Bellman Equations

In this gridworld example, once the agent selects an action,

- it always moves in the chosen direction (contrasting general MDPs where the agent doesn't always have complete control over what the next state will be), and
- the reward can be predicted with complete certainty (contrasting general MDPs where the reward is a random draw from a probability distribution).

In this simple example, we saw that the value of any state can be calculated as the sum of the immediate reward and the (discounted) value of the next state.

Alexis mentioned that for a general MDP, we have to instead work in terms of an *expectation*, since it's not often the case that the immediate reward and next state can be predicted with certainty. Indeed, we saw in an earlier lesson that the reward and next state are chosen according to the one-step dynamics of the MDP. In this case, where the reward r and next state s' are drawn from a (conditional) probability distribution $p(s',r|s,a)$, the **Bellman Expectation Equation (for v_π)** expresses the value of any state s in terms of the *expected* immediate reward and the *expected* value of the next state: