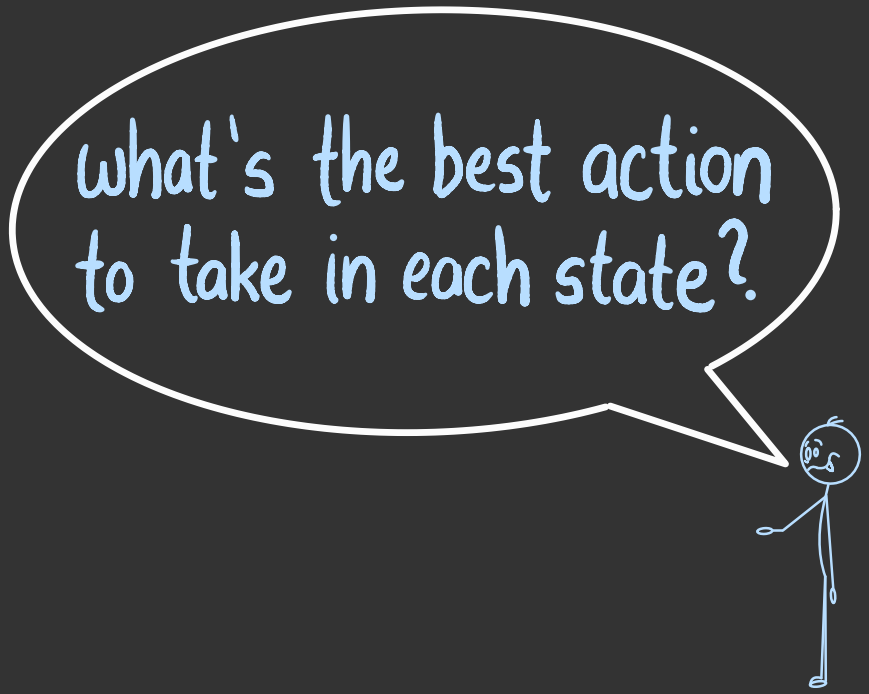
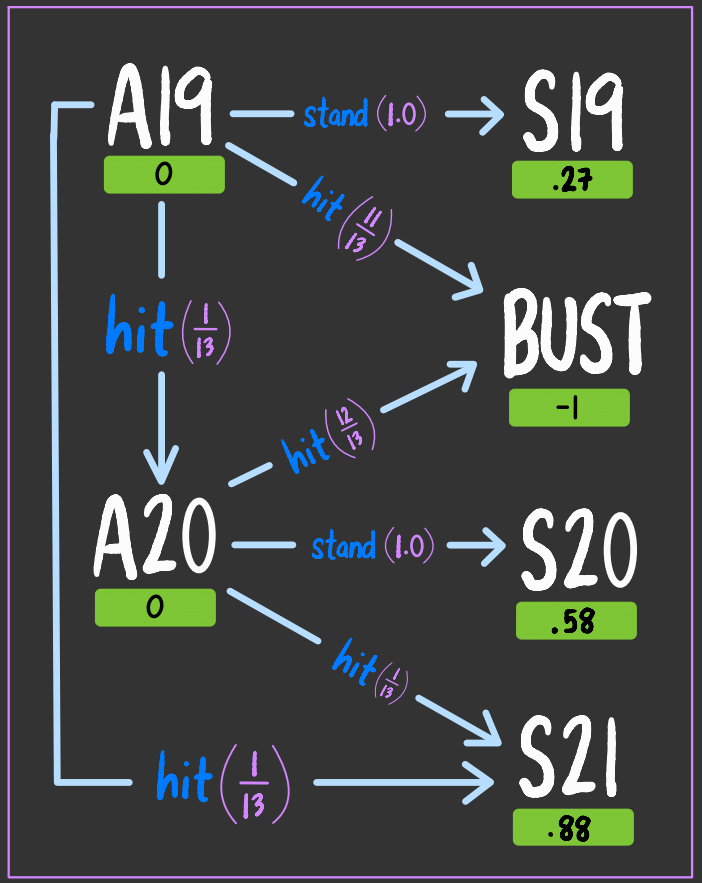
policies

CSCI 373

given a markov decision process (mdp), we might ask:

what's the best action to take in each state?





