USC Viterbi
School of Engineering

Decision Making Under Uncertainty

CSCI 699 Computational Human-Robot Interaction

Instructor: Stefanos Nikolaidis

Decision Making Problems

	Environment Deterministic	Environment Non-Deterministic
State Known		
State Uknown		

Decision Making Problems

	Environment Deterministic	Environment Non-Deterministic
State Known	A* search	
State Unknown		

Decision Making Problems

	Environment Deterministic	Environment Non-Deterministic
State Known	A* search	Markov Decision Process
State Unknown		

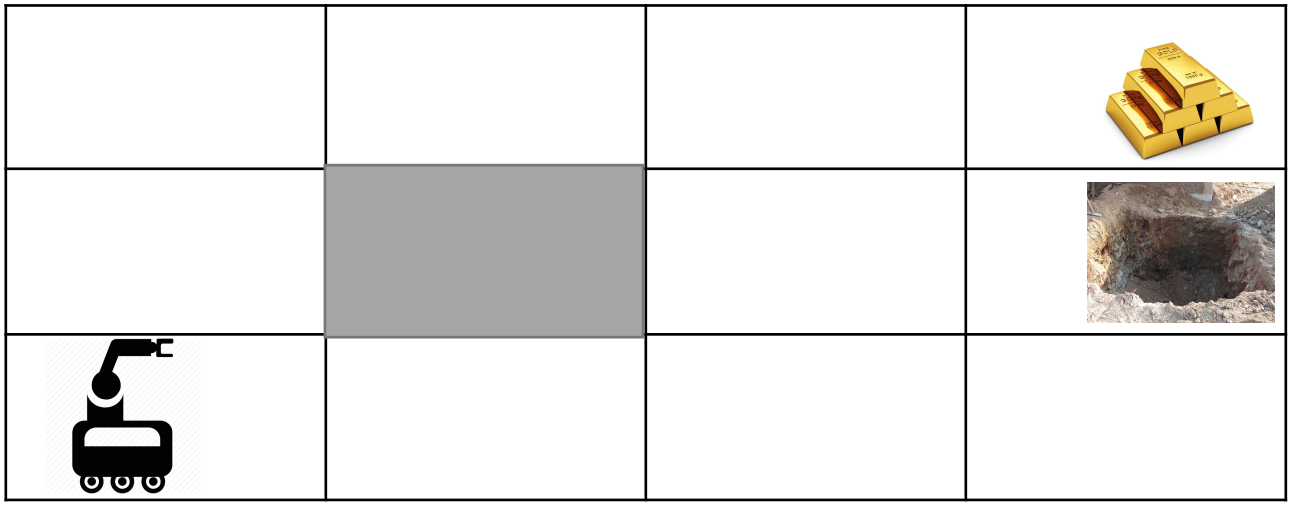
Decision Making Problems

	Environment Deterministic	Environment Non-Deterministic
State Known	A* search	Markov Decision Process
State Unknown	Partially Observable Markov Decision Process (special case)	Partially Observable Markov Decision Process

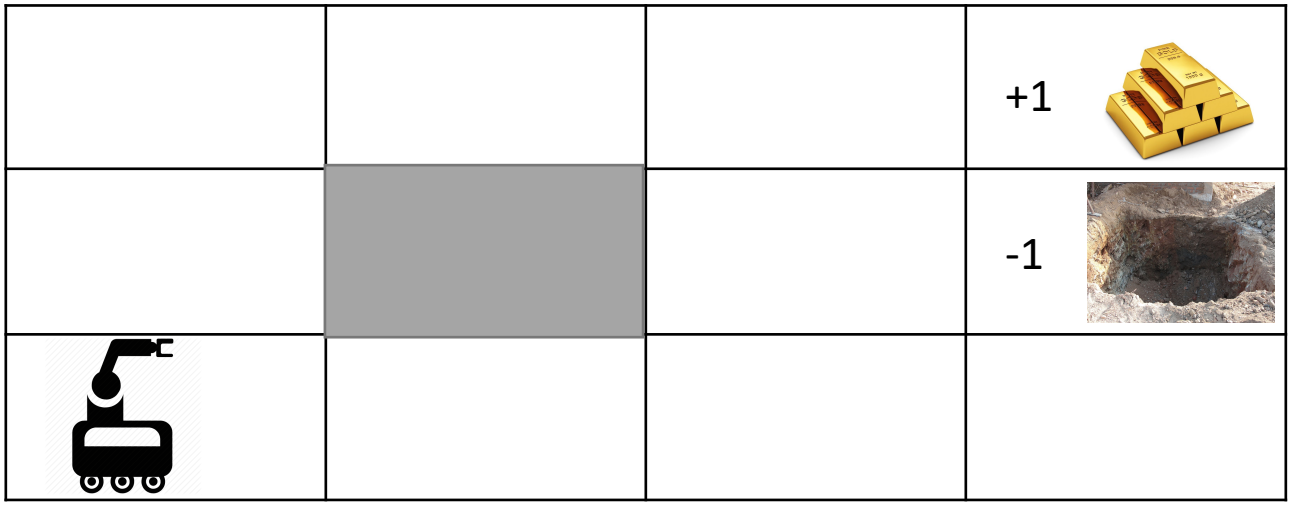
Decision Making Problems

	Environment Deterministic	Environment Non-Deterministic
State Known	A* search	Markov Decision Process
State Unknown	Partially Observable Markov Decision Process (special case)	Partially Observable Markov Decision Process

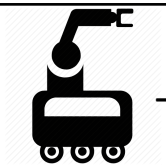
Sequential Decision Making



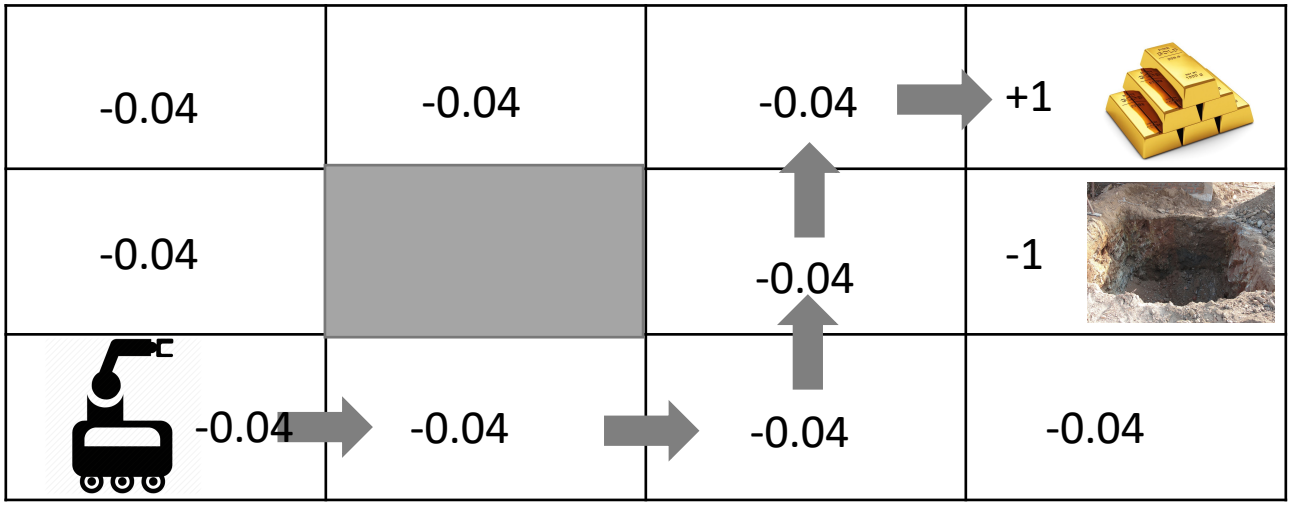
Sequential Decision Making



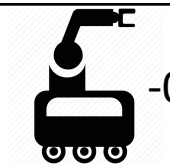
Sequential Decision Making

-0.04	-0.04	-0.04	+1 
-0.04		-0.04	-1 
 -0.04	-0.04	-0.04	-0.04

Deterministic Dynamics

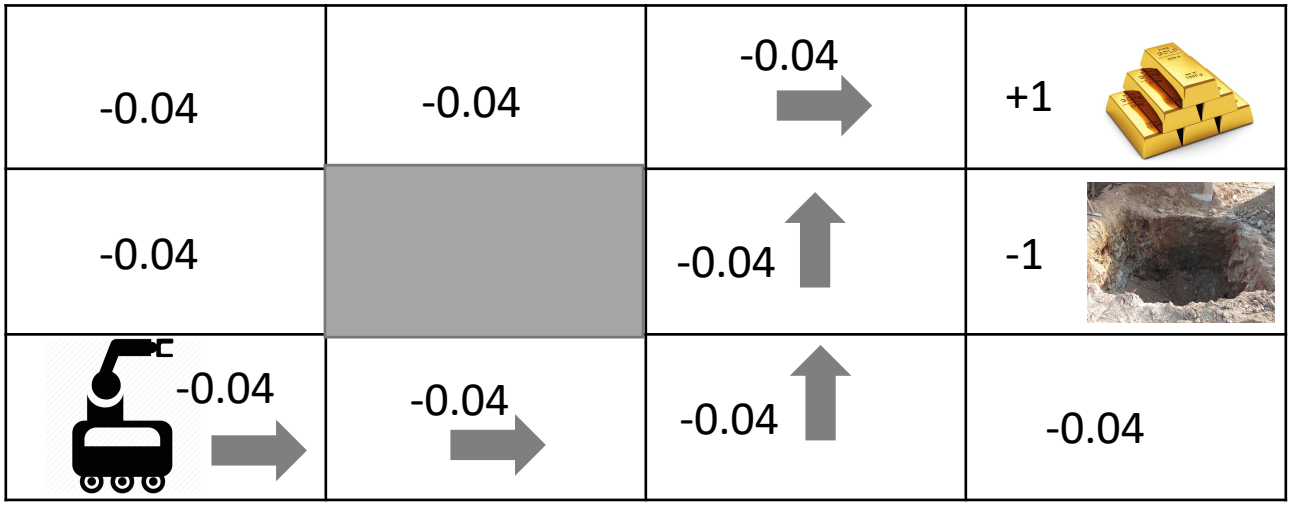


Deterministic Dynamics

-0.04	-0.04	-0.04 →	+1 
-0.04		-0.04 ↑	-1 
 -0.04 →	-0.04 →	-0.04 ↑	-0.04

How good is the path?

Deterministic Dynamics



How good is the path?

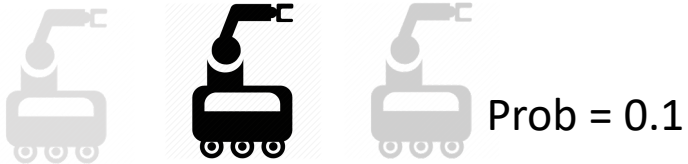
$$V(s_0) = -0.04 * 5 + 1 = 0.8$$

Stochastic Dynamics

-0.04	-0.04	-0.04	+1 
-0.04		-0.04	-1 
 -0.04	-0.04	-0.04	-0.04

Stochastic Dynamics

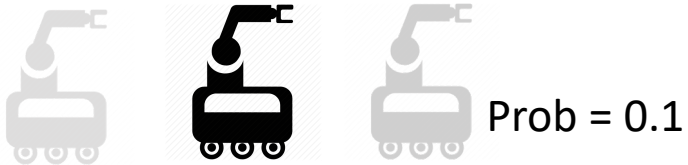
-0.04	-0.04	-0.04 →	+1 
-0.04		-0.04 ↑	-1 
 -0.04 →	-0.04 →	-0.04 ↑	-0.04



Stochastic Dynamics

-0.04	-0.04	-0.04 →	+1 
-0.04		-0.04 ↑	-1 
 -0.04 →	-0.04 →	-0.04 ↑	-0.04

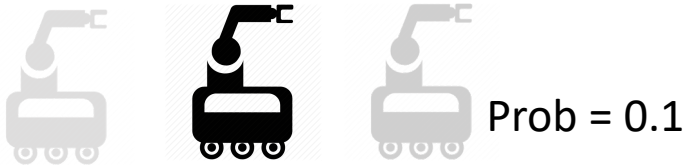
$$0.8^5 = 0.33$$



Stochastic Dynamics

-0.04	-0.04	-0.04 →	+1 
-0.04		-0.04 ↑	-1 
 -0.04 →	-0.04 →	-0.04 ↑	-0.04

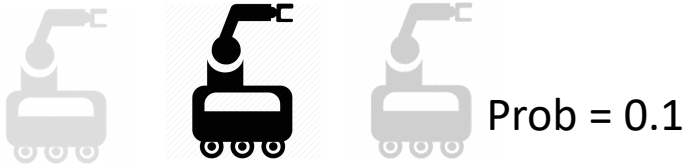
$$0.8^5 = 0.33$$



Stochastic Dynamics

-0.04	-0.04	-0.04 →	+1 
-0.04		-0.04 ↑	-1 
 -0.04 →	-0.04 →	-0.04 ↑	-0.04

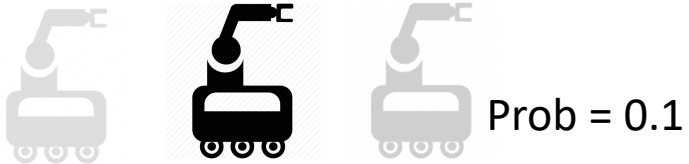
$$0.8^5 + 0.1^4 * 0.8 = 0.33$$



Stochastic Dynamics

-0.04	-0.04	-0.04 →	+1 
-0.04		-0.04 ↑	-1 
 -0.04 →	-0.04 →	-0.04 ↑	-0.04

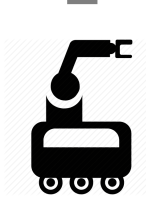
$$0.8^5 + 0.1^4 * 0.8 = 0.33$$



Stochastic Dynamics

-0.04	-0.04	-0.04 →	+1 
-0.04		-0.04 ↑	-1 
 -0.04 →	-0.04 →	-0.04 ↑	-0.04

How good is the path?

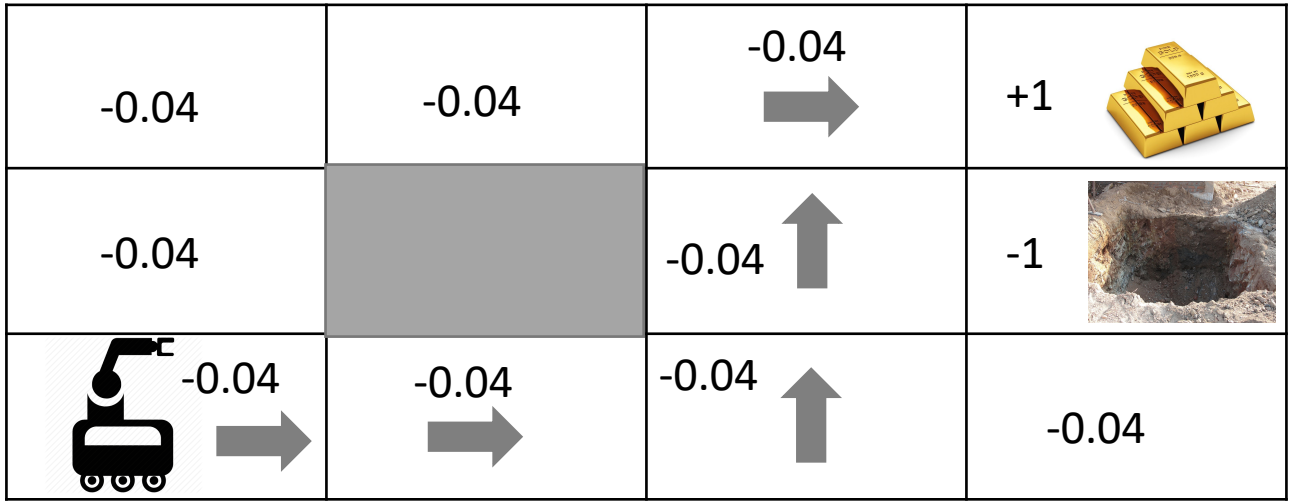



Prob = 0.8

Prob = 0.1

Prob = 0.1

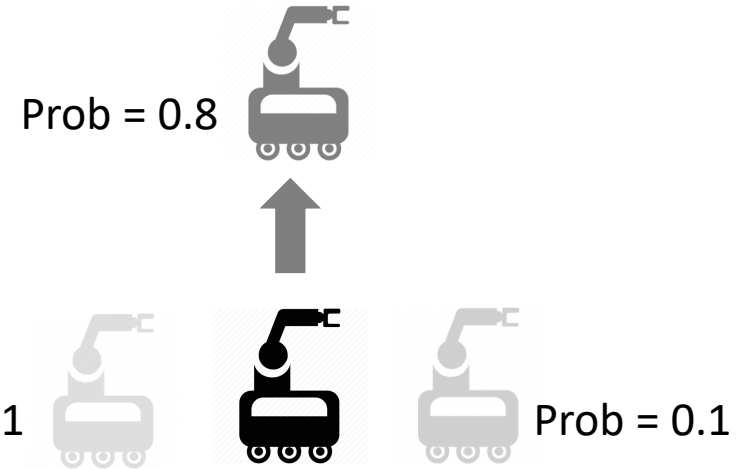
Stochastic Dynamics



$$\pi : S \rightarrow A$$

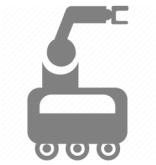
How good is the path?

$$V^{\pi_t}(x_0) = \mathbb{E}[\sum_{t=0}^{T} R(x_t)|\pi_t]$$



Stochastic Dynamics

-0.04	-0.04	-0.04 →	+1 
-0.04		-0.04 ↑	-1 
 -0.04 →	-0.04 →	-0.04 ↑	-0.04

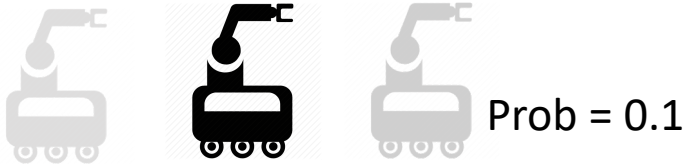


How good is the path?

$$V^{\pi_t}(x_0) = \mathbb{E}[\sum_{t=0}^{T} R(x_t)|\pi_t]$$

Stochastic Dynamics

-0.04	-0.04	-0.04 →	+1 
-0.04		-0.04 ↑	-1 
 -0.04 →	-0.04 →	-0.04 ↑	-0.04



Time Horizon

$$V^{\pi_t}(x_0) = \mathbb{E}[\sum_{t=0}^{T} R(x_t)|\pi_t]$$

Stochastic Dynamics

-0.04	-0.04	-0.04 →	+1 
-0.04		-0.04 ↑	-1 
 -0.04 →	-0.04 →	-0.04 ↑	-0.04

What if $T = 5$?

$$V^{\pi_t}(x_0) = \mathbb{E}[\sum_{t=0}^{T} R(x_t)|\pi_t]$$

Stochastic Dynamics

-0.04	-0.04	-0.04 →	+1 
-0.04		-0.04 ↑	-1 
 -0.04 →	-0.04 →	-0.04 ↑	-0.04

What if $T = 3$?

$$V^{\pi_t}(x_0) = \mathbb{E}[\sum_{t=0}^{T} R(x_t)|\pi_t]$$

Stochastic Dynamics

-0.04	-0.04	-0.04	+1 
-0.04		-0.04	-1 
 -0.04	-0.04	-0.04	-0.04

What if $T = 10000$?

$$V^{\pi_t}(x_0) = \mathbb{E}[\sum_{t=0}^{T} R(x_t)|\pi_t]$$

Stochastic Dynamics

-0.04 →	-0.04 →	-0.04 →	+1 
-0.04 ↑		-0.04	-1 
-0.04 ↑	-0.04	-0.04	-0.04

Prob = 0.8

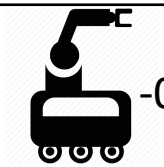
What if $T = 10000$?

$$V^{\pi_t}(x_0) = \mathbb{E}[\sum_{t=0}^{T} R(x_t)|\pi_t]$$

Prob = 0.1

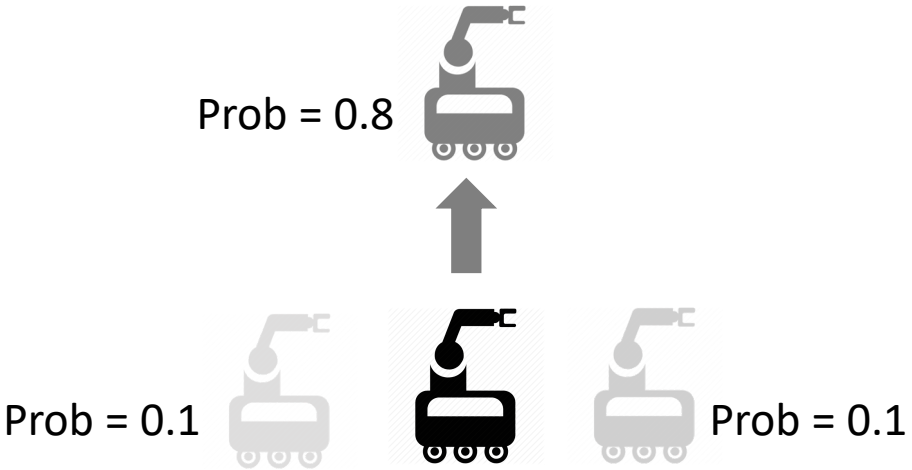
Prob = 0.1

Stochastic Dynamics

-0.04	-0.04	-0.04	+1 
-0.04		-0.04	-1 
 -0.04	-0.04	-0.04	-0.04

What if we don't want to specify a time horizon?

$$V^{\pi_t}(x_0) = \mathbb{E}[\sum_{t=0}^{T} R(x_t)|\pi_t]$$



Stochastic Dynamics

-0.04	-0.04	-0.04	+1 
-0.04		-0.04	-1 
 -0.04	-0.04	-0.04	-0.04

What if we don't want to specify a time horizon?

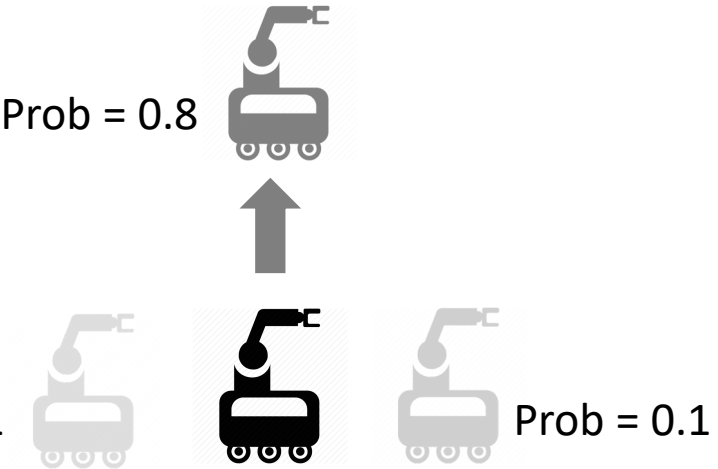
$$V^{\pi}(x_0) = \mathbb{E}[\sum_{t=0}^{\infty} R(x_t)|\pi]$$

Stochastic Dynamics

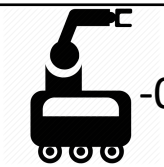
-0.04	-0.04	-0.04	+1 
-0.04		-0.04	-1 
 -0.04	-0.04	-0.04	-0.04

Sum can be unbounded!

$$V^{\pi}(x_0) = \mathbb{E}[\sum_{t=0}^{\infty} R(x_t)|\pi]$$



Stochastic Dynamics

-0.04	-0.04	-0.04	+1 
-0.04		-0.04	-1 
 -0.04	-0.04	-0.04	-0.04

discount factor

$$V^{\pi}(x_0) = \mathbb{E}[\sum_{t=0}^{\infty} \gamma^t R(x_t)|\pi]$$

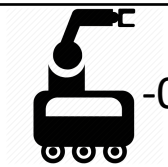
Discounted Accumulated Reward

$$\sum_{t=0}^{\infty} \gamma^t R(x_t) \leq$$

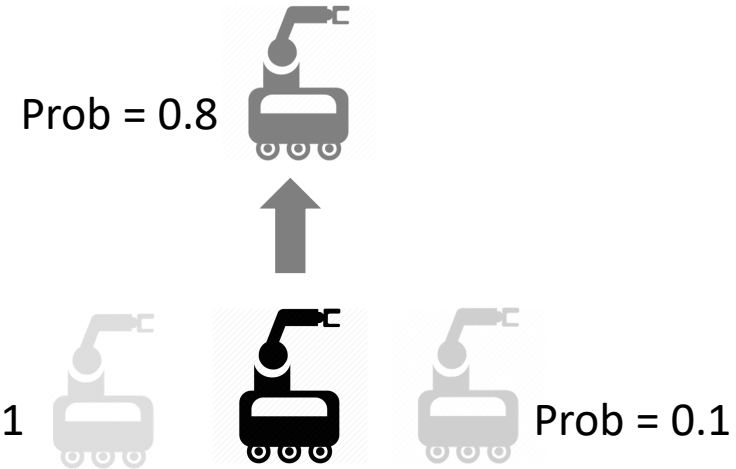
Discounted Accumulated Reward

$$\sum_{t=0}^{\infty} \gamma^t R(x_t) \leq \frac{1}{1-\gamma} R_{max}$$

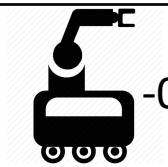
Optimal Policy

-0.04	-0.04	-0.04	+1 
-0.04		-0.04	-1 
 -0.04	-0.04	-0.04	-0.04

$$\pi^*(x_t) = \text{argmax}_\pi V^\pi(x_t)$$

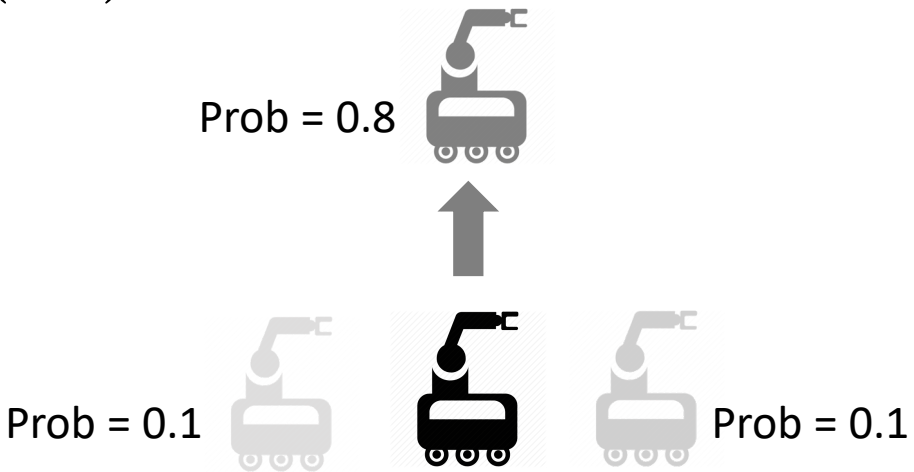


Optimal Policy

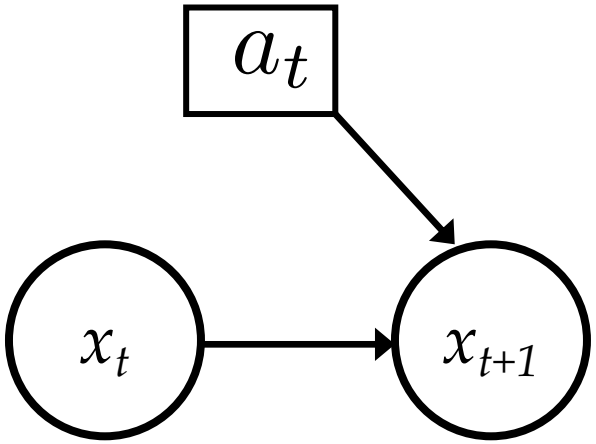
-0.04	-0.04	-0.04	+1 
-0.04		-0.04	-1 
 -0.04	-0.04	-0.04	-0.04

$$\pi^*(x_t) = \text{argmax}_\pi V^\pi(x_t)$$

Why do we need a policy?



Markov Decision Process



Markov Decision Process

- S: states
- A: actions
- T: S x A → Π(S)
- R: S → $\mathbb{R}$

Optimal Value Function

$V^*(x_t) = \max_a [R(x_t) + \gamma * \text{expected future reward after } x_t]$

Optimal Value Function

$$
\begin{aligned}
V^*(x_t) &= \max_a [R(x_t) + \gamma * \text{expected future reward after } x_t] \\
&= \max_a [R(x_t) + \gamma * \sum_{x_{t+1}} T(x_t, a, x_{t+1}) [\max_{a'} [R(x_{t+1}) + \gamma * \text{expected future reward after } x_{t+1}]]
\end{aligned}
$$

Optimal Value Function

$$V^*(x_t) = \max_a[R(x_t) + \gamma * \text{expected future reward after } x_t]$$
$$= \max_a[R(x_t) + \gamma * \sum_{x_{t+1}} T(x_t, a, x_{t+1})[\max_{a'}[R(x_{t+1}) + \gamma * \text{expected future reward after } x_{t+1}]]$$

Optimal Value Function

$$
\begin{aligned}
V^*(x_t) &= \max_a [R(x_t) + \gamma * \text{expected future reward after } x_t] \\
&= \max_a [R(x_t) + \gamma * \sum_{x_{t+1}} T(x_t, a, x_{t+1}) [\max_{a'} [R(x_{t+1}) + \gamma * \text{expected future reward after } x_{t+1}]] \\
&= \max_a [R(x_t) + \gamma * \sum_{x_{t+1}} T(x_t, a, x_{t+1}) V^*(x_{t+1})
\end{aligned}
$$

Value Iteration

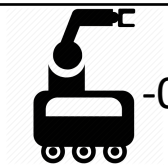
Value Iteration $V_0(s) = 0 \ \forall s \in S$
Iterate:

$$V_{t+1}(x) = \max_a [R(x) + \gamma * \sum_{x'} T(x, a, x') V_t(x')$$

Until: $\max_x |V_{t+1}(x) - V_t(x)| < \epsilon$

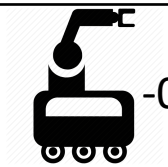
Value Iteration

$$\gamma = 1.0$$

-0.04	-0.04	-0.04	+1 
-0.04		-0.04	-1 
 -0.04	-0.04	-0.04	-0.04

$V_1(3,4) = +1 + 0$

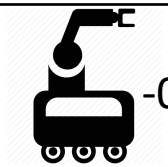
Value Iteration

-0.04	-0.04	-0.04	+1 
-0.04		-0.04	-1 
 -0.04	-0.04	-0.04	-0.04

$V_1(3,4) = +1 + 0$

$V_1(2,4) = -1 + 0$

Value Iteration

-0.04	-0.04	-0.04	+1 
-0.04		-0.04	-1 
 -0.04	-0.04	-0.04	-0.04

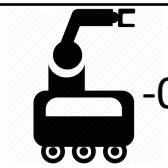
$$V_1(3,4) = +1 + 0$$

$$V_1(2,4) = -1 + 0$$

$$V_1(3,3) = -0.04 + \gamma \max_{top,right,left,down}[0.8 * V_0(3,3) + 0.1 * V_0(3,4) + 0.1 * V_0(3,2), ...,]$$

$$= -0.04 + 0 = -0.04$$

Value Iteration

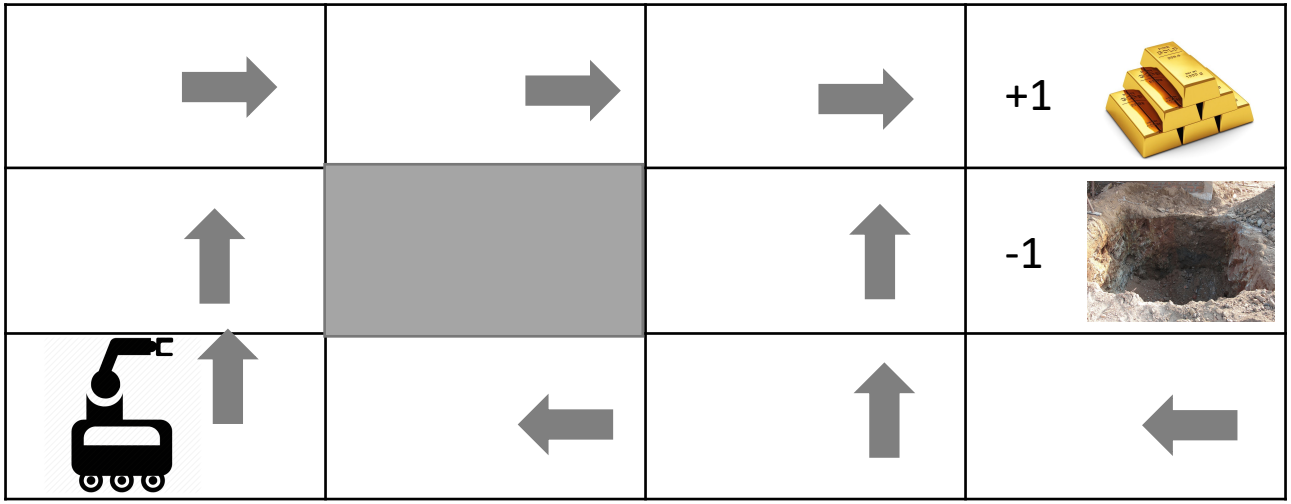
-0.04	-0.04	-0.04	+1 
-0.04		-0.04	-1 
 -0.04	-0.04	-0.04	-0.04

$$V_2(3,3) = -0.04 + 1.0 * [0.8 * (1) + 0.1 * (-0.04) + 0.1 * (-0.04)]$$
$$= -0.04 + 0.72 = 0.68$$

Value Iteration

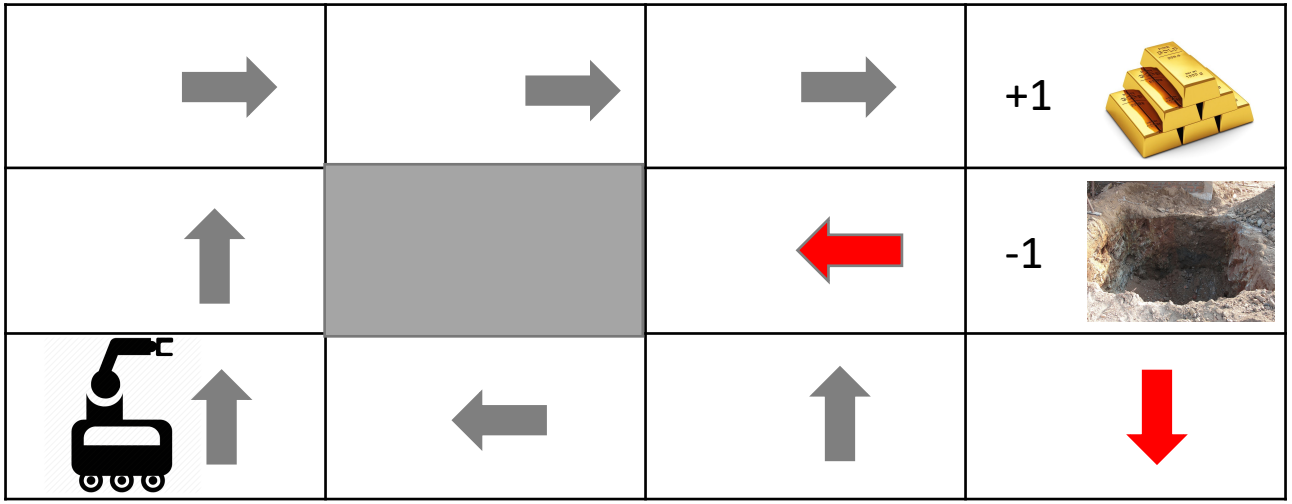
0.81 →	0.87 →	0.92 →	+1 
0.76 ↑		0.66 ↑	-1 
 ↑ 0.71	0.66 ←	0.61 ↑	0.39 ←

Value Iteration



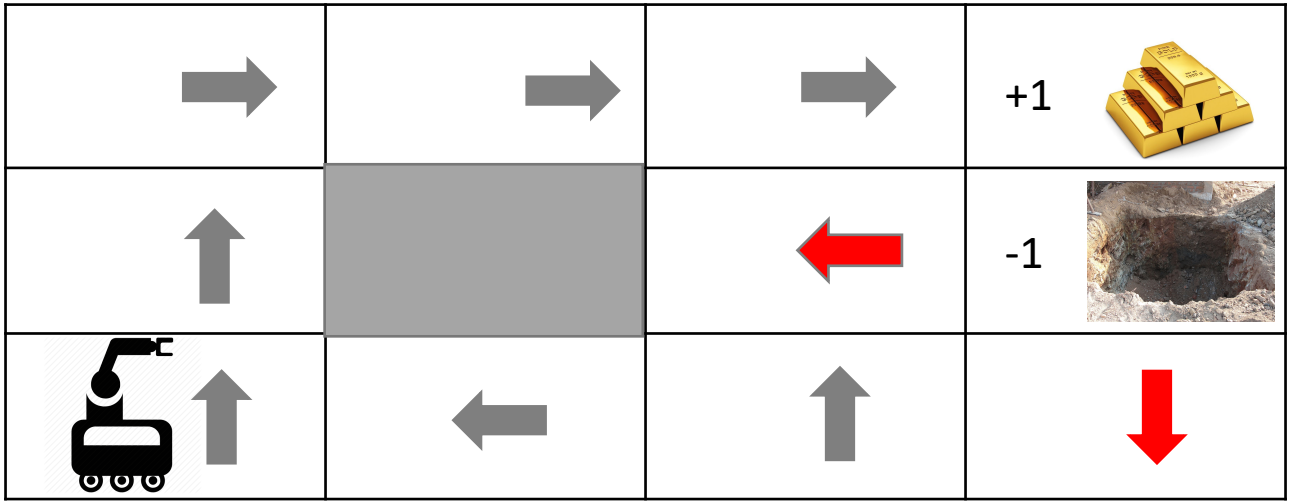
How does the policy change if $R(x_t) = -0.0001$?

Value Iteration



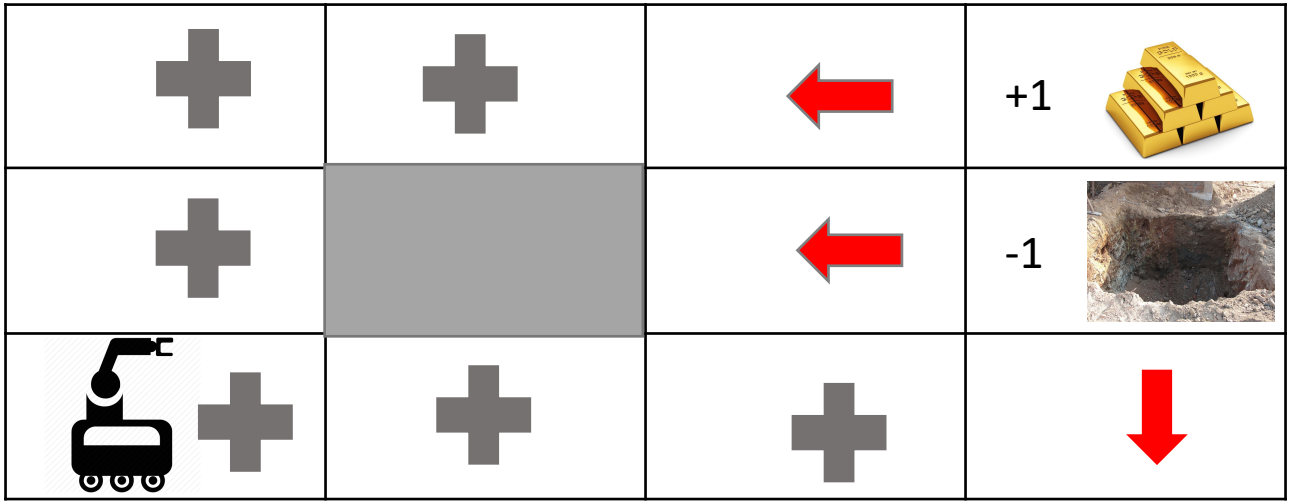
How does the policy change if $R(x_t) = -0.0001$?

Value Iteration



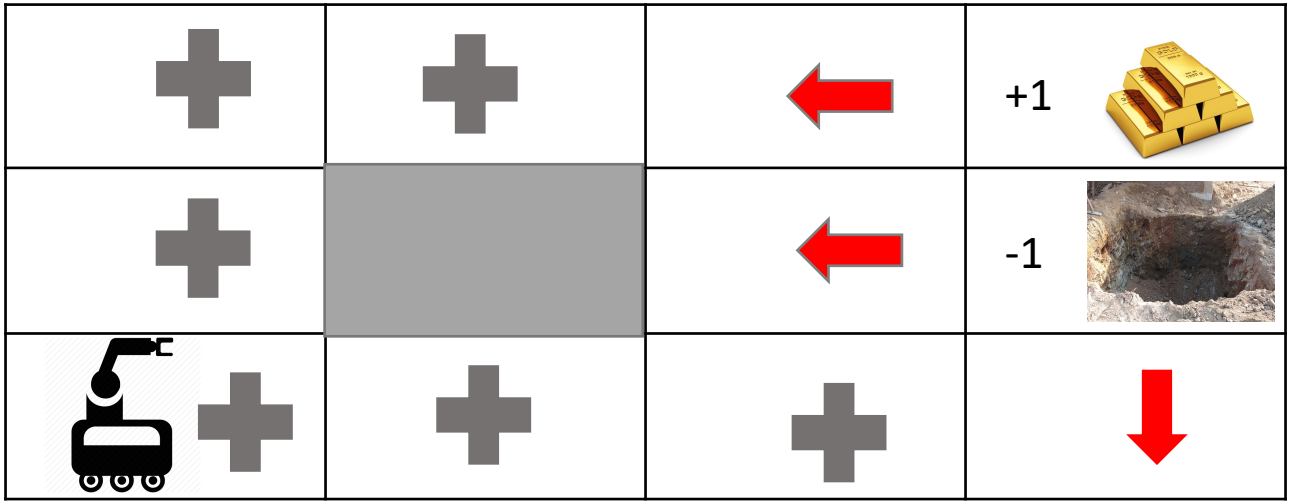
How does the policy change if $R(x_t) = +0.0001$?

Value Iteration



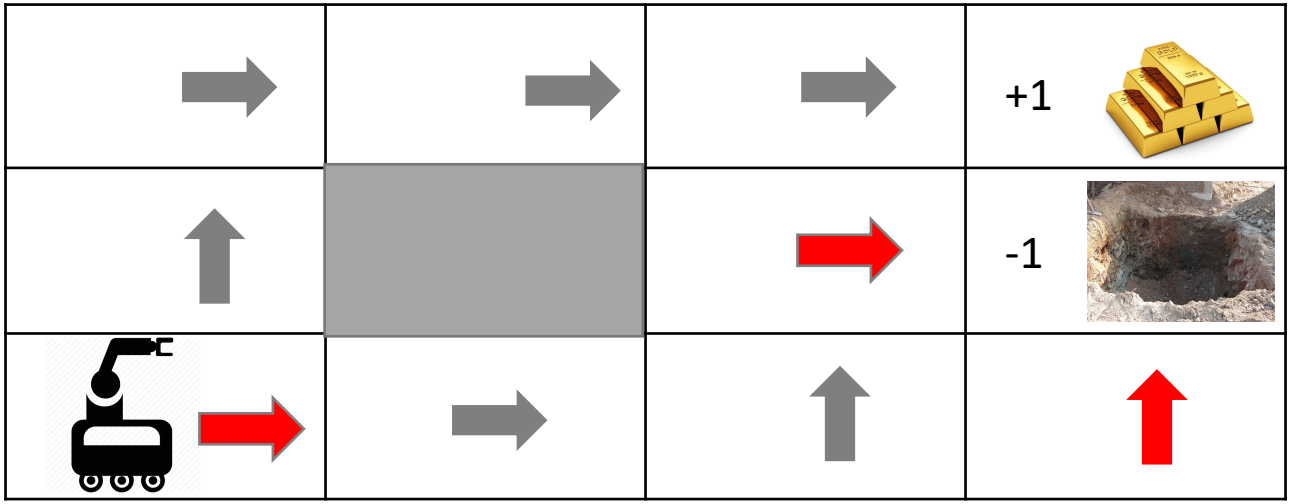
How does the policy change if $R(x_t) = +0.0001$?

Value Iteration



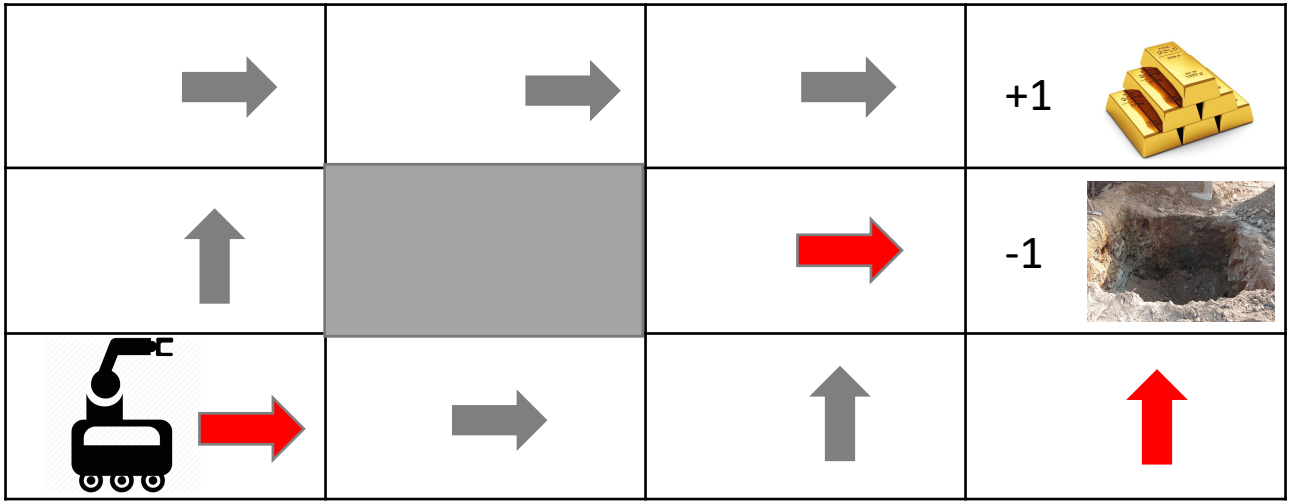
How does the policy change if $R(x_t) = -2$?

Value Iteration



How does the policy change if $R(x_t) = -2$?

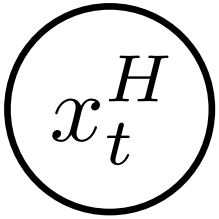
Value Iteration



How does the policy change if $R(x_t) = -2$?

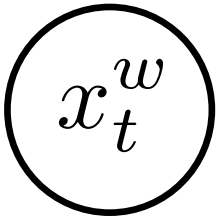

Ocean
Ocean
MoonStar
SARDINES

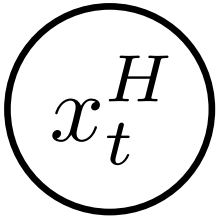
States



$$x_t^H = \{T, \neg T\}$$

States





$$x_t^H = \{T, \neg T\}$$

Dynamics

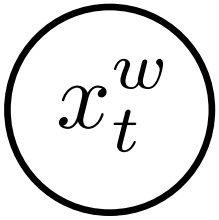
$x_t^w : B \times G$

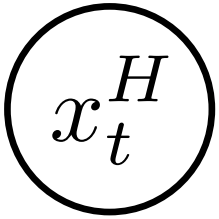
$B = \{T, R, H\}$

$T = \text{on table}$

$R = \text{picked by robot}$

$H = \text{picked by human}$





$$x_t^H = \{T, \neg T\}$$

Dynamics

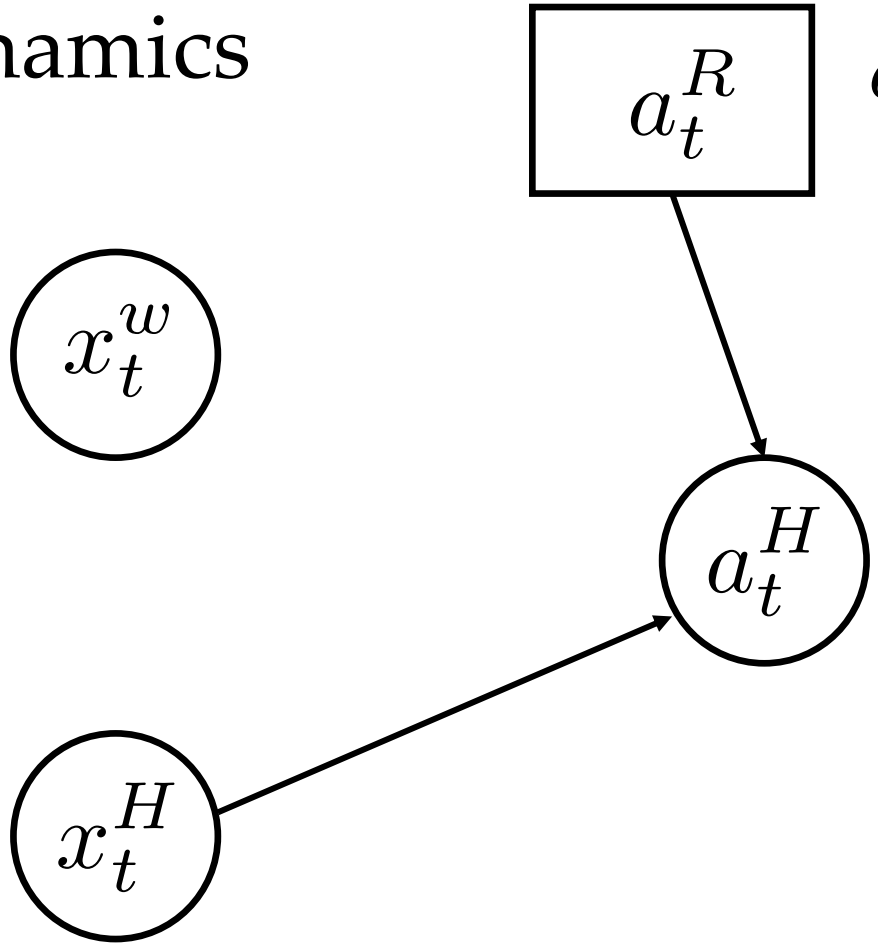
a_t^R $a_t^R = \{B, G\}$

x_t^w

x_t^H

$x_t^H = \{T, \neg T\}$

Dynamics

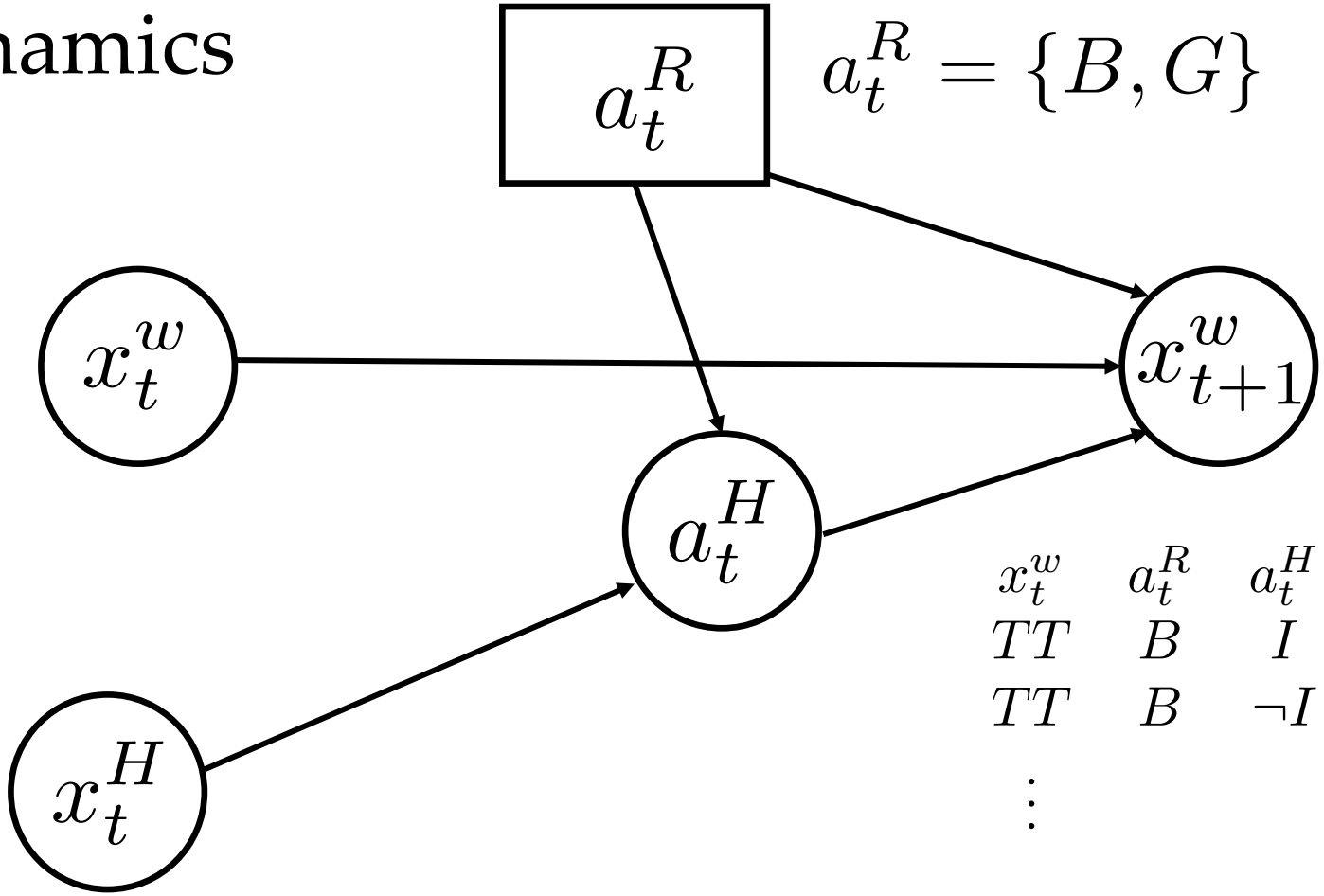


$a_t^R = \{B, G\}$

x_t^H	a_t^R	$P(a_t^H = I \mid x_t^H, a_t^R)$
T	B	0.1
T	G	0.2
$\neg T$	B	0.3
$\neg T$	G	0.8

