Markov Decision Processes

Chen-Yu Wei

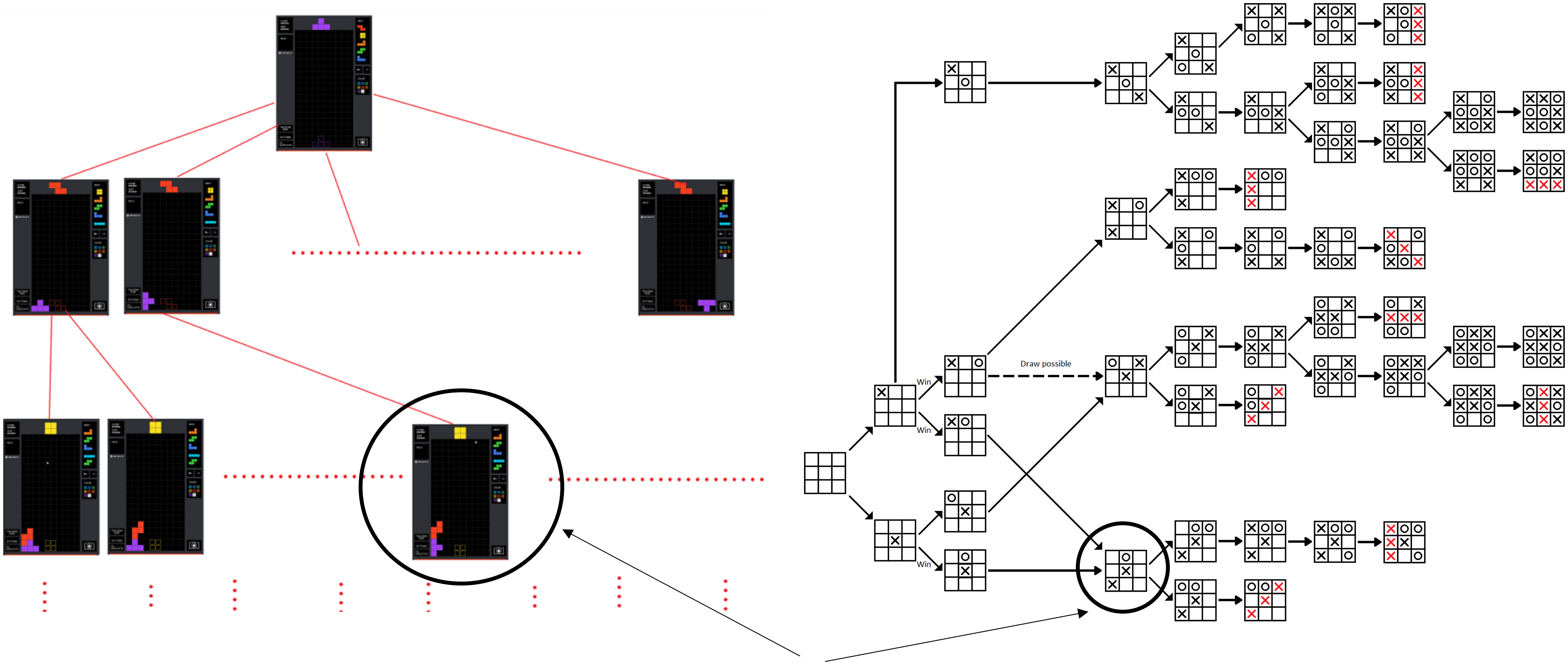
Sequence of Actions

To win the game, the learner has to take a sequence of actions $a_1 \to a_2 \to \cdots \to a_H$.

The effect of a particular action may not be revealed instantaneously.