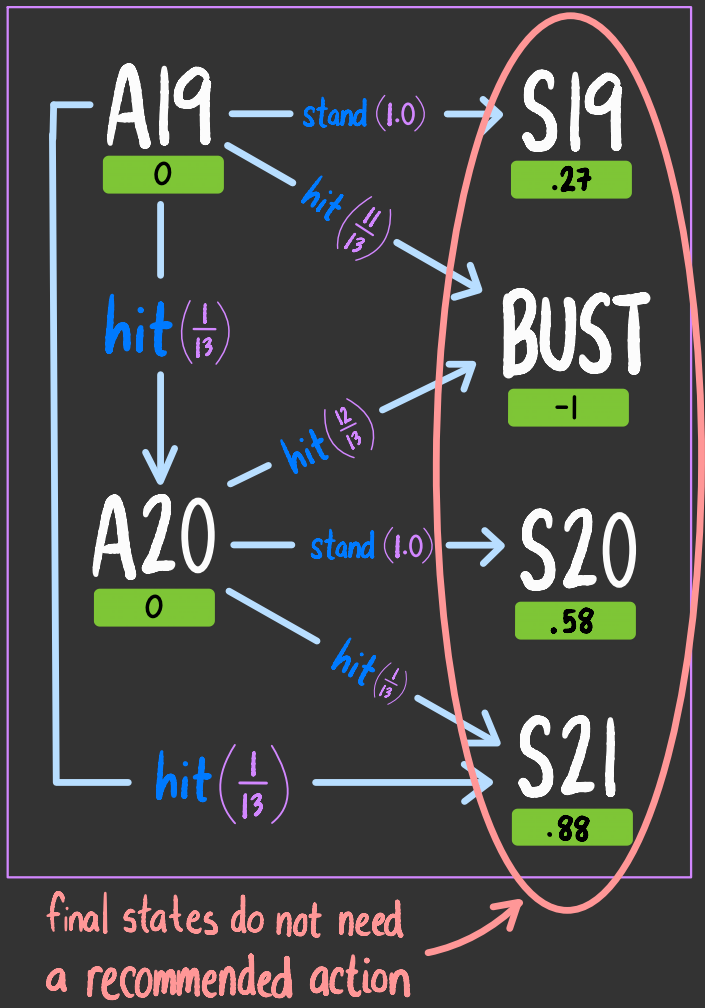
state machine

the answer to this question is called a

policy

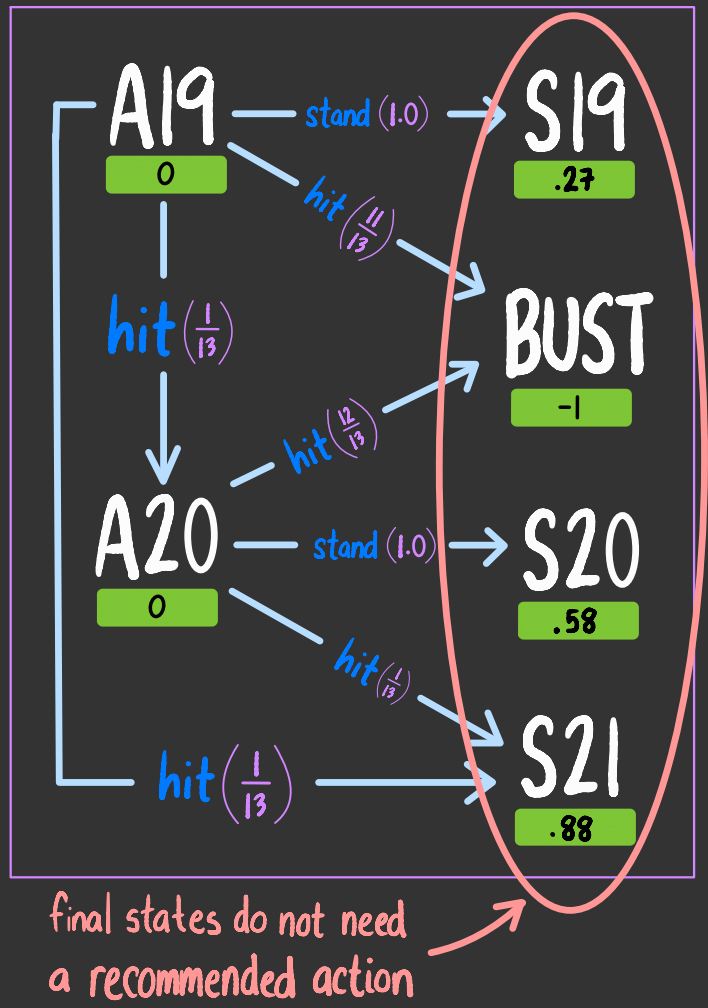
in A19, you should **stand**

in A20, you should **stand**



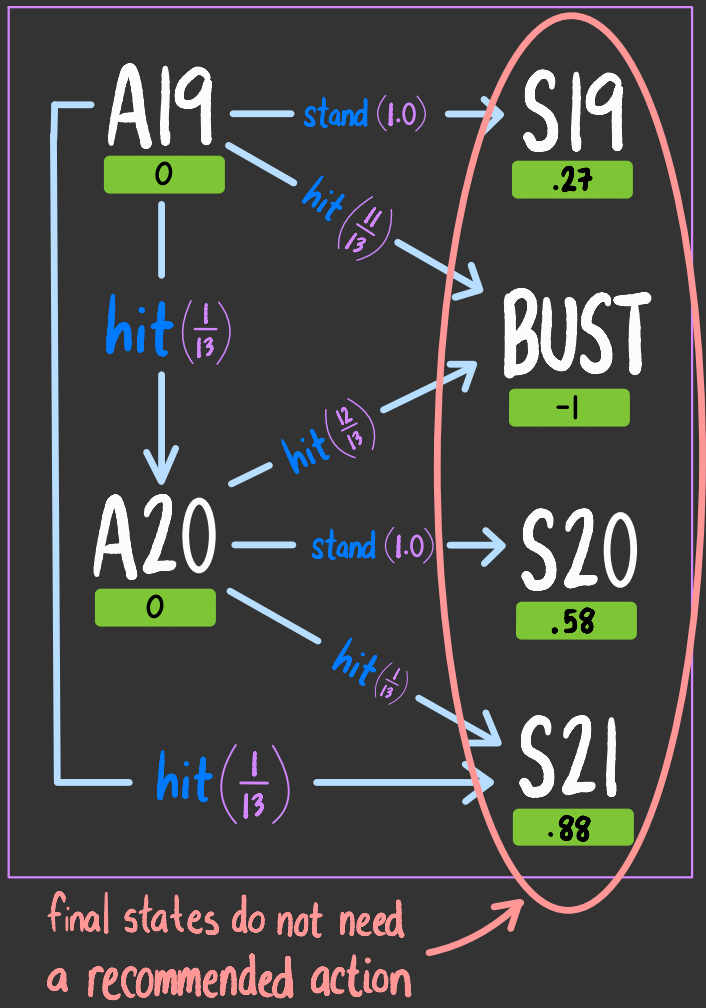
a **policy** is just a function $\pi: (Q \setminus F) \to \Sigma$ from the non-final states to actions

in A19, you should **stand**
in A20, you should **stand**



a **policy** is just a function $\pi : (Q \setminus F) \to \Sigma$ from the non-final states to actions

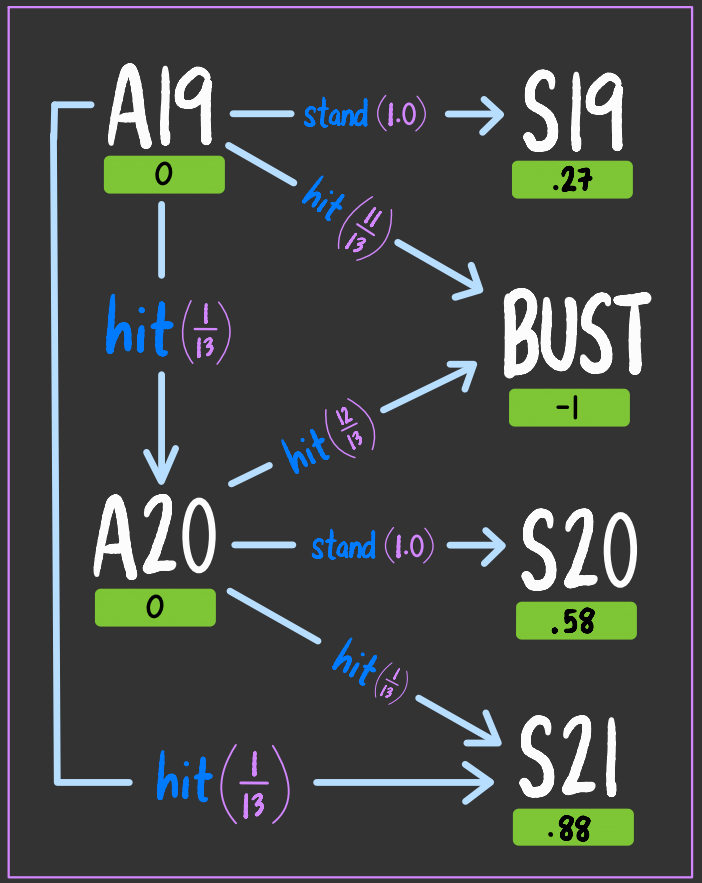
$$\pi(\text{A19}) = \text{stand}$$
$$\pi(\text{A20}) = \text{stand}$$



to determine which policy to adopt, we need a way to assess the quality of a policy

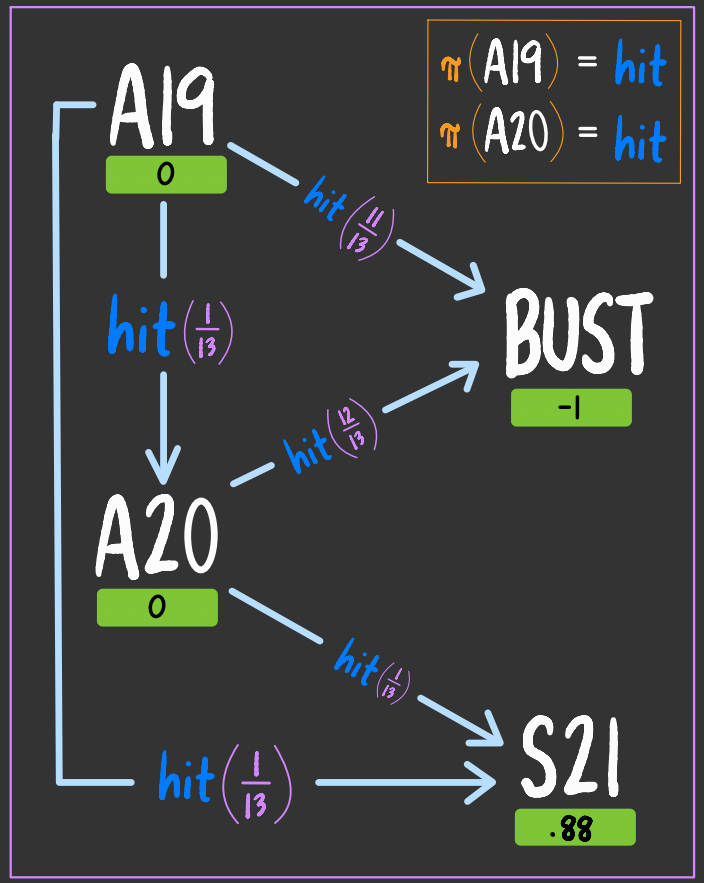
$\pi(A19) = hit$

$\pi(A20) = hit$



markov decision process (mdp)

