The Bellman Equation  
For a given state , the state-value function under a policy can be written as:

Where:

* : Current state.
* : Action taken from state .
* : Probability of taking action in state under policy .
* : Transition probability of reaching state given and .
* : Reward received after transitioning to from using action .
* : Discount factor to account for future rewards.
* : Value function of the next state .

For a given state , the policy improvement step aims to find the best action that maximizes the expected value of future rewards. The policy improvement equation can be written as:

Where:

* : Current state.
* : Action taken from state .
* : Improved policy, which selects the action that maximizes the expected return from state .
* : Value function of the next state resenting the expected return starting from .