- Some effect may be revealed instantaneously
- Some may be revealed later

Sequence of Actions

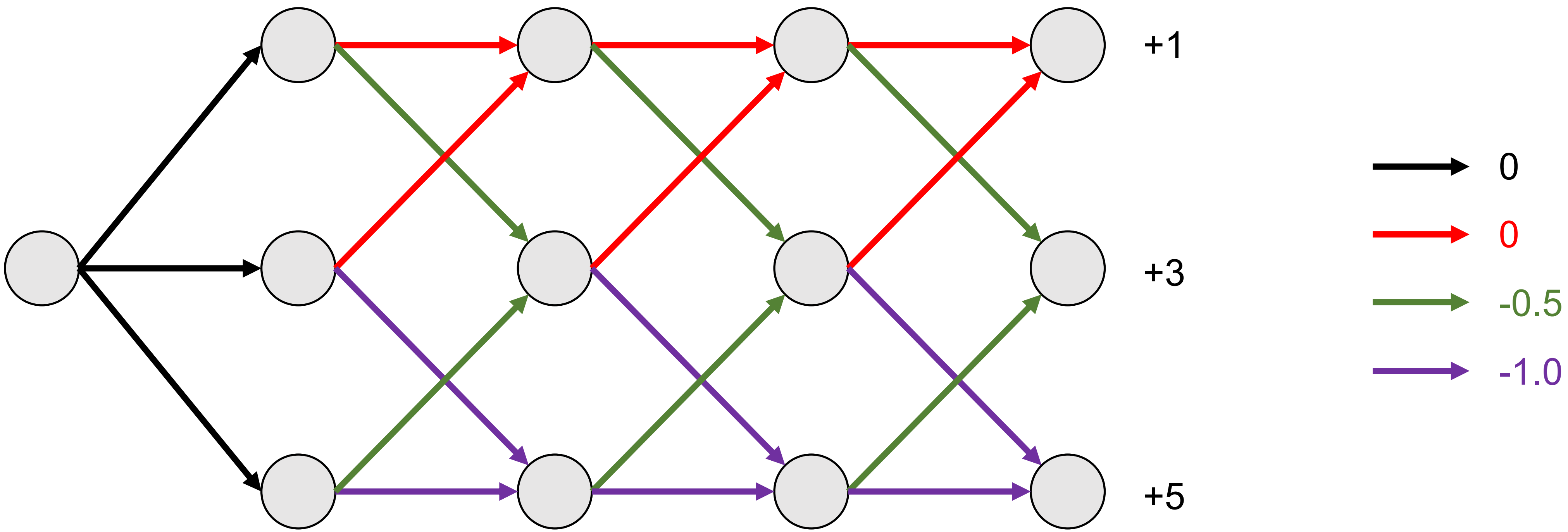


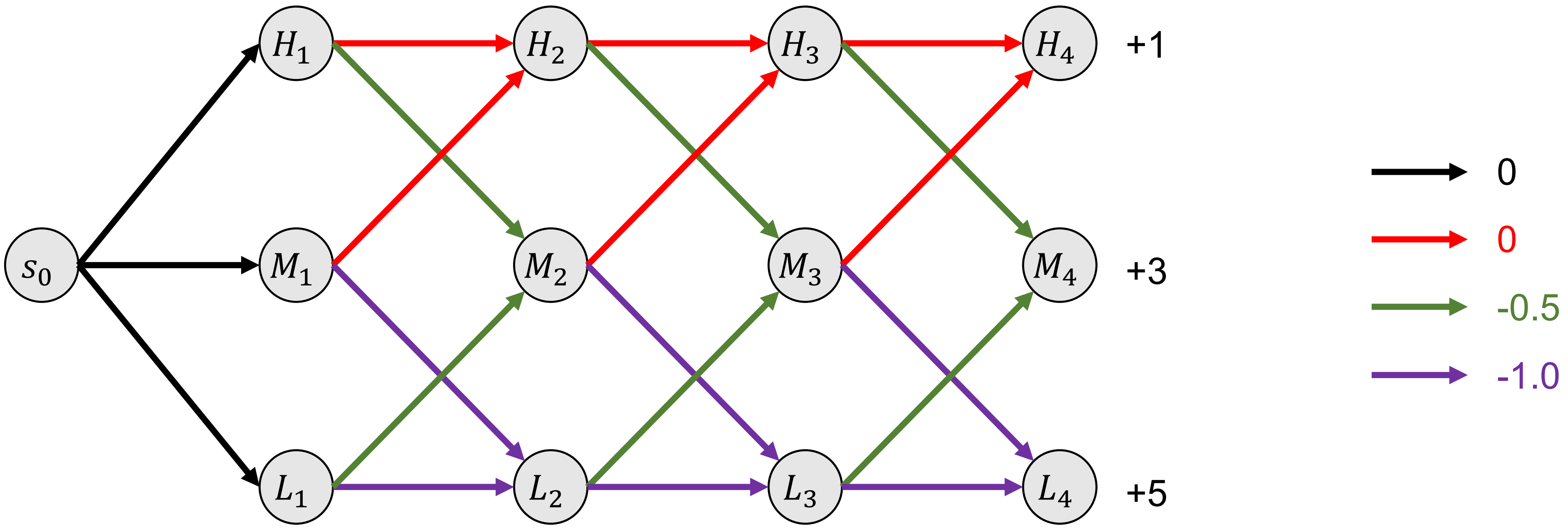
state

(a summary of the current status in a multi-stage game)

Warm-Up: Deterministic World

Which path gives us the highest **total reward**?