we quantify this as our expected lifetime reward if we adopt the policy



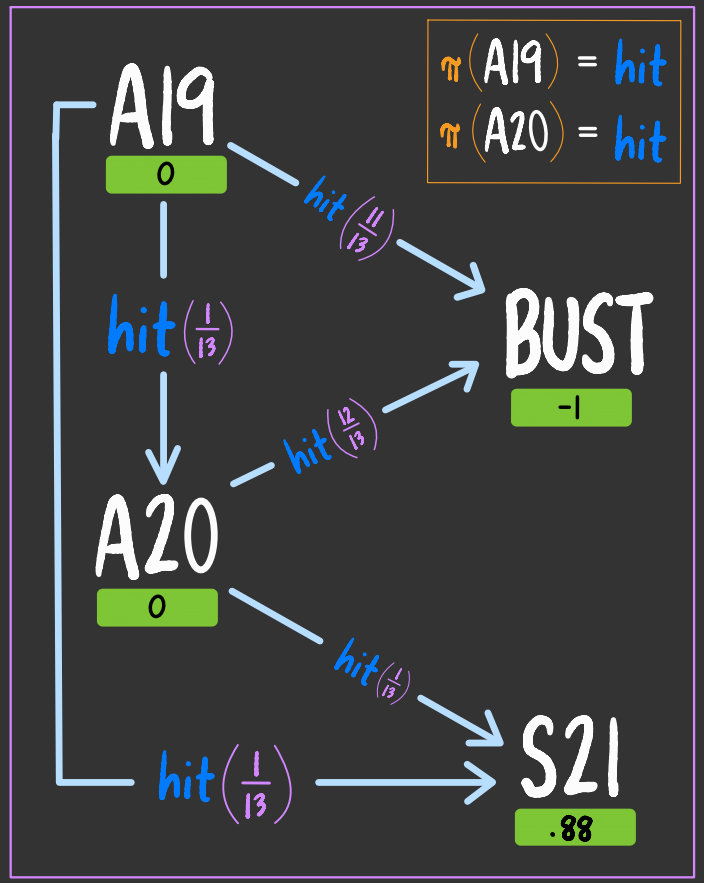
the policy permits
4 possible futures

A19 → A20 → S21

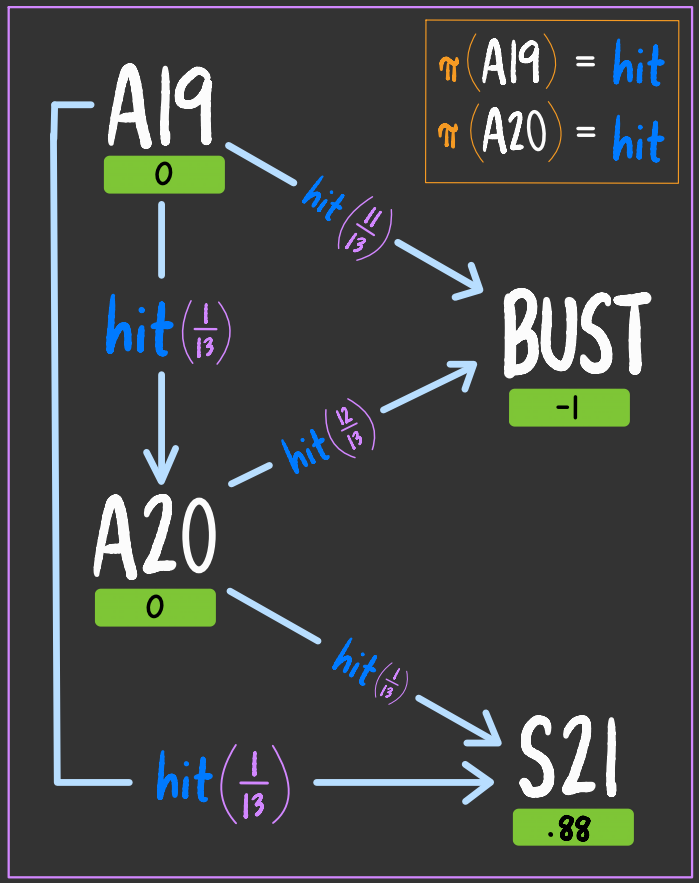
A19 → A20 → BUST

A19 → S21

A19 → BUST

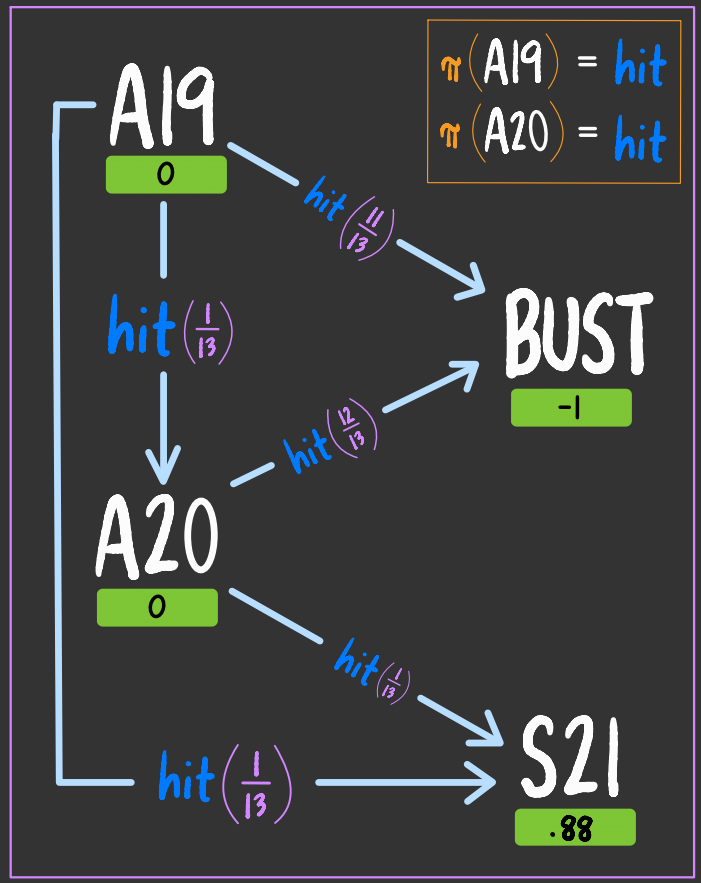


future	lifetime reward	probability
A19 → A20 → S21	?	?
A19 → A20 → BUST	?	?
A19 → S21	?	?
A19 → BUST	?	?



markov decision process (mdp)

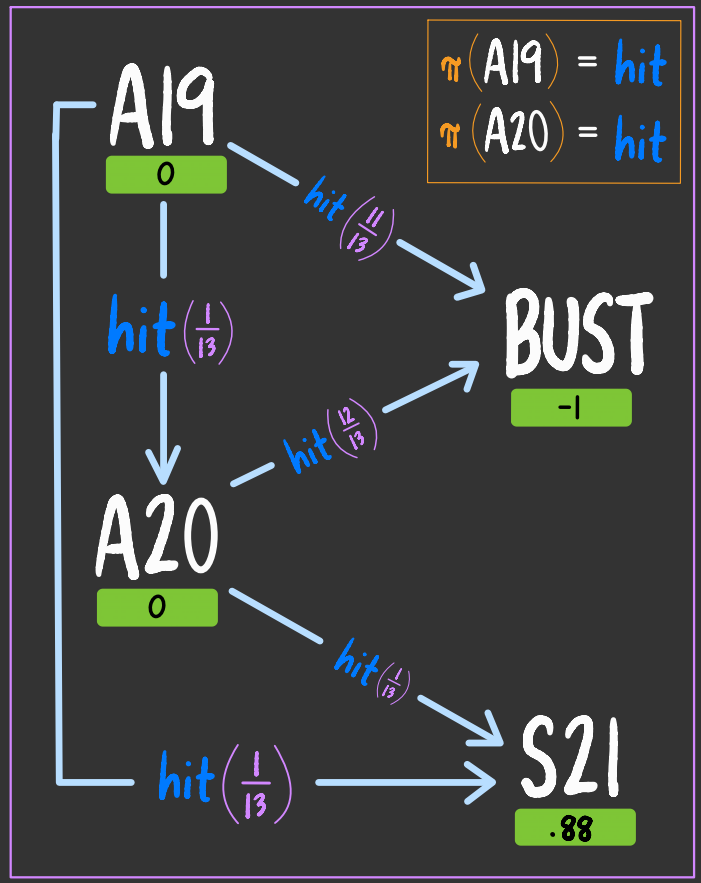
future	lifetime reward	probability
A19 → A20 → S21	.88	$\frac{1}{13} \cdot \frac{1}{13}$
A19 → A20 → BUST	-1	$\frac{1}{13} \cdot \frac{12}{13}$
A19 → S21	.88	$\frac{1}{13}$
A19 → BUST	-1	$\frac{11}{13}$



markov decision process (mdp)

future	lifetime reward		probability
A19 → A20 → S21	.88	×	$\frac{1}{13} \cdot \frac{1}{13}$
A19 → A20 → BUST	+ −1	×	$\frac{1}{13} \cdot \frac{12}{13}$
A19 → S21	+ .88	×	$\frac{1}{13}$
A19 → BUST	+ −1	×	$\frac{11}{13}$

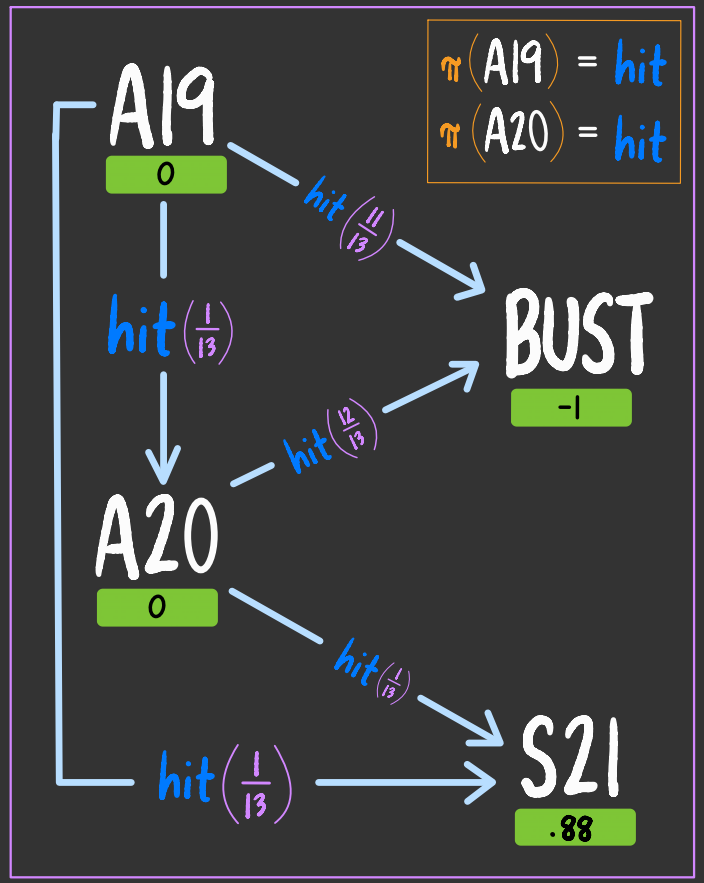
expected lifetime reward = −.84 (approx.)



