

**function** VALUE-ITERATION( $mdp, \epsilon$ ) **returns** a utility function

**inputs:**  $mdp$ , an MDP with states  $S$ , actions  $A(s)$ , transition model  $P(s' | s, a)$ , rewards  $R(s, a, s')$ , discount  $\gamma$   
 $\epsilon$ , the maximum error allowed in the utility of any state

**local variables:**  $U, U'$ , vectors of utilities for states in  $S$ , initially zero  
 $\delta$ , the maximum relative change in the utility of any state

**repeat**

$U \leftarrow U'; \delta \leftarrow 0$

**for each** state  $s$  **in**  $S$  **do**

$U'[s] \leftarrow \max_{a \in A(s)} \text{Q-VALUE}(mdp, s, a, U)$

**if**  $|U'[s] - U[s]| > \delta$  **then**  $\delta \leftarrow |U'[s] - U[s]|$

**until**  $\delta \leq \epsilon(1 - \gamma)/\gamma$

**return**  $U$