If we land up at π', which is not better but as good as π, are $\pi \& \pi'$ optimal policies?

$\Downarrow$

$V_{\pi'} = V_\pi$ for $\forall s \in S$ — (1)

$\pi'(s) = \arg\max_a q_\pi(s,a)$ — (2)

with (1), will (2) imply the following?

$$V_{\pi'}(s) = \max_a E\left[R_{t+1} + \gamma V_{\pi'}(S_{t+1}) \mid S_t = s, A_t = a\right]$$

$V_\pi(s) \le q_\pi(s, \pi'(s))$
$= V_{\pi'}(s)$

$q_\pi(s,a) = E\left[R_{t+1} + \gamma V_\pi(S_{t+1}) \mid S_t = s, A_t = a\right]$

Model-based MDP

$p(s', r \mid s, a)$ availability of model implies you know this probability — which is a 4-argument function

We first discuss policy iteration (PI)

Refer to PI pseudocode on pg. 80 from T.B.

$A = \{1, 1, 0, -1\}$

$\arg\max_i A$

$= 0, 1$ or $1, 2$

1. PE → given π, you find $V_\pi(s)$ $\forall s \in S$ for s^+
by iteratively solving Bellman equations.

2. P.I. → given $\pi \rightarrow V_\pi(s) \rightarrow \pi' \rightarrow V_{\pi'}(s) \rightarrow \pi'' \rightarrow V_{\pi''}(s) \ldots$
$\forall s$ $\forall s$

until you locate π^* & $V_{\pi^*}(s)$

Coming back to last slide —

$$V_\pi(s) = V_{\pi'}(s) = q_\pi(s, \pi'(s))$$
$$= \max_a q_\pi(s, a)$$