# Reinforcement Learning Fundamentals

## Policies and Value Functions

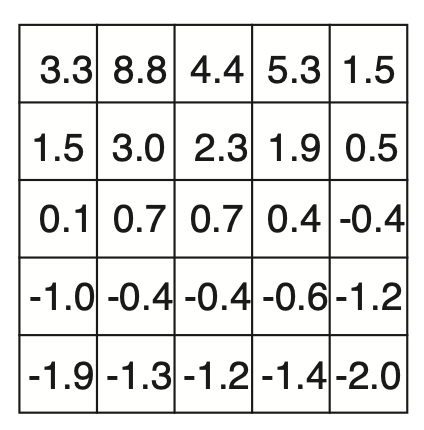
Almost all reinforcement learning algorithms involve estimating **value functions** – functions of state (or state-action pairs) that estimate how good it is for the agent to be in a given state (or how good it is to perform a given action in a given state).

***Gridworld***

Gridworld is an example illustrating the use of a finite Markov Decision Process.

The cells of the grid correspond to the states of the environment. At each cell, four actions are possible: ***north***, ***south***, ***east***, and ***west***, which deterministically cause the agent to move one cell in the respective direction on the grid. Actions that would take the agent off the grid leave its location unchanged, but also result in a reward of -1.

Other actions result in a reward of 0, except those that move the agent out of the special states and . From state , all four actions yield a reward of +10 and take the agent to . From state B



A picture containing symbol, font, white, graphics

Description automatically generatedA picture containing diagram, line

Description automatically generated