markov decision process (mdp)

future	utility		probability
A19 → A20 → S21	.88	×	$\frac{1}{13} \cdot \frac{1}{13}$
A19 → A20 → BUST	+ −1	×	$\frac{1}{13} \cdot \frac{12}{13}$
A19 → S21	+ .88	×	$\frac{1}{13}$
A19 → BUST	+ −1	×	$\frac{11}{13}$

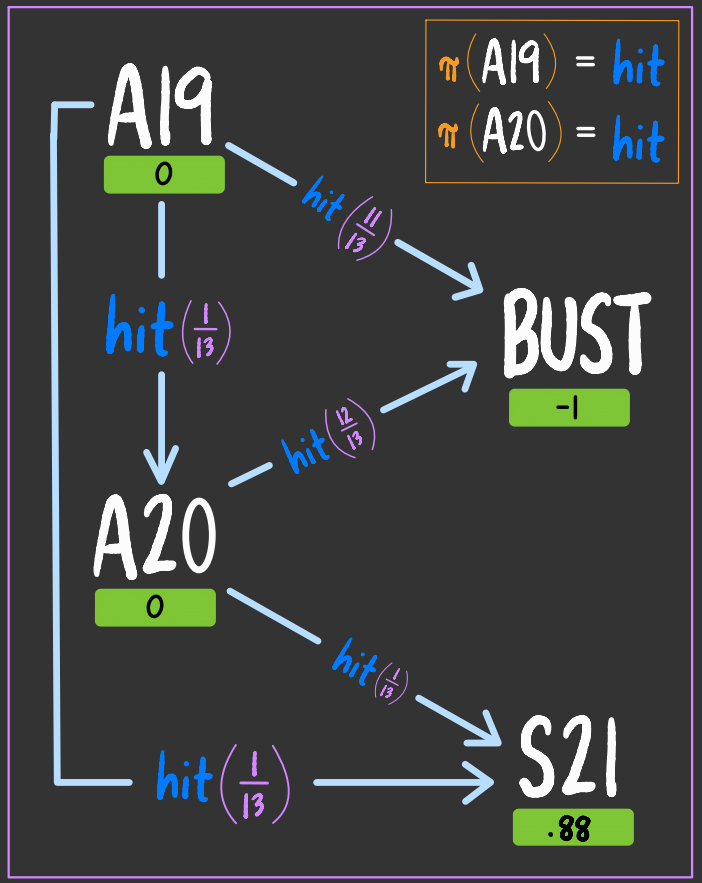
expected utility = −.84 (approx.)



markov decision process (mdp)

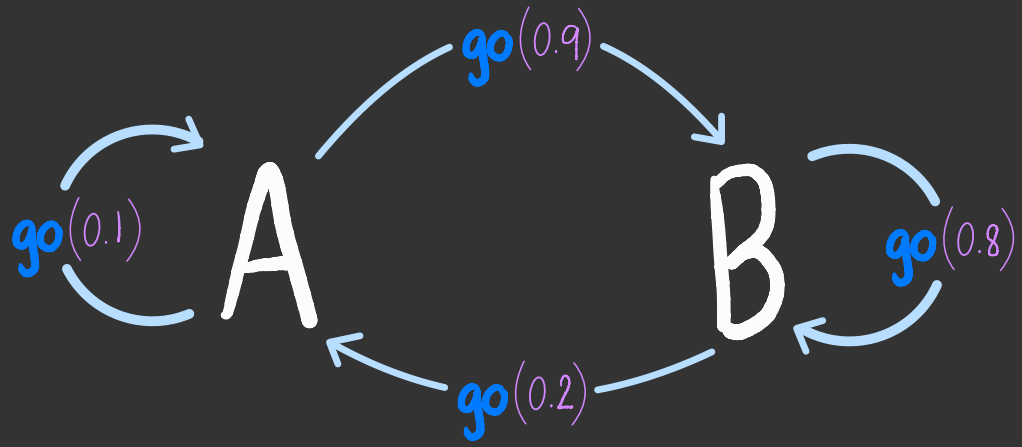
future	utility	probability
A19 → A20 → S21	.88 ×	$\frac{1}{13} \cdot \frac{1}{13}$
A19 → A20 → BUST	+ −1 ×	$\frac{1}{13} \cdot \frac{12}{13}$
A19 → S21	+ .88 ×	$\frac{1}{13}$
A19 → BUST	+ −1 ×	$\frac{11}{13}$

$$U^{\pi}(A19) = -.84 \quad \text{(approx.)}$$



markov decision process (mdp)

but for some mdps, this is not so straightforward



how would we compute the expected utility of state A?

$$U^{\pi}(A19) = U(A19 \to A20 \to S21)\, P(A19 \to A20 \to S21)$$
$$+ U(A19 \to A20 \to \text{BUST})\, P(A19 \to A20 \to \text{BUST})$$
$$+ U(A19 \to S21)\, P(A19 \to S21)$$
$$+ U(A19 \to \text{BUST})\, P(A19 \to \text{BUST})$$

future	utility	probability
$A19 \to A20 \to S21$	$.88$	$\times \frac{1}{13} \cdot \frac{1}{13}$
$A19 \to A20 \to \text{BUST}$	$+ -1$	$\times \frac{1}{13} \cdot \frac{12}{13}$
$A19 \to S21$	$+ .88$	$\times \frac{1}{13}$
$A19 \to \text{BUST}$	$+ -1$	$\times \frac{11}{13}$

$$U^{\pi}(A19) = -.84 \quad \text{(approx.)}$$

$$U^{\pi}(A19) = U(A19 \to A20 \to S21)\, P(A19 \to A20 \to S21)$$
$$+ U(A19 \to A20 \to \text{BUST})\, P(A19 \to A20 \to \text{BUST})$$
$$+ U(A19 \to S21)\, P(A19 \to S21)$$
$$+ U(A19 \to \text{BUST})\, P(A19 \to \text{BUST})$$

let:

$$p_1 = P(A19 \to A20 \to S21)$$
$$p_2 = P(A19 \to A20 \to \text{BUST})$$
$$p_3 = P(A19 \to S21)$$
$$p_4 = P(A19 \to \text{BUST})$$

$$U^{\pi}(A19) = p_1 U(A19 \to A20 \to S21) + p_2 U(A19 \to A20 \to \text{BUST}) + p_3 U(A19 \to S21) + p_4 U(A19 \to \text{BUST})$$

where:

$$p_1 = P(A19 \to A20 \to S21)$$
$$p_2 = P(A19 \to A20 \to \text{BUST})$$
$$p_3 = P(A19 \to S21)$$
$$p_4 = P(A19 \to \text{BUST})$$

$$U^{\pi}(A19) = p_1 U(A19 \to A20 \to S21) + p_2 U(A19 \to A20 \to BUST) + p_3 U(A19 \to S21) + p_4 U(A19 \to BUST)$$

where:

$$p_1 = P(A19 \to A20 \to S21)$$
$$p_2 = P(A19 \to A20 \to BUST)$$
$$p_3 = P(A19 \to S21)$$
$$p_4 = P(A19 \to BUST)$$

$$= p_1 \left(R_0(A19) + R_1(A20) + R_2(S21)\right) + p_2 \left(R_0(A19) + R_1(A20) + R_2(BUST)\right) + p_3 \left(R_0(A19) + R_1(S21)\right) + p_4 \left(R_0(A19) + R_1(BUST)\right)$$

recall:

$$R_2(S21) = \gamma R_1(S21)$$

$$U^{\pi}(A19) = p_1\left(R_0(A19) + R_1(A20) + R_2(S21)\right)$$
$$+ p_2\left(R_0(A19) + R_1(A20) + R_2(\text{BUST})\right)$$
$$+ p_3\left(R_0(A19) + R_1(S21)\right)$$
$$+ p_4\left(R_0(A19) + R_1(\text{BUST})\right)$$

where:

$$p_1 = P(A19 \to A20 \to S21)$$
$$p_2 = P(A19 \to A20 \to \text{BUST})$$
$$p_3 = P(A19 \to S21)$$
$$p_4 = P(A19 \to \text{BUST})$$

$$
\begin{aligned}
U^{\pi}(\text{A19}) = {} & p_1 \left(R_0(\text{A19}) + R_1(\text{A20}) + R_2(\text{S21}) \right) \\
& + p_2 \left(R_0(\text{A19}) + R_1(\text{A20}) + R_2(\text{BUST}) \right) \\
& + p_3 \left(R_0(\text{A19}) + R_1(\text{S21}) \right) \\
& + p_4 \left(R_0(\text{A19}) + R_1(\text{BUST}) \right) \\
= {} & p_1 R_0(\text{A19}) + p_1 R_1(\text{A20}) + p_1 R_2(\text{S21}) \\
& + p_2 R_0(\text{A19}) + p_2 R_1(\text{A20}) + p_2 R_2(\text{BUST}) \\
& + p_3 R_0(\text{A19}) + p_3 R_1(\text{S21}) \\
& + p_4 R_0(\text{A19}) + p_4 R_1(\text{BUST})
\end{aligned}
$$

where:

$$
\begin{aligned}
p_1 &= P(\text{A19} \rightarrow \text{A20} \rightarrow \text{S21}) \\
p_2 &= P(\text{A19} \rightarrow \text{A20} \rightarrow \text{BUST}) \\
p_3 &= P(\text{A19} \rightarrow \text{S21}) \\
p_4 &= P(\text{A19} \rightarrow \text{BUST})
\end{aligned}
$$

$$U^{\pi}(\text{A19}) = p_1 R_0(\text{A19}) + p_1 R_1(\text{A20}) + p_1 R_2(\text{S21})$$
$$+ p_2 R_0(\text{A19}) + p_2 R_1(\text{A20}) + p_2 R_2(\text{BUST})$$
$$+ p_3 R_0(\text{A19}) + p_3 R_1(\text{S21})$$
$$+ p_4 R_0(\text{A19}) + p_4 R_1(\text{BUST})$$

where:

$$p_1 = P(\text{A19} \to \text{A20} \to \text{S21})$$
$$p_2 = P(\text{A19} \to \text{A20} \to \text{BUST})$$
$$p_3 = P(\text{A19} \to \text{S21})$$
$$p_4 = P(\text{A19} \to \text{BUST})$$

$$U^{\pi}(A19) = \boxed{p_1 R_0(A19) + p_1 R_1(A20) + p_1 R_2(S21)} \quad \text{where: } p_1 = P(A19 \to A20 \to S21)$$
$$+ \boxed{p_2 R_0(A19) + p_2 R_1(A20) + p_2 R_2(BUST)} \qquad p_2 = P(A19 \to A20 \to BUST)$$
$$+ \boxed{p_3 R_0(A19) + p_3 R_1(S21)} \qquad p_3 = P(A19 \to S21)$$
$$+ \boxed{p_4 R_0(A19) + p_4 R_1(BUST)} \qquad p_4 = P(A19 \to BUST)$$

$$= \boxed{p_1 R_0(A19) + p_2 R_0(A19) + p_3 R_0(A19) + p_4 R_0(A19)}$$
$$+ p_1 R_1(A20) + p_1 R_2(S21)$$
$$+ p_2 R_1(A20) + p_2 R_2(BUST)$$
$$+ p_3 R_1(S21)$$
$$+ p_4 R_1(BUST)$$

$$
\begin{aligned}
U^{\pi}(\text{A19}) &= p_1 R_0(\text{A19}) + p_1 R_1(\text{A20}) + p_1 R_2(\text{S21}) \\
&+ p_2 R_0(\text{A19}) + p_2 R_1(\text{A20}) + p_2 R_2(\text{BUST}) \\
&+ p_3 R_0(\text{A19}) + p_3 R_1(\text{S21}) \\
&+ p_4 R_0(\text{A19}) + p_4 R_1(\text{BUST}) \\
&= (p_1 + p_2 + p_3 + p_4) R_0(\text{A19}) \\
&\quad + p_1 R_1(\text{A20}) + p_1 R_2(\text{S21}) \\
&\quad + p_2 R_1(\text{A20}) + p_2 R_2(\text{BUST}) \\
&\quad + p_3 R_1(\text{S21}) \\
&\quad + p_4 R_1(\text{BUST})
\end{aligned}
$$

where:

$$
\begin{aligned}
p_1 &= P(\text{A19} \to \text{A20} \to \text{S21}) \\
p_2 &= P(\text{A19} \to \text{A20} \to \text{BUST}) \\
p_3 &= P(\text{A19} \to \text{S21}) \\
p_4 &= P(\text{A19} \to \text{BUST})
\end{aligned}
$$

$$U^{\pi}(A19) = (p_1 + p_2 + p_3 + p_4) R_0(A19)$$
$$+ p_1 R_1(A20) + p_1 R_2(S21)$$
$$+ p_2 R_1(A20) + p_2 R_2(\text{BUST})$$
$$+ p_3 R_1(S21)$$
$$+ p_4 R_1(\text{BUST})$$

where:

$$p_1 = P(A19 \to A20 \to S21)$$
$$p_2 = P(A19 \to A20 \to \text{BUST})$$
$$p_3 = P(A19 \to S21)$$
$$p_4 = P(A19 \to \text{BUST})$$

$$U^{\pi}(A19) = (p_1 + p_2 + p_3 + p_4) R_0(A19)$$
$$+ p_1 R_1(A20) + p_1 R_2(S21)$$
$$+ p_2 R_1(A20) + p_2 R_2(BUST)$$
$$+ p_3 R_1(S21)$$
$$+ p_4 R_1(BUST)$$

where:
$$p_1 = P(A19 \to A20 \to S21)$$
$$+ p_2 = P(A19 \to A20 \to BUST)$$
$$+ p_3 = P(A19 \to S21)$$
$$+ p_4 = P(A19 \to BUST)$$
$$?$$

what do these sum to?

$$U^{\pi}(A19) = (p_1 + p_2 + p_3 + p_4) R_0(A19)$$
$$+ p_1 R_1(A20) + p_1 R_2(S21)$$
$$+ p_2 R_1(A20) + p_2 R_2(BUST)$$
$$+ p_3 R_1(S21)$$
$$+ p_4 R_1(BUST)$$
$$= R_0(A19)$$
$$+ p_1 R_1(A20) + p_1 R_2(S21)$$
$$+ p_2 R_1(A20) + p_2 R_2(BUST)$$
$$+ p_3 R_1(S21)$$
$$+ p_4 R_1(BUST)$$

where:
$$p_1 = P(A19 \to A20 \to S21)$$
$$+ p_2 = P(A19 \to A20 \to BUST)$$
$$+ p_3 = P(A19 \to S21)$$
$$+ p_4 = P(A19 \to BUST)$$
$$= 1$$

$$U^{\pi}(A19) = R_0(A19) + p_1 R_1(A20) + p_1 R_2(S21) + p_2 R_1(A20) + p_2 R_2(\text{BUST}) + p_3 R_1(S21) + p_4 R_1(\text{BUST})$$

where:

$$p_1 = P(A19 \to A20 \to S21)$$
$$p_2 = P(A19 \to A20 \to \text{BUST})$$
$$p_3 = P(A19 \to S21)$$
$$p_4 = P(A19 \to \text{BUST})$$

$$U^{\pi}(A19) = R_0(A19)$$
$$+ p_1 R_1(A20) + p_1 R_2(S21)$$
$$+ p_2 R_1(A20) + p_2 R_2(BUST)$$
$$+ p_3 R_1(S21)$$
$$+ p_4 R_1(BUST)$$

$$= R_0(A19)$$
$$+ p_1 \left(R_1(A20) + R_2(S21) \right)$$
$$+ p_2 \left(R_1(A20) + R_2(BUST) \right)$$
$$+ p_3 R_1(S21)$$
$$+ p_4 R_1(BUST)$$

where: $p_1 = P(A19 \to A20 \to S21)$

$p_2 = P(A19 \to A20 \to BUST)$

$p_3 = P(A19 \to S21)$

$p_4 = P(A19 \to BUST)$

$$U^{\pi}(A19) = R_0(A19) + p_1\left(R_1(A20) + R_2(S21)\right) + p_2\left(R_1(A20) + R_2(\text{BUST})\right) + p_3 R_1(S21) + p_4 R_1(\text{BUST})$$

where:

$$p_1 = P(A19 \to A20 \to S21)$$
$$p_2 = P(A19 \to A20 \to \text{BUST})$$
$$p_3 = P(A19 \to S21)$$
$$p_4 = P(A19 \to \text{BUST})$$

recall:

$$R_t(q) = \gamma R_{t-1}(q)$$

$$U^{\pi}(A19) = R_0(A19)$$
$$+ p_1\left(R_1(A20) + R_2(S21)\right)$$
$$+ p_2\left(R_1(A20) + R_2(BUST)\right)$$
$$+ p_3 R_1(S21)$$
$$+ p_4 R_1(BUST)$$
$$= R_0(A19)$$
$$+ p_1\left(\gamma R_0(A20) + \gamma R_1(S21)\right)$$
$$+ p_2\left(\gamma R_0(A20) + \gamma R_1(BUST)\right)$$
$$+ p_3 \gamma R_0(S21)$$
$$+ p_4 \gamma R_0(BUST)$$

where:
$$p_1 = P(A19 \to A20 \to S21)$$
$$p_2 = P(A19 \to A20 \to BUST)$$
$$p_3 = P(A19 \to S21)$$
$$p_4 = P(A19 \to BUST)$$

$$R_t(q) = \gamma R_{t-1}(q)$$

$$
\begin{aligned}
U^{\pi}(\text{A19}) &= R_0(\text{A19}) \\
&\quad + p_1\left(R_1(\text{A20}) + R_2(\text{S21})\right) \\
&\quad + p_2\left(R_1(\text{A20}) + R_2(\text{BUST})\right) \\
&\quad + p_3 R_1(\text{S21}) \\
&\quad + p_4 R_1(\text{BUST}) \\
&= R_0(\text{A19}) \\
&\quad + p_1\,\gamma\left(R_0(\text{A20}) + R_1(\text{S21})\right) \\
&\quad + p_2\,\gamma\left(R_0(\text{A20}) + R_1(\text{BUST})\right) \\
&\quad + p_3\,\gamma R_0(\text{S21}) \\
&\quad + p_4\,\gamma R_0(\text{BUST})
\end{aligned}
$$

where:

$$
\begin{aligned}
p_1 &= P(\text{A19} \to \text{A20} \to \text{S21}) \\
p_2 &= P(\text{A19} \to \text{A20} \to \text{BUST}) \\
p_3 &= P(\text{A19} \to \text{S21}) \\
p_4 &= P(\text{A19} \to \text{BUST})
\end{aligned}
$$

$$
R_t(q) = \gamma R_{t-1}(q)
$$

$$U^{\pi}(A19) = R_0(A19) + p_1\left(R_1(A20) + R_2(S21)\right) + p_2\left(R_1(A20) + R_2(BUST)\right) + p_3 R_1(S21) + p_4 R_1(BUST)$$

$$= R_0(A19) + \gamma\Big(p_1\left(R_0(A20) + R_1(S21)\right) + p_2\left(R_0(A20) + R_1(BUST)\right) + p_3 R_0(S21) + p_4 R_0(BUST)\Big)$$

where:

$$p_1 = P(A19 \to A20 \to S21)$$
$$p_2 = P(A19 \to A20 \to BUST)$$
$$p_3 = P(A19 \to S21)$$
$$p_4 = P(A19 \to BUST)$$

$$\boxed{R_t(q) = \gamma R_{t-1}(q)}$$

$$U^{\pi}(\text{A19}) = R_0(\text{A19}) + \gamma\Big(p_1\big(R_0(\text{A20}) + R_1(\text{S21})\big) + p_2\big(R_0(\text{A20}) + R_1(\text{BUST})\big) + p_3 R_0(\text{S21}) + p_4 R_0(\text{BUST})\Big)$$

where:

$$p_1 = P(\text{A19} \to \text{A20} \to \text{S21})$$
$$p_2 = P(\text{A19} \to \text{A20} \to \text{BUST})$$
$$p_3 = P(\text{A19} \to \text{S21})$$
$$p_4 = P(\text{A19} \to \text{BUST})$$

$$U^{\pi}(A19) = R_0(A19) + \gamma\left(p_1\left(R_0(A20) + R_1(S21)\right) + p_2\left(R_0(A20) + R_1(\text{BUST})\right) + p_3\, R_0(S21) + p_4\, R_0(\text{BUST})\right)$$

where:

$$p_1 = P(A19 \rightarrow A20 \rightarrow S21)$$
$$p_2 = P(A19 \rightarrow A20 \rightarrow \text{BUST})$$
$$p_3 = P(A19 \rightarrow S21)$$
$$p_4 = P(A19 \rightarrow \text{BUST})$$

$$U^{\pi}(A19) = R_0(A19) + \gamma\left(p_1\left(R_0(A20) + R_1(S21)\right) + p_2\left(R_0(A20) + R_1(BUST)\right) + p_3 R_0(S21) + p_4 R_0(BUST)\right)$$

$$p_1\left(R_0(A20) + R_1(S21)\right) + p_2\left(R_0(A20) + R_1(BUST)\right) + p_3 R_0(S21) + p_4 R_0(BUST)$$

where:

$$p_1 = P(A19 \to A20 \to S21)$$
$$p_2 = P(A19 \to A20 \to BUST)$$
$$p_3 = P(A19 \to S21)$$
$$p_4 = P(A19 \to BUST)$$

$$U^{\pi}(A19) = R_0(A19) + \gamma\left(p_1\left(R_0(A20) + R_1(S21)\right) + p_2\left(R_0(A20) + R_1(BUST)\right) + p_3 R_0(S21) + p_4 R_0(BUST)\right)$$

$$p_1\left(R_0(A20) + R_1(S21)\right) + p_2\left(R_0(A20) + R_1(BUST)\right) + p_3 R_0(S21) + p_4 R_0(BUST)$$

$$\begin{aligned} = \; & P(A19 \to A20 \to S21)\left(R_0(A20) + R_1(S21)\right) \\ & + P(A19 \to A20 \to BUST)\left(R_0(A20) + R_1(BUST)\right) \\ & + P(A19 \to S21) R_0(S21) \\ & + P(A19 \to BUST) R_0(BUST) \end{aligned}$$

$$U^{\pi}(\text{A19}) = R_0(\text{A19}) + \gamma\left(p_1\left(R_0(\text{A20}) + R_1(\text{S21})\right) + p_2\left(R_0(\text{A20}) + R_1(\text{BUST})\right) + p_3 R_0(\text{S21}) + p_4 R_0(\text{BUST})\right)$$

$$p_1\left(R_0(\text{A20}) + R_1(\text{S21})\right) + p_2\left(R_0(\text{A20}) + R_1(\text{BUST})\right) + p_3 R_0(\text{S21}) + p_4 R_0(\text{BUST})$$

$$= P(\text{A19} \to \text{A20} \to \text{S21})\left(R_0(\text{A20}) + R_1(\text{S21})\right) + P(\text{A19} \to \text{A20} \to \text{BUST})\left(R_0(\text{A20}) + R_1(\text{BUST})\right) + P(\text{A19} \to \text{S21}) R_0(\text{S21}) + P(\text{A19} \to \text{BUST}) R_0(\text{BUST})$$

$$= P(\text{A19} \to \text{A20}) P(\text{A20} \to \text{S21})\left(R_0(\text{A20}) + R_1(\text{S21})\right) + P(\text{A19} \to \text{A20}) P(\text{A20} \to \text{BUST})\left(R_0(\text{A20}) + R_1(\text{BUST})\right) + P(\text{A19} \to \text{S21}) R_0(\text{S21}) + P(\text{A19} \to \text{BUST}) R_0(\text{BUST})$$

$$P(\text{BUST} \mid \text{A20}, \text{HIT})$$

$$U^{\pi}(A19) = R_0(A19) + \gamma\left(p_1\left(R_0(A20) + R_1(S21)\right) + p_2\left(R_0(A20) + R_1(BUST)\right) + p_3 R_0(S21) + p_4 R_0(BUST)\right)$$

$$p_1\left(R_0(A20) + R_1(S21)\right) + p_2\left(R_0(A20) + R_1(BUST)\right) + p_3 R_0(S21) + p_4 R_0(BUST)$$

$$\begin{aligned} &= P(A19 \to A20 \to S21)\left(R_0(A20) + R_1(S21)\right) \\ &+ P(A19 \to A20 \to BUST)\left(R_0(A20) + R_1(BUST)\right) \\ &+ P(A19 \to S21) R_0(S21) \\ &+ P(A19 \to BUST) R_0(BUST) \end{aligned}$$

$$\begin{aligned} &= P(A19 \to A20) P(A20 \to S21)\left(R_0(A20) + R_1(S21)\right) \\ &+ P(A19 \to A20) P(A20 \to BUST)\left(R_0(A20) + R_1(BUST)\right) \\ &+ P(A19 \to S21) R_0(S21) \\ &+ P(A19 \to BUST) R_0(BUST) \end{aligned}$$

$$P\left(BUST \mid A20, \pi(A20)\right)$$

$$U^{\pi}(A19) = R_0(A19) + \gamma\left(p_1\left(R_0(A20) + R_1(S21)\right) + p_2\left(R_0(A20) + R_1(\text{BUST})\right) + p_3 R_0(S21) + p_4 R_0(\text{BUST})\right)$$

$$p_1\left(R_0(A20) + R_1(S21)\right) + p_2\left(R_0(A20) + R_1(\text{BUST})\right) + p_3 R_0(S21) + p_4 R_0(\text{BUST})$$

$$\begin{aligned} = \; & P(A19 \to A20)\,P(A20 \to S21)\left(R_0(A20) + R_1(S21)\right) \\ & + P(A19 \to A20)\,P(A20 \to \text{BUST})\left(R_0(A20) + R_1(\text{BUST})\right) \\ & + P(A19 \to S21)\,R_0(S21) \\ & + P(A19 \to \text{BUST})\,R_0(\text{BUST}) \end{aligned}$$

$$U^{\pi}(A19) = R_0(A19) + \gamma\left(p_1\left(R_0(A20) + R_1(S21)\right) + p_2\left(R_0(A20) + R_1(BUST)\right) + p_3 R_0(S21) + p_4 R_0(BUST)\right)$$

$$p_1\left(R_0(A20) + R_1(S21)\right) + p_2\left(R_0(A20) + R_1(BUST)\right) + p_3 R_0(S21) + p_4 R_0(BUST)$$

$$= \begin{aligned} &P(A19 \to A20)P(A20 \to S21)\left(R_0(A20) + R_1(S21)\right) \\ &+ P(A19 \to A20)P(A20 \to BUST)\left(R_0(A20) + R_1(BUST)\right) \\ &+ P(A19 \to S21)R_0(S21) \\ &+ P(A19 \to BUST)R_0(BUST) \end{aligned} = \begin{aligned} &P(A19 \to A20)\Big(P(A20 \to S21)\left(R_0(A20) + R_1(S21)\right) \\ &\quad + P(A20 \to BUST)\left(R_0(A20) + R_1(BUST)\right)\Big) \\ &+ P(A19 \to S21)R_0(S21) \\ &+ P(A19 \to BUST)R_0(BUST) \end{aligned}$$

$$U^{\pi}(A19) = R_0(A19) + \gamma\left(p_1\left(R_0(A20) + R_1(S21)\right) + p_2\left(R_0(A20) + R_1(\text{BUST})\right) + p_3 R_0(S21) + p_4 R_0(\text{BUST})\right)$$

$$p_1\left(R_0(A20) + R_1(S21)\right) + p_2\left(R_0(A20) + R_1(\text{BUST})\right) + p_3 R_0(S21) + p_4 R_0(\text{BUST})$$

$$= P(A19 \to A20)\left(P(A20 \to S21)\left(R_0(A20) + R_1(S21)\right) + P(A20 \to \text{BUST})\left(R_0(A20) + R_1(\text{BUST})\right)\right) + P(A19 \to S21) R_0(S21) + P(A19 \to \text{BUST}) R_0(\text{BUST})$$

$$U^{\pi}(A19) = R_0(A19) + \gamma\left(p_1\left(R_0(A20) + R_1(S21)\right) + p_2\left(R_0(A20) + R_1(BUST)\right) + p_3 R_0(S21) + p_4 R_0(BUST)\right)$$

$$p_1\left(R_0(A20) + R_1(S21)\right) + p_2\left(R_0(A20) + R_1(BUST)\right) + p_3 R_0(S21) + p_4 R_0(BUST)$$

$$= P(A19 \to A20)\left(P(A20 \to S21)\left(R_0(A20) + R_1(S21)\right) + P(A20 \to BUST)\left(R_0(A20) + R_1(BUST)\right)\right) + P(A19 \to S21) R_0(S21) + P(A19 \to BUST) R_0(BUST)$$

the lifetime reward of $A20 \to BUST$

$$= P(A19 \to A20)\left(P(A20 \to S21)\, U(A20 \to S21) + P(A20 \to BUST)\, U(A20 \to BUST)\right) + P(A19 \to S21) R_0(S21) + P(A19 \to BUST) R_0(BUST)$$

$$U^{\pi}(A19) = R_0(A19) + \gamma\left(p_1\left(R_0(A20) + R_1(S21)\right) + p_2\left(R_0(A20) + R_1(\text{BUST})\right) + p_3 R_0(S21) + p_4 R_0(\text{BUST})\right)$$

$$p_1\left(R_0(A20) + R_1(S21)\right) + p_2\left(R_0(A20) + R_1(\text{BUST})\right) + p_3 R_0(S21) + p_4 R_0(\text{BUST})$$

$$= P(A19 \to A20)\left(P(A20 \to S21)\,U(A20 \to S21) + P(A20 \to \text{BUST})\,U(A20 \to \text{BUST})\right) + P(A19 \to S21)\,R_0(S21) + P(A19 \to \text{BUST})\,R_0(\text{BUST})$$

$$U^{\pi}(A19) = R_0(A19) + \gamma\left(p_1\left(R_0(A20) + R_1(S21)\right) + p_2\left(R_0(A20) + R_1(\text{BUST})\right) + p_3 R_0(S21) + p_4 R_0(\text{BUST})\right)$$

$$p_1\left(R_0(A20) + R_1(S21)\right) + p_2\left(R_0(A20) + R_1(\text{BUST})\right) + p_3 R_0(S21) + p_4 R_0(\text{BUST})$$

$$= P(A19 \to A20)\left(P(A20 \to S21)\,U(A20 \to S21) + P(A20 \to \text{BUST})\,U(A20 \to \text{BUST})\right) + P(A19 \to S21)\,R_0(S21) + P(A19 \to \text{BUST})\,R_0(\text{BUST})$$

← expected utility of policy π in state A20

$$U^{\pi}(A19) = R_0(A19) + \gamma\left(p_1\left(R_0(A20) + R_1(S21)\right) + p_2\left(R_0(A20) + R_1(BUST)\right) + p_3 R_0(S21) + p_4 R_0(BUST)\right)$$

$$p_1\left(R_0(A20) + R_1(S21)\right) + p_2\left(R_0(A20) + R_1(BUST)\right) + p_3 R_0(S21) + p_4 R_0(BUST)$$

$$= P(A19 \to A20)\left(P(A20 \to S21)\,U(A20 \to S21) + P(A20 \to BUST)\,U(A20 \to BUST)\right) \leftarrow U^{\pi}(A20)$$
$$+ P(A19 \to S21)\,R_0(S21)$$
$$+ P(A19 \to BUST)\,R_0(BUST)$$

$$U^{\pi}(A19) = R_0(A19) + \gamma\left(p_1\left(R_0(A20) + R_1(S21)\right) + p_2\left(R_0(A20) + R_1(BUST)\right) + p_3 R_0(S21) + p_4 R_0(BUST)\right)$$

$$p_1\left(R_0(A20) + R_1(S21)\right) + p_2\left(R_0(A20) + R_1(BUST)\right) + p_3 R_0(S21) + p_4 R_0(BUST)$$

$$= P(A19 \to A20)\left(P(A20 \to S21)\,U(A20 \to S21) + P(A20 \to BUST)\,U(A20 \to BUST)\right) + P(A19 \to S21)\,R_0(S21) + P(A19 \to BUST)\,R_0(BUST)$$

$$= P(A19 \to A20)\,U^{\pi}(A20) + P(A19 \to S21)\,R_0(S21) + P(A19 \to BUST)\,R_0(BUST)$$

$$U^{\pi}(A19) = R_0(A19) + \gamma\left(p_1\left(R_0(A20) + R_1(S21)\right) + p_2\left(R_0(A20) + R_1(BUST)\right) + p_3 R_0(S21) + p_4 R_0(BUST)\right)$$

$$p_1\left(R_0(A20) + R_1(S21)\right) + p_2\left(R_0(A20) + R_1(BUST)\right) + p_3 R_0(S21) + p_4 R_0(BUST)$$

$$\begin{aligned} & P(A19 \to A20)\, U^{\pi}(A20) \\ = & + P(A19 \to S21)\, R_0(S21) \\ & + P(A19 \to BUST)\, R_0(BUST) \end{aligned}$$

$$U^{\pi}(A19) = R_0(A19) + \gamma\left(p_1\left(R_0(A20) + R_1(S21)\right) + p_2\left(R_0(A20) + R_1(BUST)\right) + p_3 R_0(S21) + p_4 R_0(BUST)\right)$$

$$p_1\left(R_0(A20) + R_1(S21)\right) + p_2\left(R_0(A20) + R_1(BUST)\right) + p_3 R_0(S21) + p_4 R_0(BUST)$$

$$P(A19 \to A20)\, U^{\pi}(A20)$$

$$= + P(A19 \to S21)\, \boxed{R_0(S21)} \leftarrow \text{expected utility of policy } \pi \text{ in state S21}$$

$$+ P(A19 \to BUST)\, \boxed{R_0(BUST)} \leftarrow \text{expected utility of policy } \pi \text{ in state BUST}$$

$$U^{\pi}(A19) = R_0(A19) + \gamma\left(p_1\left(R_0(A20) + R_1(S21)\right) + p_2\left(R_0(A20) + R_1(BUST)\right) + p_3 R_0(S21) + p_4 R_0(BUST)\right)$$

$$p_1\left(R_0(A20) + R_1(S21)\right) + p_2\left(R_0(A20) + R_1(BUST)\right) + p_3 R_0(S21) + p_4 R_0(BUST)$$

$$= P(A19 \to A20)\, U^{\pi}(A20) + P(A19 \to S21)\, U^{\pi}(S21) + P(A19 \to BUST)\, U^{\pi}(BUST)$$

$$U^{\pi}(A19) = R_0(A19) + \gamma\left(p_1\left(R_0(A20) + R_1(S21)\right) + p_2\left(R_0(A20) + R_1(BUST)\right) + p_3 R_0(S21) + p_4 R_0(BUST)\right)$$

$$p_1\left(R_0(A20) + R_1(S21)\right) + p_2\left(R_0(A20) + R_1(BUST)\right) + p_3 R_0(S21) + p_4 R_0(BUST)$$

$$= P(A19 \to A20)\, U^{\pi}(A20) + P(A19 \to S21)\, U^{\pi}(S21) + P(A19 \to BUST)\, U^{\pi}(BUST)$$