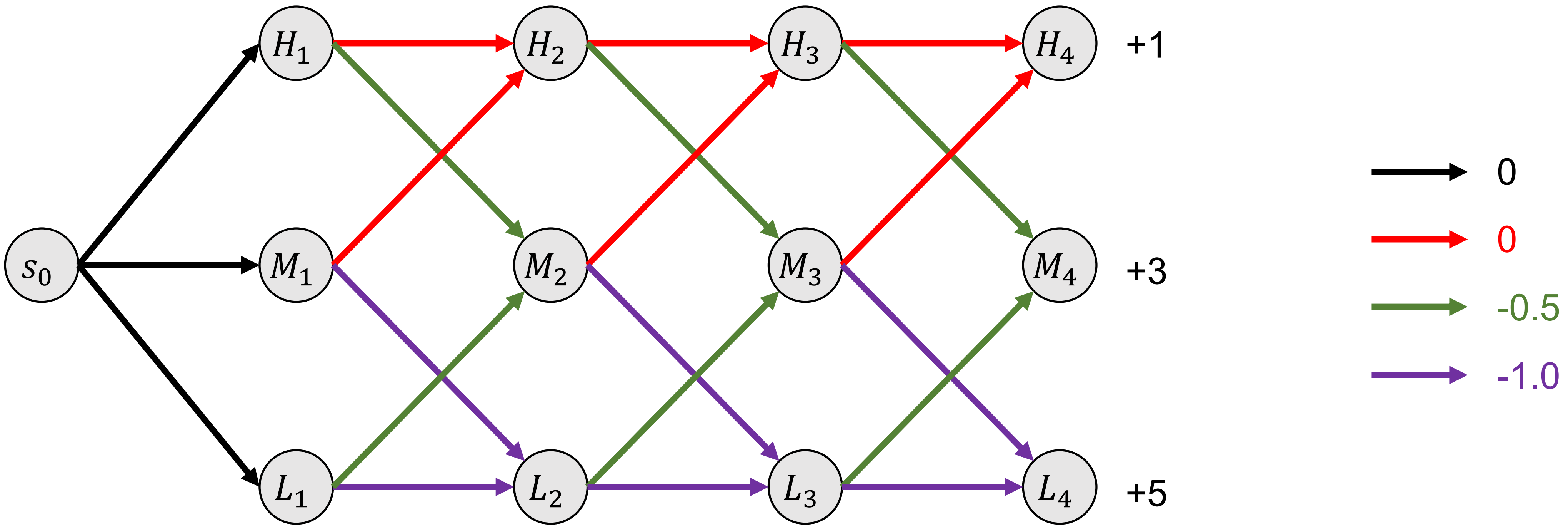




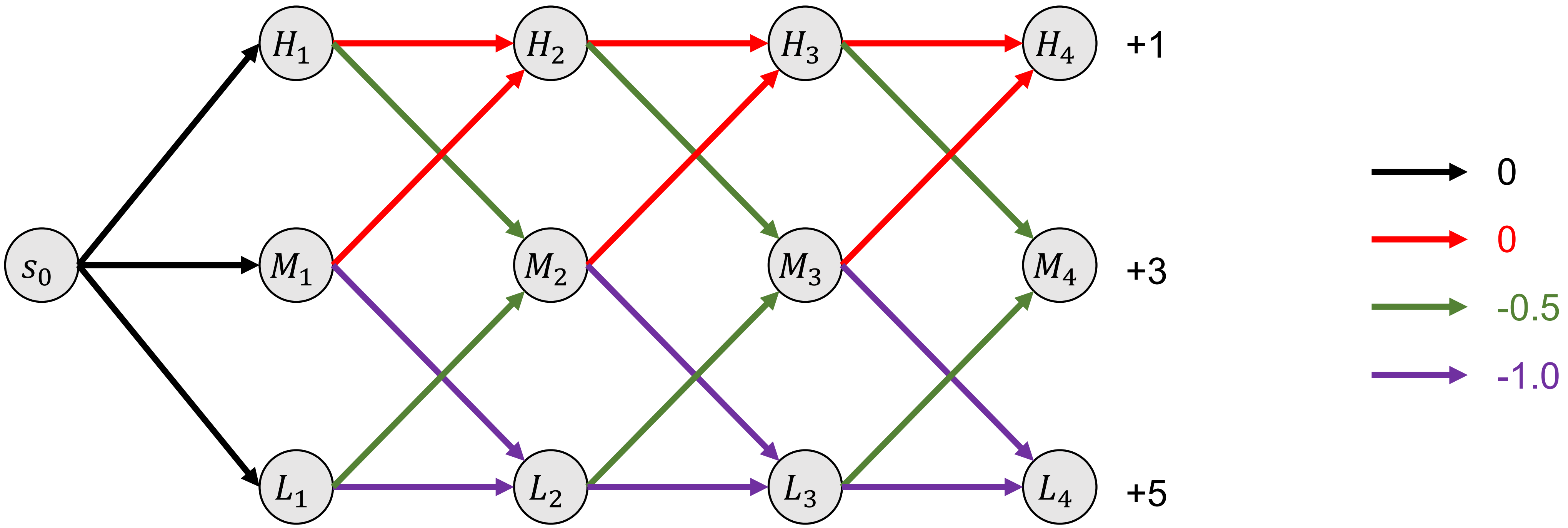
$V^\star(s)$ ≔ maximum total reward starting from state s

$Q^\star(s,a)$ ≔ maximum total reward starting from state s and taking action a **for one step**, and then following the optimal strategy

$\pi^\star(s)$ ≔ optimal decision on state s



$V^\star(H_4) =$	$Q^\star(H_3, R) =$	$V^\star(H_3) =$	$Q^\star(H_2, R) =$	$V^\star(H_2) =$
	$Q^\star(H_3, G) =$		$Q^\star(H_2, G) =$	
$V^\star(M_4) =$	$Q^\star(M_3, R) =$	$V^\star(M_3) =$	$Q^\star(M_2, R) =$	$V^\star(M_2) =$
	$Q^\star(M_3, P) =$		$Q^\star(M_2, P) =$	
$V^\star(L_4) =$	$Q^\star(L_3, G) =$	$V^\star(L_3) =$	$Q^\star(L_2, G) =$	$V^\star(L_2) =$
	$Q^\star(L_3, P) =$		$Q^\star(L_2, P) =$	

