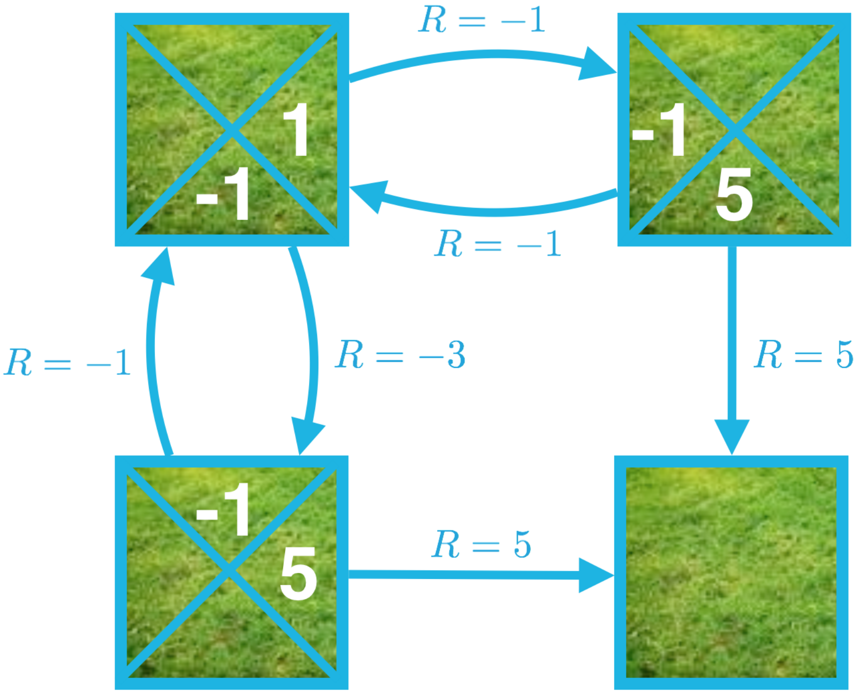
Take the time now to verify that the below image corresponds to the **action-value function** for the same policy.



As an example, consider $q_\pi(s_1, \text{right})$. This action value can be calculated as

$$q_\pi(s_1, \text{right}) = -1 + v_\pi(s_2) = -1 + 2 = 1,$$

where we just use the fact that we can express the value of the state-action pair s_1, right as the sum of two quantities: (1) the immediate reward after moving right and landing on state s_2, and (2) the cumulative reward obtained if the agent begins in state s_2 and follows the policy.

Please now use the state-value function v_π to calculate $q_\pi(s_1, \text{down})$, $q_\pi(s_2, \text{left})$, $q_\pi(s_2, \text{down})$, $q_\pi(s_3, \text{up})$, and $q_\pi(s_3, \text{right})$.

For More Complex Environments