substitute back in

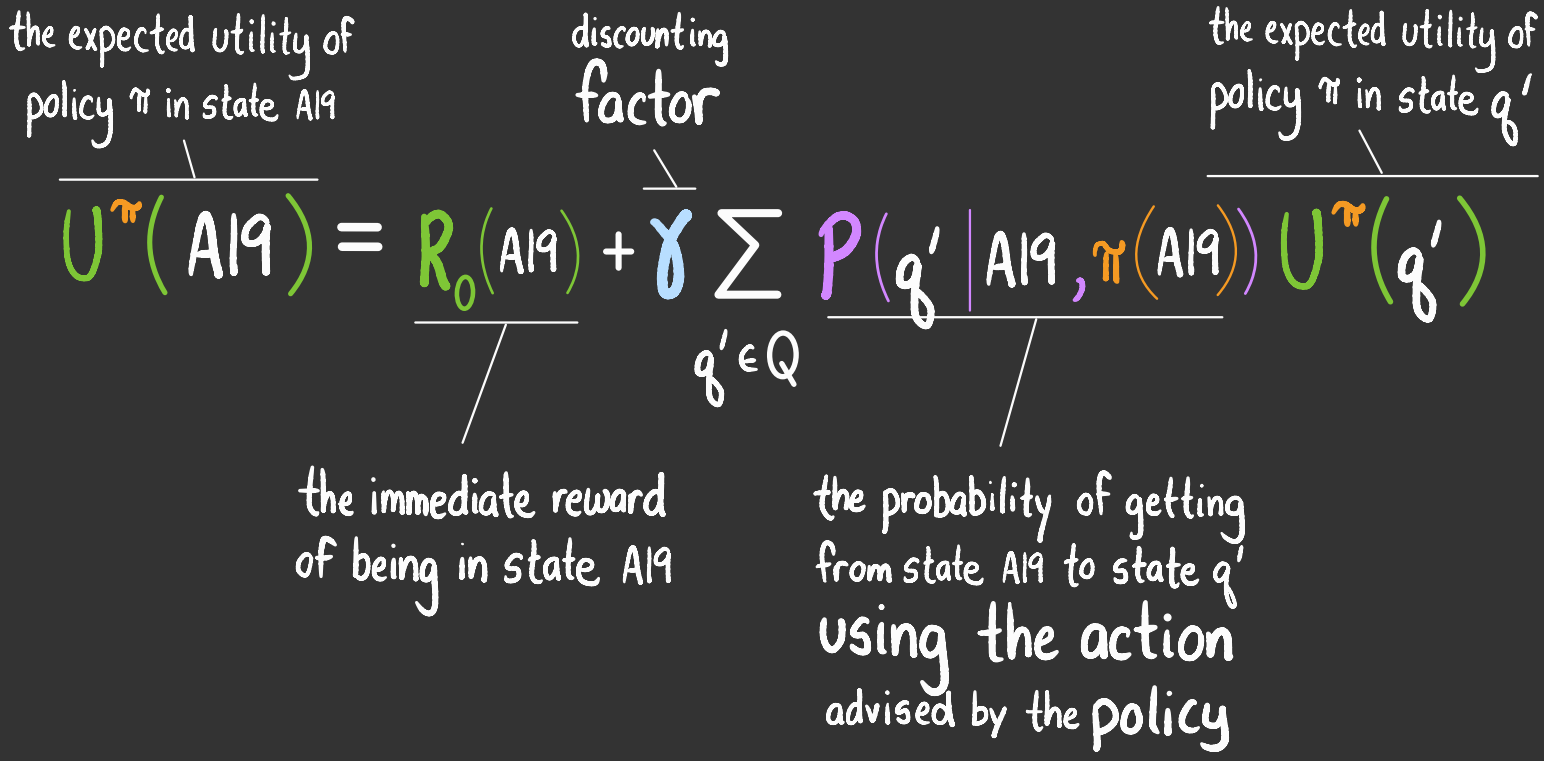
$$U^{\pi}(A19) = R_0(A19) + \gamma \left(P(A19 \to A20)\, U^{\pi}(A20) + P(A19 \to S21)\, U^{\pi}(S21) + P(A19 \to BUST)\, U^{\pi}(BUST) \right)$$

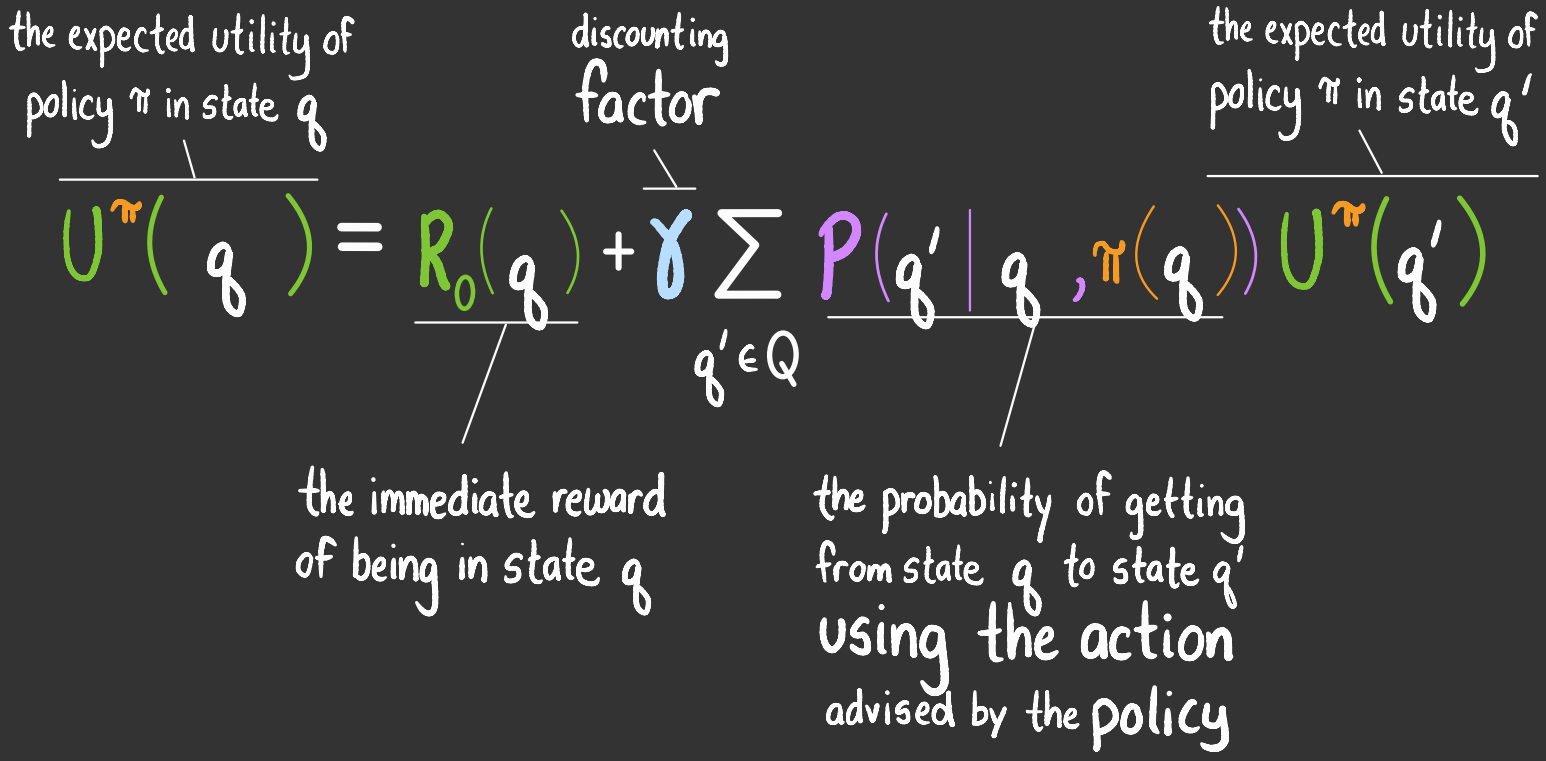
$$U^{\pi}(A19) = R_0(A19) + \gamma\left(P(A19 \to A20)\,U^{\pi}(A20) + P(A19 \to S21)\,U^{\pi}(S21) + P(A19 \to BUST)\,U^{\pi}(BUST)\right)$$

$$= R_0(A19) + \gamma \sum_{q' \in Q} P\left(A19 \xrightarrow{\pi(A19)} q'\right)\; U^{\pi}(q')$$

$$= R_0(A19) + \gamma \sum_{q' \in Q} P(q' \mid A19, \pi(A19))\, U^{\pi}(q')$$

$$U^{\pi}(A19) = R_0(A19) + \gamma \sum_{q' \in Q} P(q' \mid A19, \pi(A19)) U^{\pi}(q')$$


the expected utility of policy π in state A19
discounting factor
the expected utility of policy π in state q'
$U^{\pi}(A19) = R_0(A19) + \gamma \sum_{q' \in Q} P(q' \mid A19, \pi(A19)) U^{\pi}(q')$
the immediate reward of being in state A19
the probability of getting from state A19 to state q' using the action advised by the policy


the expected utility of policy π in state q
discounting factor
the expected utility of policy π in state q'
$U^{\pi}(q) = R_0(q) + \gamma \sum_{q' \in Q} P(q' \mid q, \pi(q)) U^{\pi}(q')$
the immediate reward of being in state q
the probability of getting from state q to state q' using the action advised by the policy

$$U^{\pi}(q) = R_0(q) + \gamma \sum_{q' \in Q} P(q' \mid q, \pi(q)) U^{\pi}(q')$$

this is the expected utility of policy π if we start in state q