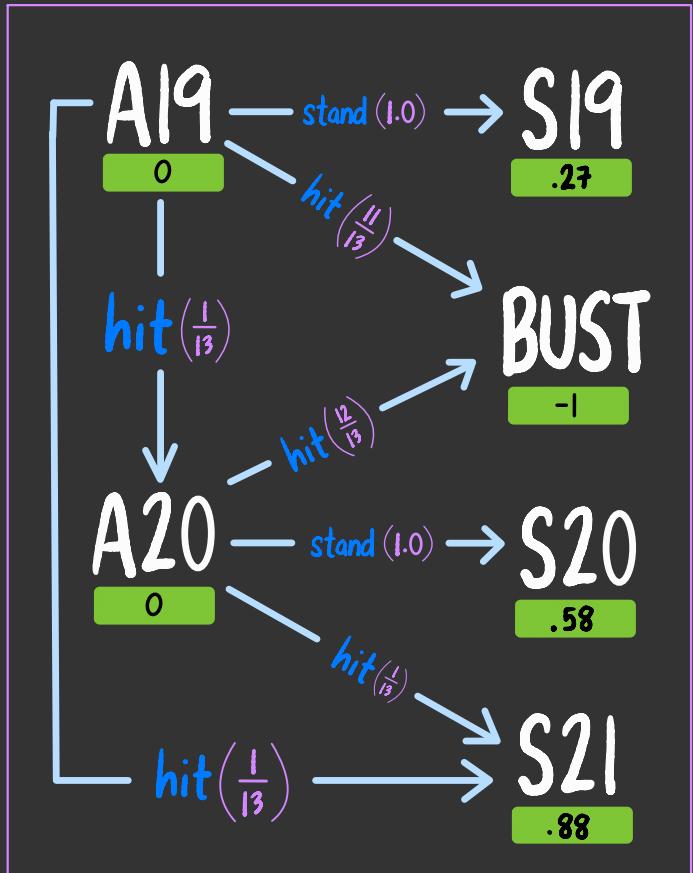


policies

CSCI  
373

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

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

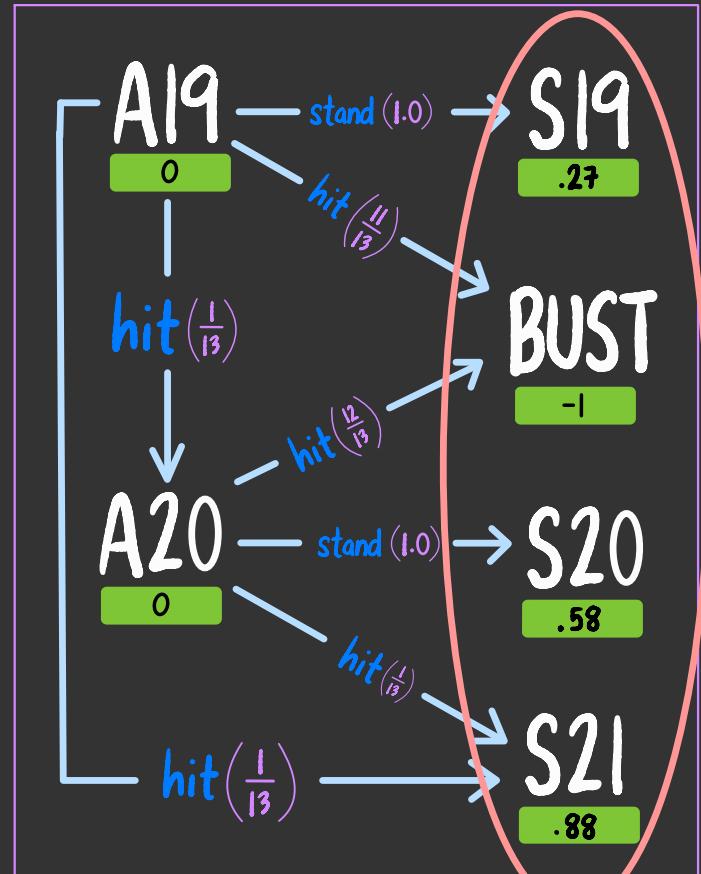


state machine

the answer to this question is called a

policy

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

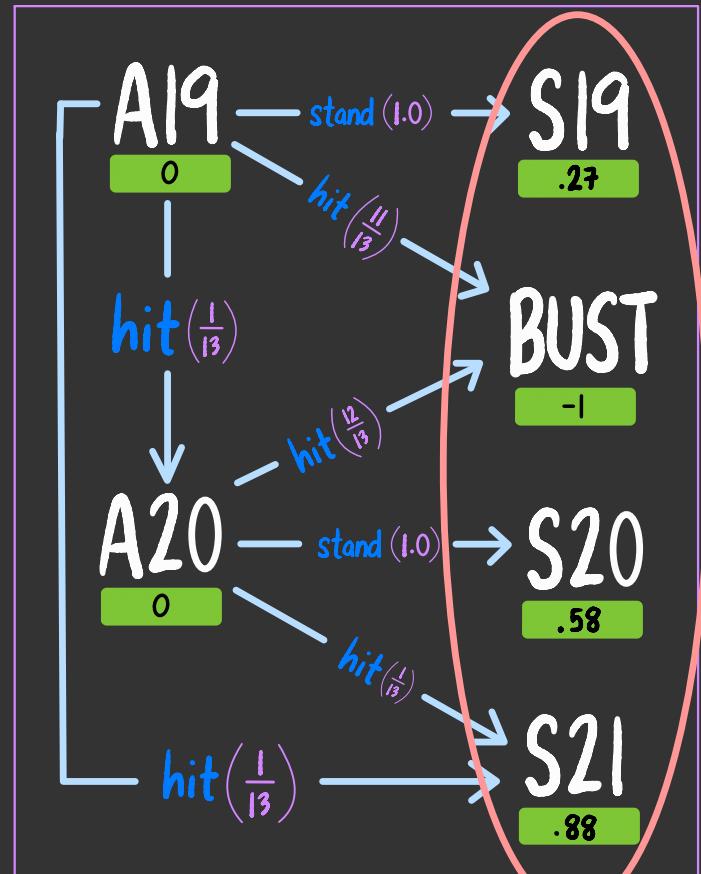


final states do not need  
a recommended action

a policy is just a function

$\pi: (Q \setminus F) \rightarrow \Sigma$  from the  
non-final states to actions