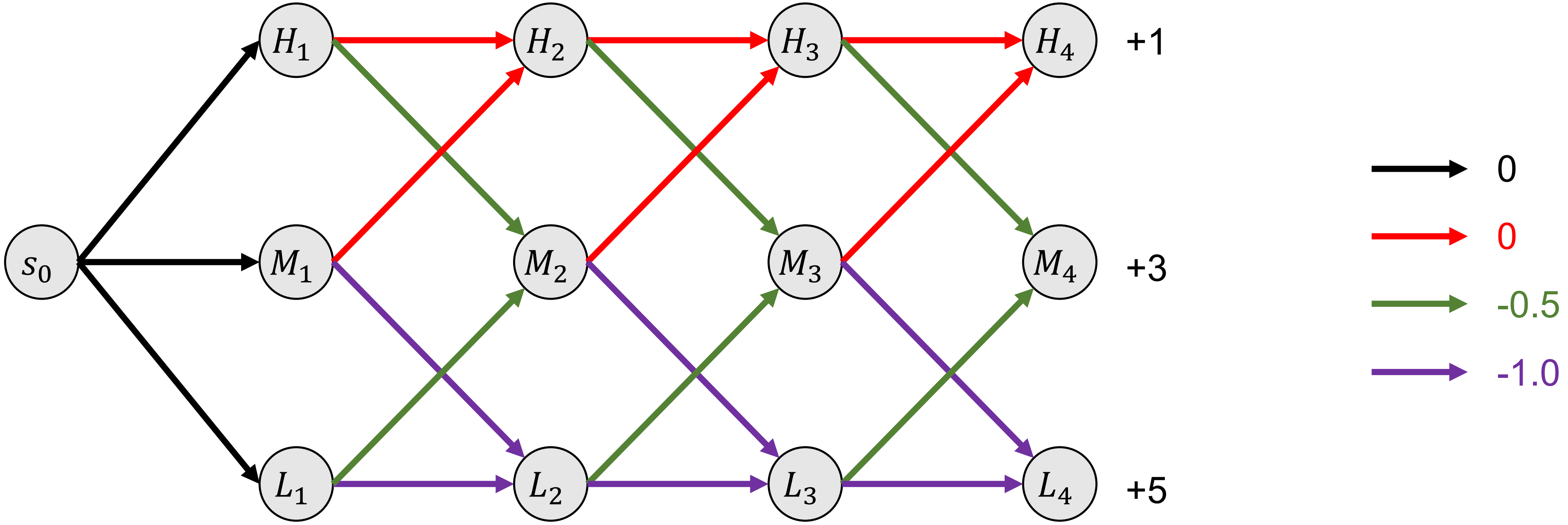


General rule:

$$Q^\star(s,a) = R(s,a) + V^\star(\text{next\_state}(s,a))$$

$$V^\star(s) = \max_a Q^\star(s,a)$$

$$\pi^\star(s) = \operatorname*{argmax}_a Q^\star(s,a)$$



General algorithm:

Repeat until Q, V no longer changes:

$Q(s,a) \leftarrow R(s,a) + V(\text{next\_state}(s,a))$ for all (s,a)

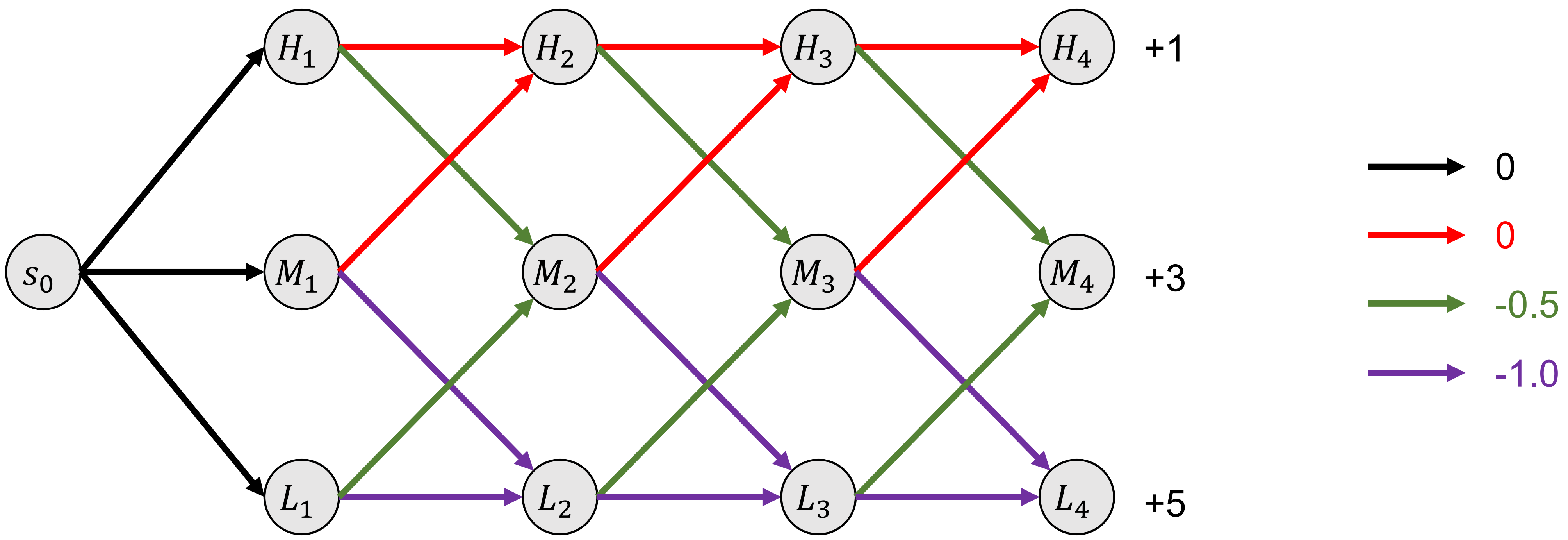
$V(s) = \max_a Q(s,a)$ for all s

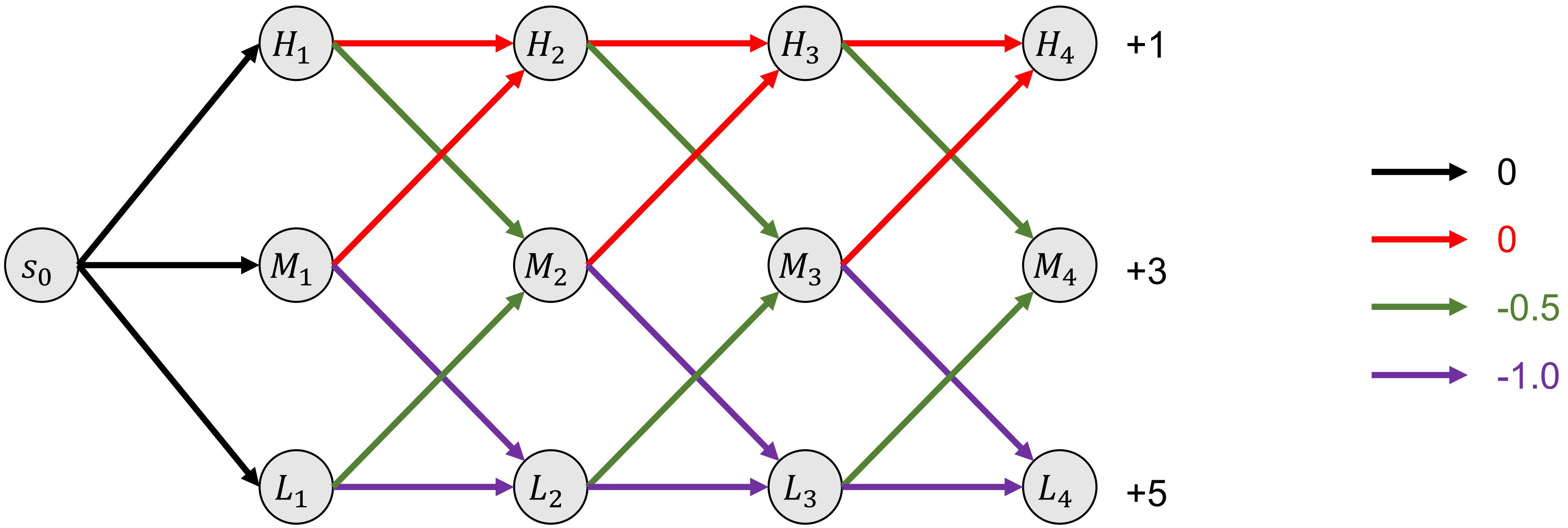
$\pi^\star(s) = \operatorname{argmax}_a Q(s,a)$

Stochastic Worlds

Now, suppose taking an action does **not** lead to the desired state deterministically.

Instead, with probability 0.8, it goes to the state as specified in the figure;
with probability 0.1 each, it goes to the other two states.





$V^\star(H_4) =$

$V^\star(M_4) =$

$V^\star(L_4) =$

$Q^\star(H_3, R) =$

$Q^\star(H_3, G) =$

$Q^\star(M_3, R) =$

$Q^\star(M_3, P) =$

$Q^\star(L_3, G) =$

$Q^\star(L_3, P) =$

$V^\star(H_3) =$

$V^\star(M_3) =$

$V^\star(L_3) =$

$Q^\star(H_2, R) =$

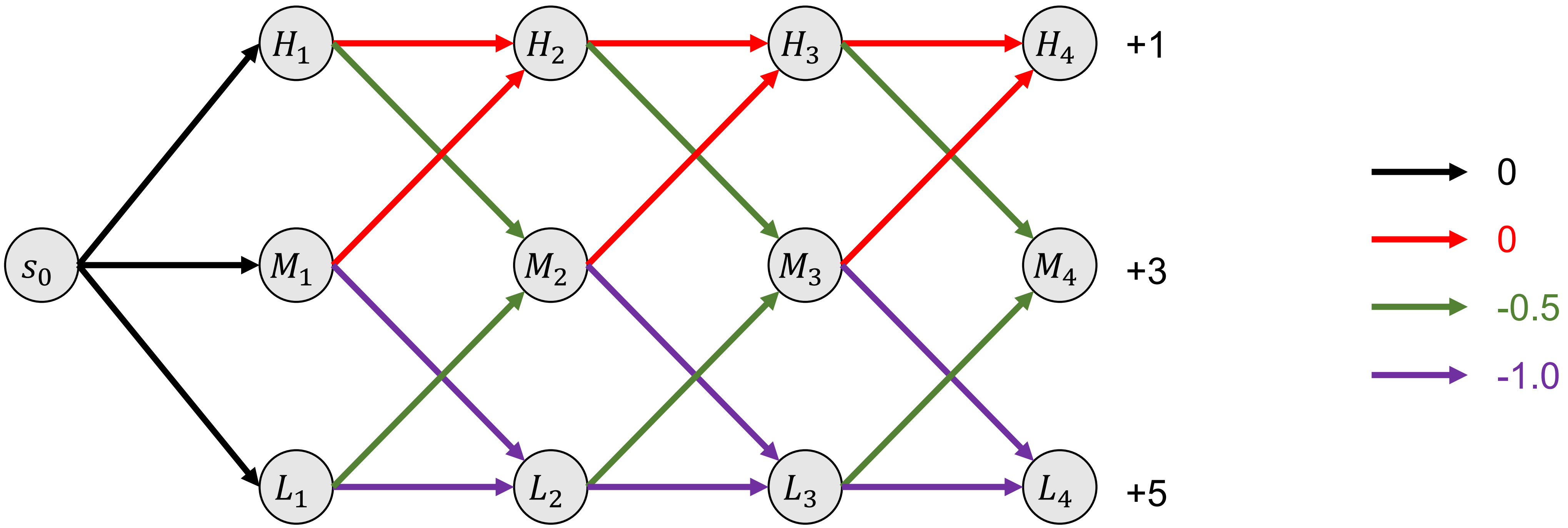
$Q^\star(H_2, G) =$

$Q^\star(M_2, R) =$

$Q^\star(M_2, P) =$

$Q^\star(L_2, G) =$

$Q^\star(L_2, P) =$

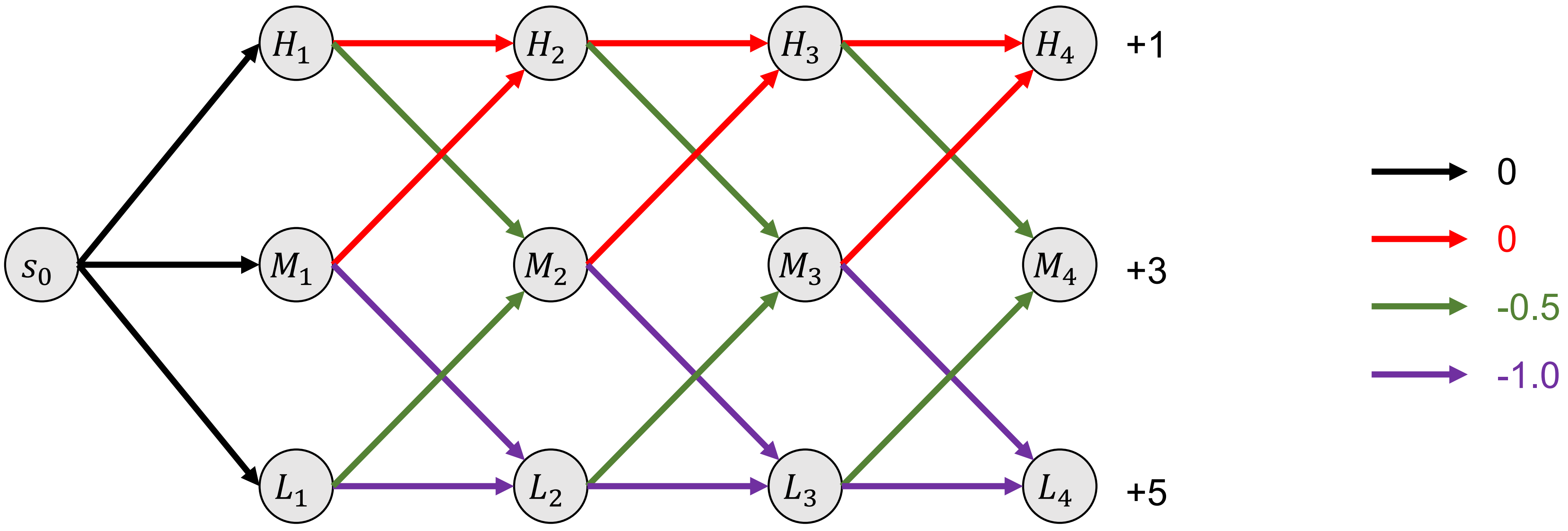


General rule:

$$Q^\star(s,a) = R(s,a) + \sum_{s'} P(s'|s,a)\, V^\star(s')$$

$$V^\star(s) = \max_a Q^\star(s,a)$$

$$\pi^\star(s) = \operatorname*{argmax}_a Q^\star(s,a)$$



General algorithm (Value Iteration):

Repeat until Q, V no longer changes:

$$Q(s,a) \leftarrow R(s,a) + \sum_{s'} P(s'|s,a)\, V^\star(s') \qquad \text{for all } (s,a)$$

$$V(s) = \max_a Q(s,a) \qquad \text{for all } s$$

$$\pi^\star(s) = \operatorname*{argmax}_a Q(s,a)$$