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



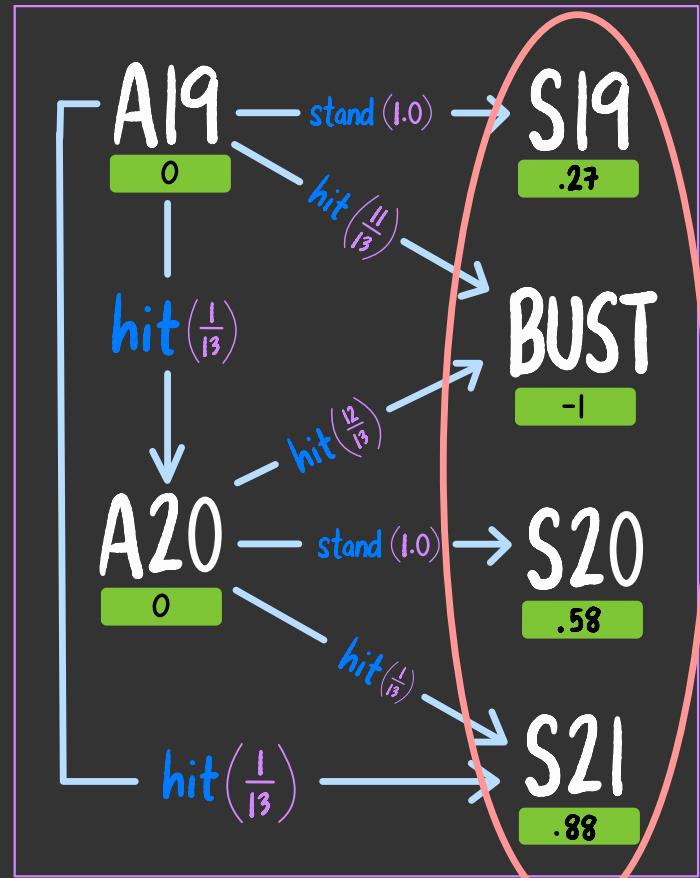
final states do not need  
a recommended action

a policy is just a function

$\pi: (Q \setminus F) \rightarrow \Sigma$  from the  
non-final states to actions

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

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

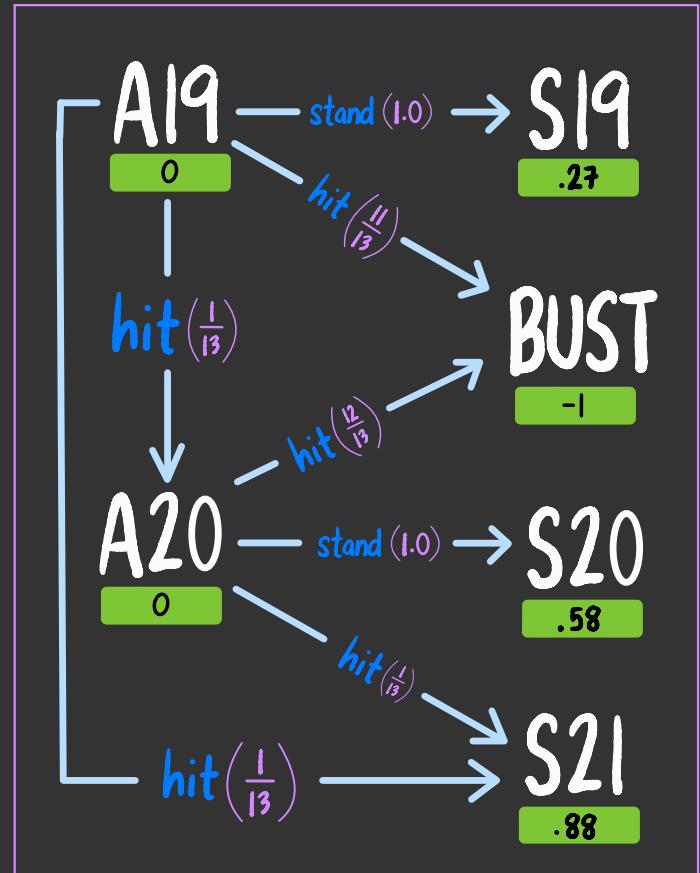


final states do not need  
a recommended action

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

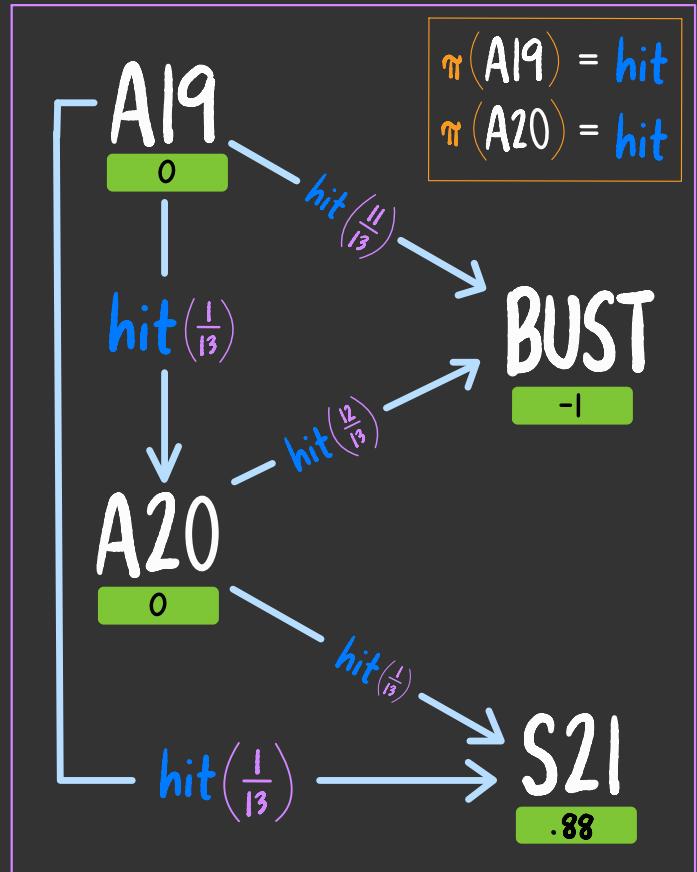
$$\begin{aligned}\pi(A19) &= \text{hit} \\ \pi(A20) &= \text{hit}\end{aligned}$$

how good  
is this?



markov decision process (mdp)

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



markov decision process (mdp)

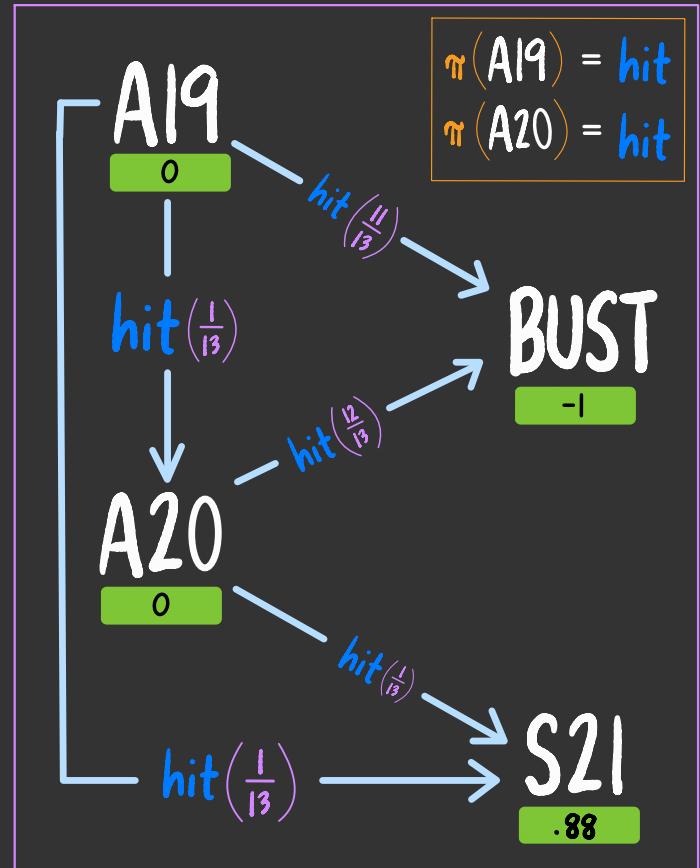
the policy permits  
4 possible futures

A19 → A20 → S21

A19 → A20 → BUST

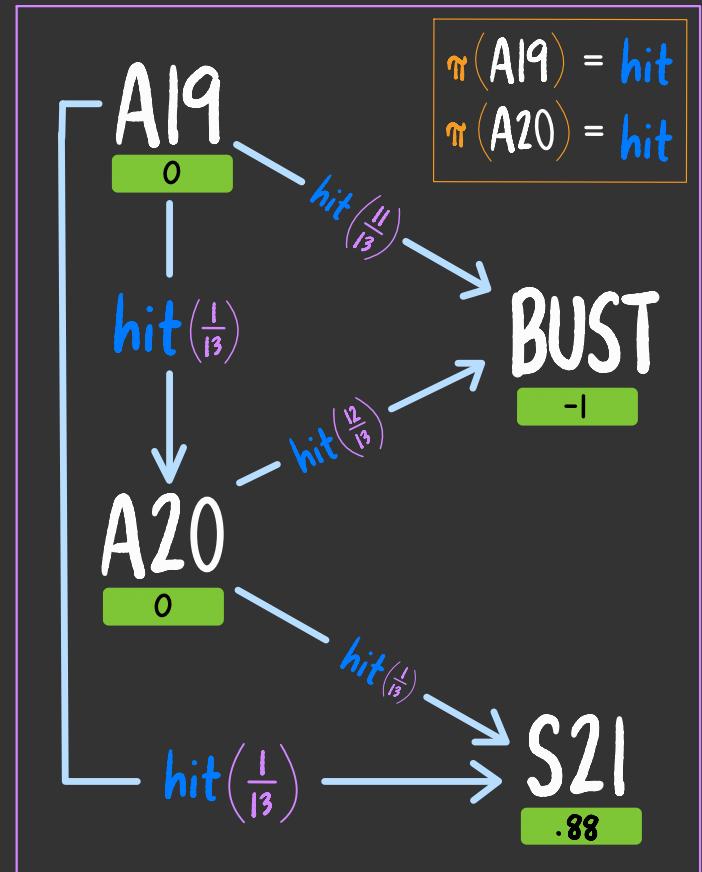
A19 → S21

A19 → BUST



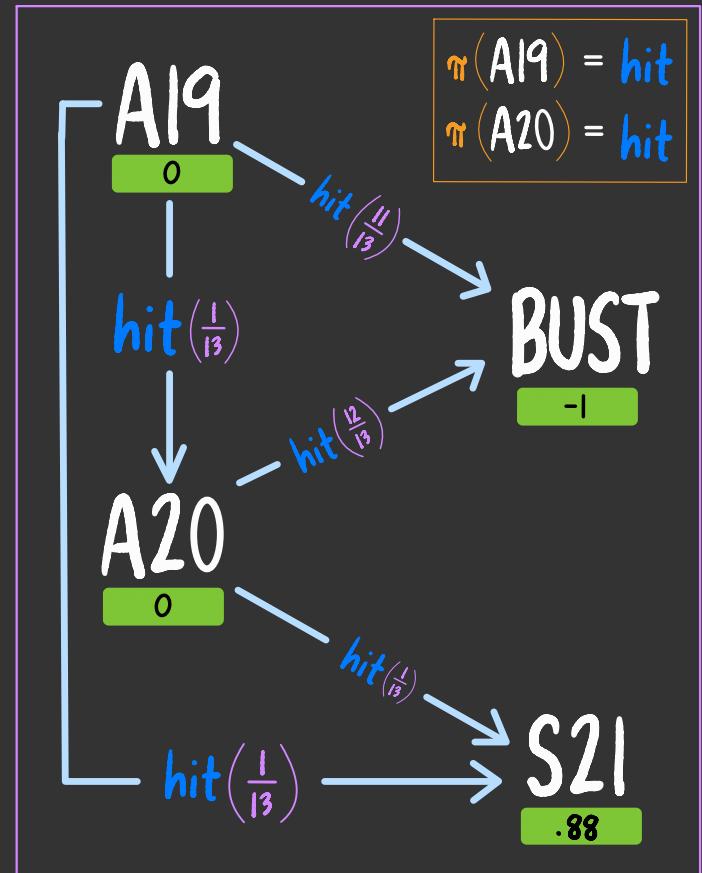
markov decision process (mdp)

future	lifetime reward	probability
$A19 \rightarrow A20 \rightarrow S21$	?	?
$A19 \rightarrow A20 \rightarrow \text{BUST}$	?	?
$A19 \rightarrow S21$	?	?
$A19 \rightarrow \text{BUST}$	?	?



markov decision process (mdp)

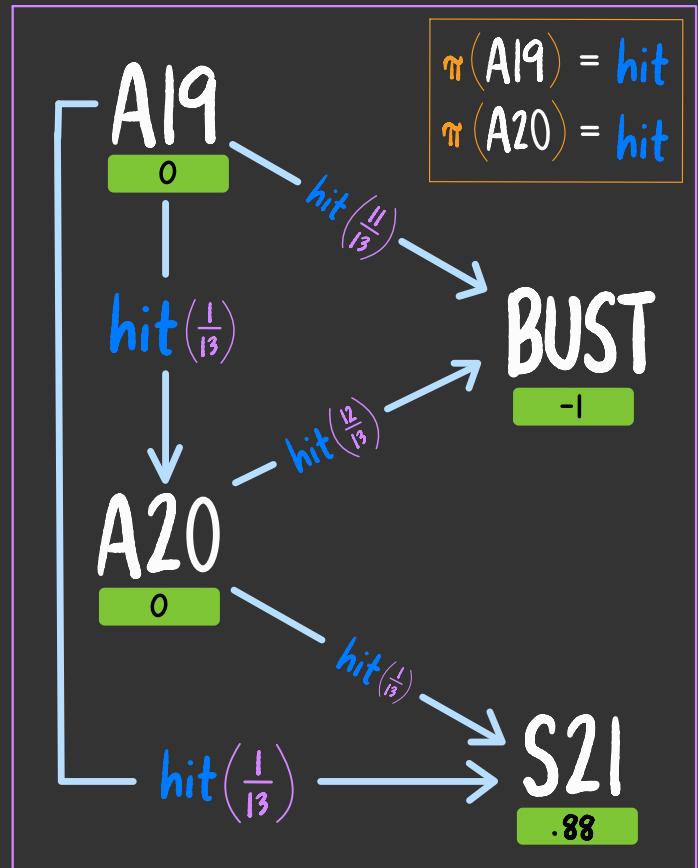
future	lifetime reward	probability
$A19 \rightarrow A20 \rightarrow S21$	.88	$\frac{1}{13} \cdot \frac{1}{13}$
$A19 \rightarrow A20 \rightarrow \text{BUST}$	-1	$\frac{1}{13} \cdot \frac{12}{13}$
$A19 \rightarrow S21$	.88	$\frac{1}{13}$
$A19 \rightarrow \text{BUST}$	-1	$\frac{11}{13}$



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 × $\frac{11}{13}$	

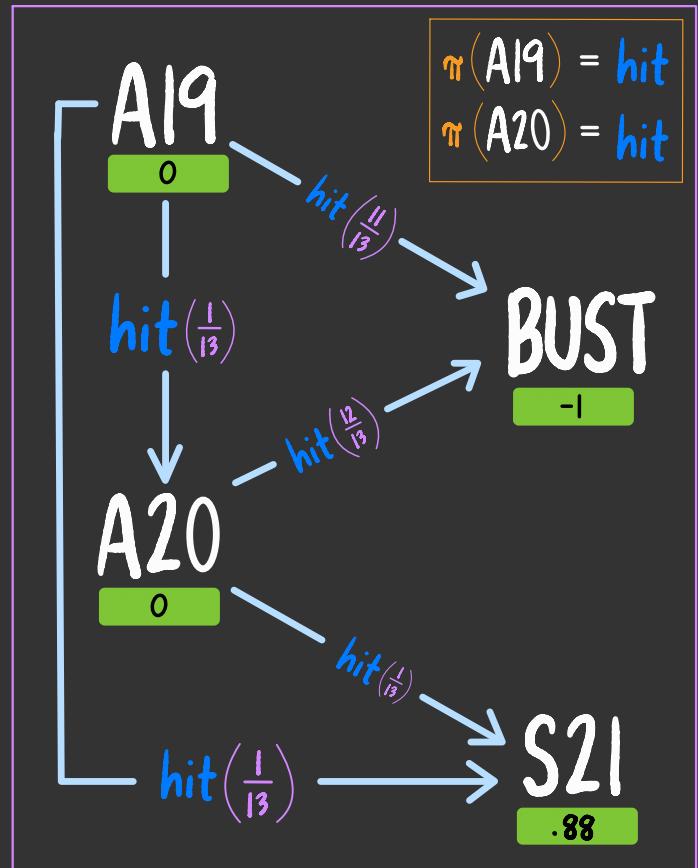
expected lifetime reward =  $-.84$   
(approx.)



markov decision process (mdp)

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

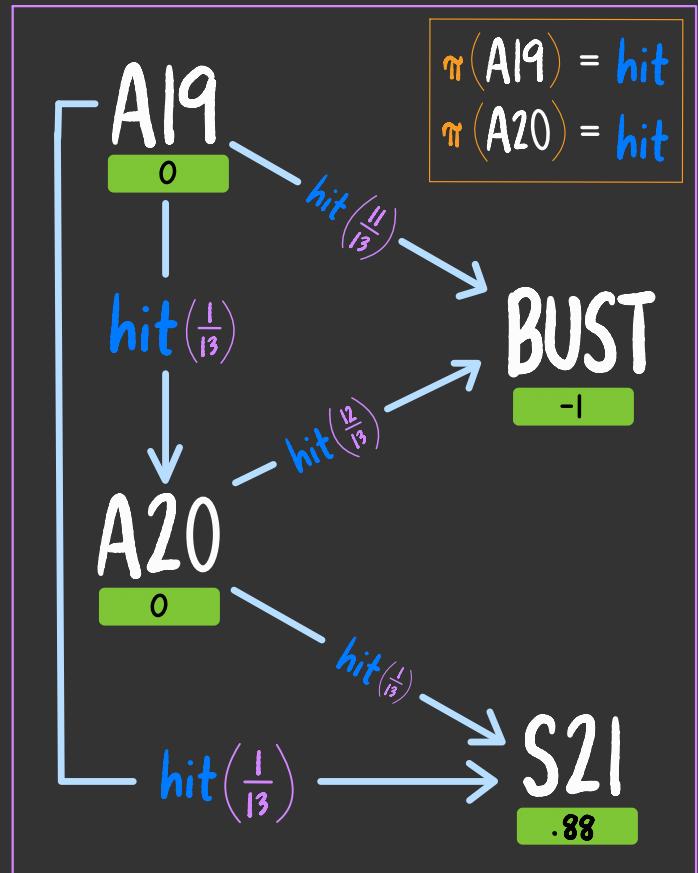
$$\text{expected utility} = -.84 \quad (\text{approx.})$$



markov decision process (mdp)

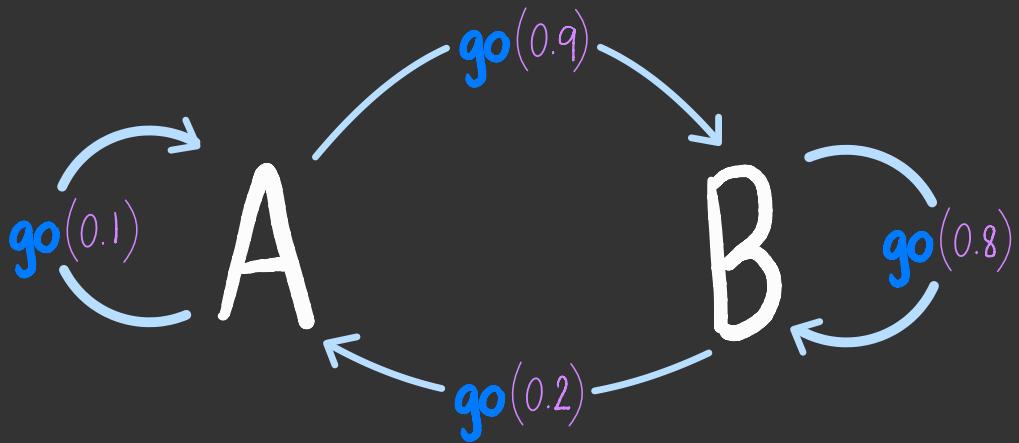
future	utility	probability
$A19 \rightarrow A20 \rightarrow S21$	.88	$\times \frac{1}{13} \cdot \frac{1}{13}$
$A19 \rightarrow A20 \rightarrow \text{BUST}$	+ -1	$\times \frac{1}{13} \cdot \frac{12}{13}$
$A19 \rightarrow S21$	+ .88	$\times \frac{1}{13}$
$A19 \rightarrow \text{BUST}$	+ -1	$\times \frac{11}{13}$

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



markov decision process (mdp)

but for some mdps, this is not so straightforward



how would we compute the  
expected utility of state A?

$$\begin{aligned}
 U^\pi(A19) = & \quad \bigcup(A19 \rightarrow A20 \rightarrow S21) P(A19 \rightarrow A20 \rightarrow S21) \\
 & + \bigcup(A19 \rightarrow A20 \rightarrow \text{BUST}) P(A19 \rightarrow A20 \rightarrow \text{BUST}) \\
 & + \bigcup(A19 \rightarrow S21) P(A19 \rightarrow S21) \\
 & + \bigcup(A19 \rightarrow \text{BUST}) P(A19 \rightarrow \text{BUST})
 \end{aligned}$$

future	utility	probability
$A19 \rightarrow A20 \rightarrow S21$	.88	$\times \frac{1}{13} \cdot \frac{1}{13}$
$A19 \rightarrow A20 \rightarrow \text{BUST}$	+ -1	$\times \frac{1}{13} \cdot \frac{12}{13}$
$A19 \rightarrow S21$	+ .88	$\times \frac{1}{13}$
$A19 \rightarrow \text{BUST}$	+ -1	$\times \frac{11}{13}$
<hr/>		
$U^\pi(A19)$	=	- .84
		(approx.)

$$\begin{aligned}
 U(\text{A19}) = & \cup (\text{A19} \rightarrow \text{A20} \rightarrow \text{S21}) P(\text{A19} \rightarrow \text{A20} \rightarrow \text{S21}) \\
 & + \cup (\text{A19} \rightarrow \text{A20} \rightarrow \text{BUST}) P(\text{A19} \rightarrow \text{A20} \rightarrow \text{BUST}) \\
 & + \cup (\text{A19} \rightarrow \text{S21}) P(\text{A19} \rightarrow \text{S21}) \\
 & + \cup (\text{A19} \rightarrow \text{BUST}) P(\text{A19} \rightarrow \text{BUST})
 \end{aligned}$$

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

$$\begin{aligned} U(\text{A1q}) &= P_1 \cup (\text{A1q} \rightarrow \text{A20} \rightarrow \text{S21}) \\ &+ P_2 \cup (\text{A1q} \rightarrow \text{A20} \rightarrow \text{BUST}) \\ &+ P_3 \cup (\text{A1q} \rightarrow \text{S21}) \\ &+ P_4 \cup (\text{A1q} \rightarrow \text{BUST}) \end{aligned}$$

where:  $P_1 = P(\text{A1q} \rightarrow \text{A20} \rightarrow \text{S21})$   
 $P_2 = P(\text{A1q} \rightarrow \text{A20} \rightarrow \text{BUST})$   
 $P_3 = P(\text{A1q} \rightarrow \text{S21})$   
 $P_4 = P(\text{A1q} \rightarrow \text{BUST})$

$$\begin{aligned}
 U(\text{A1q}) &= p_1 U(\text{A1q} \rightarrow \text{A20} \rightarrow \text{S21}) \\
 &\quad + p_2 U(\text{A1q} \rightarrow \text{A20} \rightarrow \text{BUST}) \\
 &\quad + p_3 U(\text{A1q} \rightarrow \text{S21}) \\
 &\quad + p_4 U(\text{A1q} \rightarrow \text{BUST}) \\
 \\ 
 &= p_1 \left( R_0(\text{A1q}) + R_1(\text{A20}) + R_2(\text{S21}) \right) \\
 &\quad + p_2 \left( R_0(\text{A1q}) + R_1(\text{A20}) + R_2(\text{BUST}) \right) \\
 &\quad + p_3 \left( R_0(\text{A1q}) + R_1(\text{S21}) \right) \\
 &\quad + p_4 \left( R_0(\text{A1q}) + R_1(\text{BUST}) \right)
 \end{aligned}$$

where:

- $p_1 = P(\text{A1q} \rightarrow \text{A20} \rightarrow \text{S21})$
- $p_2 = P(\text{A1q} \rightarrow \text{A20} \rightarrow \text{BUST})$
- $p_3 = P(\text{A1q} \rightarrow \text{S21})$
- $p_4 = P(\text{A1q} \rightarrow \text{BUST})$

recall:

$$R_2(\text{S21}) = \gamma R_1(\text{S21})$$

$$\begin{aligned}
 U(\text{A1q}) = & p_1 (R_0(\text{A1q}) + R_1(\text{A20}) + R_2(\text{S2I})) \\
 & + p_2 (R_0(\text{A1q}) + R_1(\text{A20}) + R_2(\text{BUST})) \\
 & + p_3 (R_0(\text{A1q}) + R_1(\text{S2I})) \\
 & + p_4 (R_0(\text{A1q}) + R_1(\text{BUST}))
 \end{aligned}$$

where:  $p_1 = P(\text{A1q} \rightarrow \text{A20} \rightarrow \text{S2I})$   
 $p_2 = P(\text{A1q} \rightarrow \text{A20} \rightarrow \text{BUST})$   
 $p_3 = P(\text{A1q} \rightarrow \text{S2I})$   
 $p_4 = P(\text{A1q} \rightarrow \text{BUST})$

$$\begin{aligned}
U(\text{A1q}) &= p_1 \left( R_0(\text{A1q}) + R_1(\text{A20}) + R_2(\text{S2I}) \right) \quad \text{where: } p_1 = P(\text{A1q} \rightarrow \text{A20} \rightarrow \text{S2I}) \\
&\quad + p_2 \left( R_0(\text{A1q}) + R_1(\text{A20}) + R_2(\text{BUST}) \right) \\
&\quad + p_3 \left( R_0(\text{A1q}) + R_1(\text{S2I}) \right) \\
&\quad + p_4 \left( R_0(\text{A1q}) + R_1(\text{BUST}) \right) \\
&= p_1 R_0(\text{A1q}) + p_1 R_1(\text{A20}) + p_1 R_2(\text{S2I}) \\
&\quad + p_2 R_0(\text{A1q}) + p_2 R_1(\text{A20}) + p_2 R_2(\text{BUST}) \\
&\quad + p_3 R_0(\text{A1q}) + p_3 R_1(\text{S2I}) \\
&\quad + p_4 R_0(\text{A1q}) + p_4 R_1(\text{BUST})
\end{aligned}$$

$$\begin{aligned}
 U(\text{A1q}) &= p_1 R_0(\text{A1q}) + p_1 R_1(\text{A20}) + p_1 R_2(\text{S21}) \quad \text{where: } p_1 = P(\text{A1q} \rightarrow \text{A20} \rightarrow \text{S21}) \\
 &\quad + p_2 R_0(\text{A1q}) + p_2 R_1(\text{A20}) + p_2 R_2(\text{BUST}) \quad p_2 = P(\text{A1q} \rightarrow \text{A20} \rightarrow \text{BUST}) \\
 &\quad + p_3 R_0(\text{A1q}) + p_3 R_1(\text{S21}) \quad p_3 = P(\text{A1q} \rightarrow \text{S21}) \\
 &\quad + p_4 R_0(\text{A1q}) + p_4 R_1(\text{BUST}) \quad p_4 = P(\text{A1q} \rightarrow \text{BUST})
 \end{aligned}$$

$$U^*(A1q) = \boxed{p_1 R_0(A1q) + p_1 R_1(A20) + p_1 R_2(S21)} \quad \text{where: } p_1 = P(A1q \rightarrow A20 \rightarrow S21) \\ + p_2 R_0(A1q) + p_2 R_1(A20) + p_2 R_2(BUST) \quad p_2 = P(A1q \rightarrow A20 \rightarrow BUST) \\ + p_3 R_0(A1q) + p_3 R_1(S21) \quad p_3 = P(A1q \rightarrow S21) \\ + p_4 R_0(A1q) + p_4 R_1(BUST) \quad p_4 = P(A1q \rightarrow BUST)$$

$$= \boxed{p_1 R_0(A1q) + p_2 R_0(A1q) + p_3 R_0(A1q) + p_4 R_0(A1q)}$$

$$+ p_1 R_1(A20) + p_1 R_2(S21) \\ + p_2 R_1(A20) + p_2 R_2(BUST) \\ + p_3 R_1(S21) \\ + p_4 R_1(BUST)$$

$$\begin{aligned}
U(A|q) &= p_1 R_0(A|q) + p_1 R_1(A20) + p_1 R_2(S21) \quad \text{where: } p_1 = P(A|q \rightarrow A20 \rightarrow S21) \\
&\quad + p_2 R_0(A|q) + p_2 R_1(A20) + p_2 R_2(BUST) \quad p_2 = P(A|q \rightarrow A20 \rightarrow BUST) \\
&\quad + p_3 R_0(A|q) + p_3 R_1(S21) \quad p_3 = P(A|q \rightarrow S21) \\
&\quad + p_4 R_0(A|q) + p_4 R_1(BUST) \quad p_4 = P(A|q \rightarrow BUST) \\
&= (p_1 + p_2 + p_3 + p_4) R_0(A|q) \\
&\quad + p_1 R_1(A20) + p_1 R_2(S21) \\
&\quad + p_2 R_1(A20) + p_2 R_2(BUST) \\
&\quad + p_3 R_1(S21) \\
&\quad + p_4 R_1(BUST)
\end{aligned}$$

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

where:

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

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

where:

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

---

?

what do these sum to?

$$U^*(A19) = (p_1 + p_2 + p_3 + p_4) R_0(A19)$$

$$+ p_1 R_1(A20) + p_1 R_2(S21)$$

$$+ p_2 R_1(A20) + p_2 R_2(BUST)$$

$$+ p_3 R_1(S21)$$

$$+ p_4 R_1(BUST)$$

$$= R_0(A19)$$

$$+ p_1 R_1(A20) + p_1 R_2(S21)$$

$$+ p_2 R_1(A20) + p_2 R_2(BUST)$$

$$+ p_3 R_1(S21)$$

$$+ p_4 R_1(BUST)$$

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

+  $p_2 = P(A19 \rightarrow A20 \rightarrow BUST)$

+  $p_3 = P(A19 \rightarrow S21)$

+  $p_4 = P(A19 \rightarrow BUST)$

1

$$U^*(A|q) = R_0(A|q)$$

$$+ p_1 R_1(A20) + p_1 R_2(S21)$$

$$+ p_2 R_1(A20) + p_2 R_2(BUST)$$

$$+ p_3 R_1(S21)$$

$$+ p_4 R_1(BUST)$$

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

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

$p_3 = P(A19 \rightarrow S21)$

$p_4 = P(A19 \rightarrow BUST)$

$$\begin{aligned}
 U(A|q) &= R_0(A|q) \\
 &\quad + p_1 R_1(A20) + p_1 R_2(S21) \\
 &\quad + p_2 R_1(A20) + p_2 R_2(BUST) \\
 &\quad + p_3 R_1(S21) \\
 &\quad + p_4 R_1(BUST) \\
 \\ 
 &= R_0(A|q) \\
 &\quad + p_1 (R_1(A20) + R_2(S21)) \\
 &\quad + p_2 (R_1(A20) + R_2(BUST)) \\
 &\quad + p_3 R_1(S21) \\
 &\quad + p_4 R_1(BUST)
 \end{aligned}$$

where:  $p_1 = P(A|q \rightarrow A20 \rightarrow S21)$   
 $p_2 = P(A|q \rightarrow A20 \rightarrow BUST)$   
 $p_3 = P(A|q \rightarrow S21)$   
 $p_4 = P(A|q \rightarrow BUST)$

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

where:

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

recall:

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

$$\begin{aligned}
U(\text{A1q}) &= R_0(\text{A1q}) \\
&\quad + p_1 (R_1(\text{A20}) + R_2(\text{S21})) \\
&\quad + p_2 (R_1(\text{A20}) + R_2(\text{BUST})) \\
&\quad + p_3 R_1(\text{S21}) \\
&\quad + p_4 R_1(\text{BUST}) \\
&= R_0(\text{A1q}) \\
&\quad + p_1 (\gamma R_0(\text{A20}) + \gamma R_1(\text{S21})) \\
&\quad + p_2 (\gamma R_0(\text{A20}) + \gamma R_1(\text{BUST})) \\
&\quad + p_3 \gamma R_0(\text{S21}) \\
&\quad + p_4 \gamma R_0(\text{BUST})
\end{aligned}$$

where :  $p_1 = P(\text{A1q} \rightarrow \text{A20} \rightarrow \text{S21})$   
 $p_2 = P(\text{A1q} \rightarrow \text{A20} \rightarrow \text{BUST})$   
 $p_3 = P(\text{A1q} \rightarrow \text{S21})$   
 $p_4 = P(\text{A1q} \rightarrow \text{BUST})$

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

$$\begin{aligned}
U^*(A1q) &= R_0(A1q) \\
&\quad + p_1 (R_1(A20) + R_2(S21)) \\
&\quad + p_2 (R_1(A20) + R_2(BUST)) \\
&\quad + p_3 R_1(S21) \\
&\quad + p_4 R_1(BUST) \\
&= R_0(A1q) \\
&\quad + p_1 \gamma (R_0(A20) + R_1(S21)) \\
&\quad + p_2 \gamma (R_0(A20) + R_1(BUST)) \\
&\quad + p_3 \gamma R_0(S21) \\
&\quad + p_4 \gamma R_0(BUST)
\end{aligned}$$

where :  $p_1 = P(A1q \rightarrow A20 \rightarrow S21)$   
 $p_2 = P(A1q \rightarrow A20 \rightarrow BUST)$   
 $p_3 = P(A1q \rightarrow S21)$   
 $p_4 = P(A1q \rightarrow BUST)$

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

$$\begin{aligned}
U^*(A|q) &= R_0(A|q) \\
&\quad + p_1 (R_1(A20) + R_2(S21)) \\
&\quad + p_2 (R_1(A20) + R_2(BUST)) \\
&\quad + p_3 R_1(S21) \\
&\quad + p_4 R_1(BUST) \\
&= R_0(A|q) \\
&\quad + \gamma (p_1 (R_0(A20) + R_1(S21)) \\
&\quad + p_2 (R_0(A20) + R_1(BUST)) \\
&\quad + p_3 R_0(S21) \\
&\quad + p_4 R_0(BUST))
\end{aligned}$$

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

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

$$U^*(A|q) = R_0(A|q) + \gamma \left( p_1 (R_0(A20) + R_1(S21)) + p_2 (R_0(A20) + R_1(BUST)) + p_3 R_0(S21) + p_4 R_0(BUST) \right)$$

where :  $p_1 = P(A|q \rightarrow A20 \rightarrow S21)$   
 $p_2 = P(A|q \rightarrow A20 \rightarrow BUST)$   
 $p_3 = P(A|q \rightarrow S21)$   
 $p_4 = P(A|q \rightarrow BUST)$

$$U^*(A|q) = R_0(A|q) + \gamma \left( p_1 (R_0(A20) + R_1(S21)) + p_2 (R_0(A20) + R_1(BUST)) + p_3 R_0(S21) + p_4 R_0(BUST) \right)$$

where:  $p_1 = P(A|q \rightarrow A20 \rightarrow S21)$

$p_2 = P(A|q \rightarrow A20 \rightarrow BUST)$

$p_3 = P(A|q \rightarrow S21)$

$p_4 = P(A|q \rightarrow BUST)$

$$U^*(A|q) = R_0(A|q) + \gamma \left( p_1 (R_0(A20) + R_1(S21)) + p_2 (R_0(A20) + R_1(BUST)) + p_3 R_0(S21) + p_4 R_0(BUST) \right)$$


---


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

where:

- $p_1 = P(A|q \rightarrow A20 \rightarrow S21)$
- $p_2 = P(A|q \rightarrow A20 \rightarrow BUST)$
- $p_3 = P(A|q \rightarrow S21)$
- $p_4 = P(A|q \rightarrow BUST)$

$$J^*(A|q) = R_0(A|q) + \gamma \left( p_1 (R_0(A20) + R_1(S21)) + p_2 (R_0(A20) + R_1(BUST)) + p_3 R_0(S21) + p_4 R_0(BUST) \right)$$

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

$$P(A|q \rightarrow A20 \rightarrow S21) (R_0(A20) + R_1(S21))$$

$$+ P(A|q \rightarrow A20 \rightarrow BUST) (R_0(A20) + R_1(BUST))$$

$$+ P(A|q \rightarrow S21) R_0(S21)$$

$$+ P(A|q \rightarrow BUST) R_0(BUST)$$

=

$$J^*(A|q) = R_0(A|q) + \gamma \left( p_1 (R_0(A20) + R_1(S21)) + p_2 (R_0(A20) + R_1(BUST)) + p_3 R_0(S21) + p_4 R_0(BUST) \right)$$

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

$$\begin{aligned}
 & P(A|q \rightarrow A20 \rightarrow S21) (R_0(A20) + R_1(S21)) \\
 & + P(A|q \rightarrow A20 \rightarrow BUST) (R_0(A20) + R_1(BUST)) \\
 & + P(A|q \rightarrow S21) R_0(S21) \\
 & + P(A|q \rightarrow BUST) R_0(BUST)
 \end{aligned}
 \quad =
 \quad
 \begin{aligned}
 & P(A|q \rightarrow A20) P(A20 \rightarrow S21) (R_0(A20) + R_1(S21)) \\
 & + P(A|q \rightarrow A20) \underline{P(A20 \rightarrow BUST)} (R_0(A20) + R_1(BUST)) \\
 & + P(A|q \rightarrow S21) \overbrace{R_0(S21)}^{+ P(A|q \rightarrow BUST) R_0(BUST)} \\
 & + P(A|q \rightarrow BUST) R_0(BUST)
 \end{aligned}$$
$$P(BUST | A20, HIT)$$

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

$$P_1(R_o(\text{A20}) + R_i(\text{S2I})) + P_2(R_o(\text{A20}) + R_i(\text{BUST})) + P_3 R_o(\text{S2I}) + P_4 R_o(\text{BUST})$$

$$\begin{aligned}
 & P(A19 \rightarrow A20 \rightarrow S21) (R_o(A20) + R_i(S21)) \\
 = & + P(A19 \rightarrow A20 \rightarrow \text{BUST}) (R_o(A20) + R_i(\text{BUST})) \\
 & + P(A19 \rightarrow S21) R_o(S21) \\
 & + P(A19 \rightarrow \text{BUST}) R_o(\text{BUST})
 \end{aligned}
 \quad
 \begin{aligned}
 & P(A19 \rightarrow A20) P(A20 \rightarrow S21) (R_o(A20) + R_i(S21)) \\
 = & + P(A19 \rightarrow A20) \underline{P(A20 \rightarrow \text{BUST})} (R_o(A20) + R_i(\text{BUST})) \\
 & + P(A19 \rightarrow S21) R_o(S21) \\
 & + P(A19 \rightarrow \text{BUST}) R_o(\text{BUST})
 \end{aligned}$$


 $P(\text{BUST} | A20, \pi(A20))$

$$J^*(A|q) = R_0(A|q) + \gamma \left( p_1 (R_0(A20) + R_1(S21)) + p_2 (R_0(A20) + R_1(BUST)) + p_3 R_0(S21) + p_4 R_0(BUST) \right)$$

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

$$P(A|q \rightarrow A20) P(A20 \rightarrow S21) (R_0(A20) + R_1(S21))$$

$$+ P(A|q \rightarrow A20) P(A20 \rightarrow BUST) (R_0(A20) + R_1(BUST))$$

$$+ P(A|q \rightarrow S21) R_0(S21)$$

$$+ P(A|q \rightarrow BUST) R_0(BUST)$$

=

$$J^*(A|q) = R_0(A|q) + \gamma \left( p_1 (R_0(A20) + R_1(S21)) + p_2 (R_0(A20) + R_1(BUST)) + p_3 R_0(S21) + p_4 R_0(BUST) \right)$$

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

$$\begin{aligned}
 & P(A|q \rightarrow A20) P(A20 \rightarrow S21) (R_0(A20) + R_1(S21)) \\
 & + P(A|q \rightarrow A20) P(A20 \rightarrow BUST) (R_0(A20) + R_1(BUST)) \\
 & + P(A|q \rightarrow S21) R_0(S21) \\
 & + P(A|q \rightarrow BUST) R_0(BUST)
 \end{aligned}
 \quad =
 \quad
 \begin{aligned}
 & P(A|q \rightarrow A20) \left( P(A20 \rightarrow S21) (R_0(A20) + R_1(S21)) \right. \\
 & \quad \left. + P(A20 \rightarrow BUST) (R_0(A20) + R_1(BUST)) \right) \\
 & + P(A|q \rightarrow S21) R_0(S21) \\
 & + P(A|q \rightarrow BUST) R_0(BUST)
 \end{aligned}$$

$$J^*(A|q) = R_0(A|q) + \gamma \left( p_1 (R_0(A20) + R_1(S21)) + p_2 (R_0(A20) + R_1(BUST)) + p_3 R_0(S21) + p_4 R_0(BUST) \right)$$

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

$$= P(A|q \rightarrow A20) \left( P(A20 \rightarrow S21) (R_0(A20) + R_1(S21)) + P(A20 \rightarrow BUST) (R_0(A20) + R_1(BUST)) \right) \\ + P(A|q \rightarrow S21) R_0(S21) \\ + P(A|q \rightarrow BUST) R_0(BUST)$$

$$\begin{aligned} \mathcal{U}(\text{A1q}) &= R_0(\text{A1q}) \\ &+ \gamma \left( p_1 (R_0(\text{A20}) + R_1(\text{S2I})) + p_2 (R_0(\text{A20}) + R_1(\text{BUST})) + p_3 R_0(\text{S2I}) + p_4 R_0(\text{BUST}) \right) \end{aligned}$$

$$p_1 (R_0(\text{A20}) + R_1(\text{S2I})) + p_2 (R_0(\text{A20}) + R_1(\text{BUST})) + p_3 R_0(\text{S2I}) + p_4 R_0(\text{BUST})$$

$$\begin{aligned}
&= P(\text{A1q} \rightarrow \text{A20}) \left( P(\text{A20} \rightarrow \text{S2I}) (R_0(\text{A20}) + R_1(\text{S2I})) \right. \\
&\quad \left. + P(\text{A20} \rightarrow \text{BUST}) (R_0(\text{A20}) + R_1(\text{BUST})) \right) = \\
&= P(\text{A1q} \rightarrow \text{A20}) \left( P(\text{A20} \rightarrow \text{S2I}) \cup (\text{A20} \rightarrow \text{S2I}) \right. \\
&\quad \left. + P(\text{A20} \rightarrow \text{BUST}) \cup (\text{A20} \rightarrow \text{BUST}) \right) \\
&+ P(\text{A1q} \rightarrow \text{S2I}) R_0(\text{S2I}) \\
&+ P(\text{A1q} \rightarrow \text{BUST}) R_0(\text{BUST})
\end{aligned}$$

the lifetime  
 reward of  
 $\text{A20} \rightarrow \text{BUST}$

$$\boxed{U^*(A|q) = R_0(A|q) + \gamma \left( p_1 (R_0(A20) + R_1(S21)) + p_2 (R_0(A20) + R_1(BUST)) + p_3 R_0(S21) + p_4 R_0(BUST) \right)}$$

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

$$= P(A|q \rightarrow A20) \left( P(A20 \rightarrow S21) U(A20 \rightarrow S21) + P(A20 \rightarrow BUST) U(A20 \rightarrow BUST) \right) \\ + P(A|q \rightarrow S21) R_0(S21) + P(A|q \rightarrow BUST) R_0(BUST)$$

$$U^\pi(A|q) = R_0(A|q) + \gamma \left( p_1 (R_0(A20) + R_1(S21)) + p_2 (R_0(A20) + R_1(BUST)) + p_3 R_0(S21) + p_4 R_0(BUST) \right)$$

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

$$= P_{(A|q \rightarrow A20)} \left[ P_{(A20 \rightarrow S21)} U_{(A20 \rightarrow S21)} + P_{(A20 \rightarrow BUST)} U_{(A20 \rightarrow BUST)} \right] \leftarrow \begin{array}{l} \text{expected utility} \\ \text{of policy } \pi \text{ in state } A20 \end{array}$$

$$+ P_{(A|q \rightarrow S21)} R_0(S21)$$

$$+ P_{(A|q \rightarrow BUST)} R_0(BUST)$$

$$\begin{aligned} \cup^{\pi}(A|q) &= R_0(A|q) \\ &+ \gamma \left( p_1 (R_0(A20) + R_1(S21)) + p_2 (R_0(A20) + R_1(BUST)) + p_3 R_0(S21) + p_4 R_0(BUST) \right) \end{aligned}$$

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

$$\begin{aligned} &= P(A|q \rightarrow A20) \left[ P(A20 \rightarrow S21) \cup (A20 \rightarrow S21) \right. \\ &\quad \left. + P(A20 \rightarrow BUST) \cup (A20 \rightarrow BUST) \right] \leftarrow \cup^{\pi}(A20) \\ &+ P(A|q \rightarrow S21) R_0(S21) \\ &+ P(A|q \rightarrow BUST) R_0(BUST) \end{aligned}$$

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

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

$$= P(A|q \rightarrow A20) \left( P(A20 \rightarrow S21) U(A20 \rightarrow S21) + P(A20 \rightarrow BUST) U(A20 \rightarrow BUST) \right) = P(A|q \rightarrow A20) U^{\pi}(A20) \\ + P(A|q \rightarrow S21) R_0(S21) + P(A|q \rightarrow BUST) R_0(BUST)$$

$$U^*(A|q) = R_0(A|q) + \gamma \left( p_1 (R_0(A20) + R_1(S21)) + p_2 (R_0(A20) + R_1(BUST)) + p_3 R_0(S21) + p_4 R_0(BUST) \right)$$

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

$$\begin{aligned} & P(A|q \rightarrow A20) U^*(A20) \\ &= + P(A|q \rightarrow S21) R_0(S21) \\ &+ P(A|q \rightarrow BUST) R_0(BUST) \end{aligned}$$

$$U^\pi(A|q) = R_0(A|q) + \gamma \left( p_1 (R_0(A20) + R_1(S21)) + p_2 (R_0(A20) + R_1(BUST)) + p_3 R_0(S21) + p_4 R_0(BUST) \right)$$

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

$$P(A|q \rightarrow A20) U^\pi(A20)$$

$$= + P(A|q \rightarrow S21) \boxed{R_0(S21)} \leftarrow \text{expected utility of policy } \pi \text{ in state } S21$$

$$+ P(A|q \rightarrow BUST) \boxed{R_0(BUST)} \leftarrow \text{expected utility of policy } \pi \text{ in state } BUST$$

$$U^*(A|q) = R_0(A|q) + \gamma \left( p_1 (R_0(A20) + R_1(S21)) + p_2 (R_0(A20) + R_1(BUST)) + p_3 R_0(S21) + p_4 R_0(BUST) \right)$$

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

$$\begin{aligned} & P(A|q \rightarrow A20) U^*(A20) \\ &= + P(A|q \rightarrow S21) U^*(S21) \\ &+ P(A|q \rightarrow BUST) U^*(BUST) \end{aligned}$$

$$U^*(A|q) = R_0(A|q) + \gamma \left( p_1 (R_0(A20) + R_1(S21)) + p_2 (R_0(A20) + R_1(BUST)) + p_3 R_0(S21) + p_4 R_0(BUST) \right)$$

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

$$= P(A|q \rightarrow A20) U^*(A20) + P(A|q \rightarrow S21) U^*(S21) + P(A|q \rightarrow BUST) U^*(BUST)$$

substitute  
back in

$$\begin{aligned} U^*(A|q) &= R_0(A|q) \\ &+ \gamma \left( P_{(A|q \rightarrow A20)} U^*(A20) + P_{(A|q \rightarrow S21)} U^*(S21) + P_{(A|q \rightarrow BUST)} U^*(BUST) \right) \end{aligned}$$

$$U^\pi(A|q) = R_0(A|q) + \gamma \left( P_{(A|q \rightarrow A20)} U^\pi(A20) + P_{(A|q \rightarrow S21)} U^\pi(S21) + P_{(A|q \rightarrow BUST)} U^\pi(BUST) \right)$$

$$= R_0(A|q) + \gamma \sum_{q' \in Q} P_{(A|q \xrightarrow{\pi(A|q)} q')} U^\pi(q')$$

$$= R_0(A|q) + \gamma \sum_{q' \in Q} P_{(q' | A|q, \pi(A|q))} U^\pi(q')$$

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

the expected utility of  
policy  $\pi$  in state  $A|q$

discounting  
factor

the expected utility of  
policy  $\pi$  in state  $q'$

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

the immediate reward  
of being in state  $A|q$

the probability of getting  
from state  $A|q$  to state  $q'$   
using the action  
advised by the policy

the expected utility of  
policy  $\pi$  in state  $q$

discounting  
factor

the expected utility of  
policy  $\pi$  in state  $q'$

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

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

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

this is the expected utility of  
policy  $\pi$  if we start in state  $q$