$x_t^H = \{T, \neg T\}$

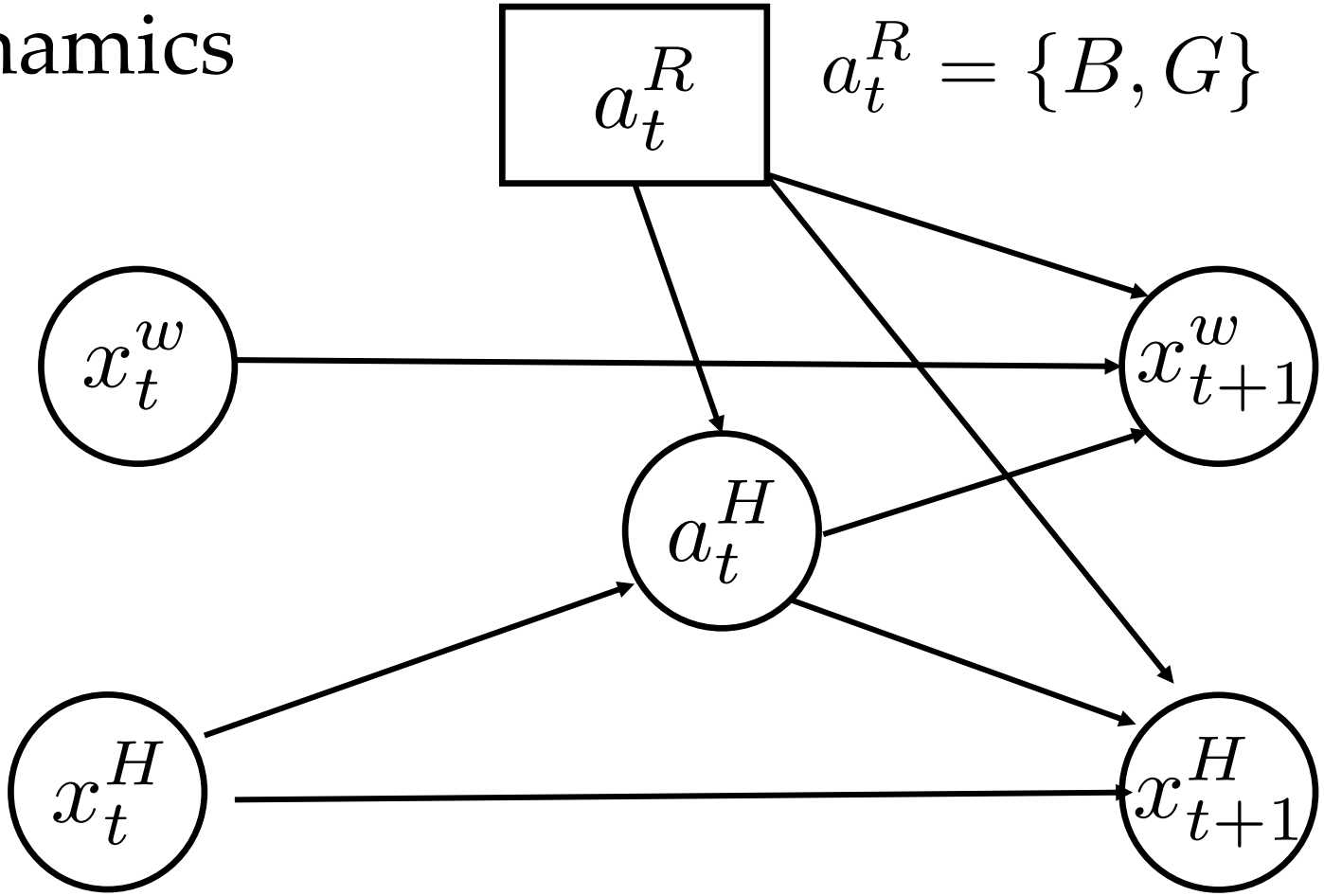
Dynamics



x_t^w	a_t^R	a_t^H	x_{t+1}^w
TT	B	I	HT
TT	B	$\neg I$	RT
$\vdots$			

$$x_t^H = \{T, \neg T\}$$

Dynamics



$$x_t^H = \{T, \neg T\}$$

x_t^H	a_t^R	a_t^H	$P(x_t^H = T \mid x_t^H, a_t^H, a_t^R)$
T	B	I	1.0
T	B	$\neg I$	1.0
$\vdots$			
$-T$	B	$\neg I$	0.8
$-T$	G	$\neg I$	0.9

Reward

$$R(*) = 0$$
$$R(RR) = 10$$
$$R(TR) = 5$$

Formulation

How can we formulate this as an MDP?

MDP: <X, T, A, R>

Formulation

How can we formulate this as an MDP?

MDP: <**X**, T, A, R>

Formulation

How can we formulate this as an MDP?

MDP: <**X**, T, A, R>

$$X^W \times X^H$$

Transition Function

MDP: <X, T, A, R>

Transition Function

$$T_s = T(x_t, a_t^R, x_{t+1})$$

Transition Function

$$
\begin{aligned}
T_s &= T(x_t, a_t^R, x_{t+1}) \\
&= P(x_{t+1}|x_t, a_t^R)
\end{aligned}
$$

Transition Function

$$
\begin{aligned}
T_s &= T(x_t, a_t^R, x_{t+1}) \\
&= P(x_{t+1}|x_t, a_t^R) \\
&= P(x_{t+1}^w, x_{t+1}^H|x_t^w, x_t^H, a_t^R)
\end{aligned}
$$

Transition Function

$$
\begin{aligned}
T_s &= T(x_t, a_t^R, x_{t+1}) \\
&= P(x_{t+1}|x_t, a_t^R) \\
&= P(x_{t+1}^w, x_{t+1}^H|x_t^w, x_t^H, a_t^R) \\
&= \sum_{a_t^H} P(x_{t+1}^w, x_{t+1}^H, a_t^H|x_t^w, x_t^H, a_t^R)
\end{aligned}
$$

Transition Function

$$
\begin{aligned}
T_s &= T(x_t, a_t^R, x_{t+1}) \\
&= P(x_{t+1}|x_t, a_t^R) \\
&= P(x_{t+1}^w, x_{t+1}^H|x_t^w, x_t^H, a_t^R) \\
&= \sum_{a_t^H} P(x_{t+1}^w, x_{t+1}^H, a_t^H|x_t^w, x_t^H, a_t^R) \\
&= \sum_{a_t^H} P(_{t+1}^w, x_{t+1}^H|x_t^w, x_t^H, a_t^R, a_t^H) P(a_t^H|x_t^w, x_t^H, a_t^R)
\end{aligned}
$$

Transition Function

$$
\begin{aligned}
T_s &= T(x_t, a_t^R, x_{t+1}) \\
&= P(x_{t+1}|x_t, a_t^R) \\
&= P(x_{t+1}^w, x_{t+1}^H|x_t^w, x_t^H, a_t^R) \\
&= \sum_{a_t^H} P(x_{t+1}^w, x_{t+1}^H, a_t^H|x_t^w, x_t^H, a_t^R) \\
&= \sum_{a_t^H} P(_{t+1}^w, x_{t+1}^H|x_t^w, x_t^H, a_t^R, a_t^H) P(a_t^H|x_t^w, x_t^H, a_t^R) \\
&= \sum_{a_t^H} P(x_{t+1}^w, x_{t+1}^H|x_t^w, x_t^H, a_t^R, a_t^H) P(a_t^H|x_t^H, a_t^R)
\end{aligned}
$$

Transition Function

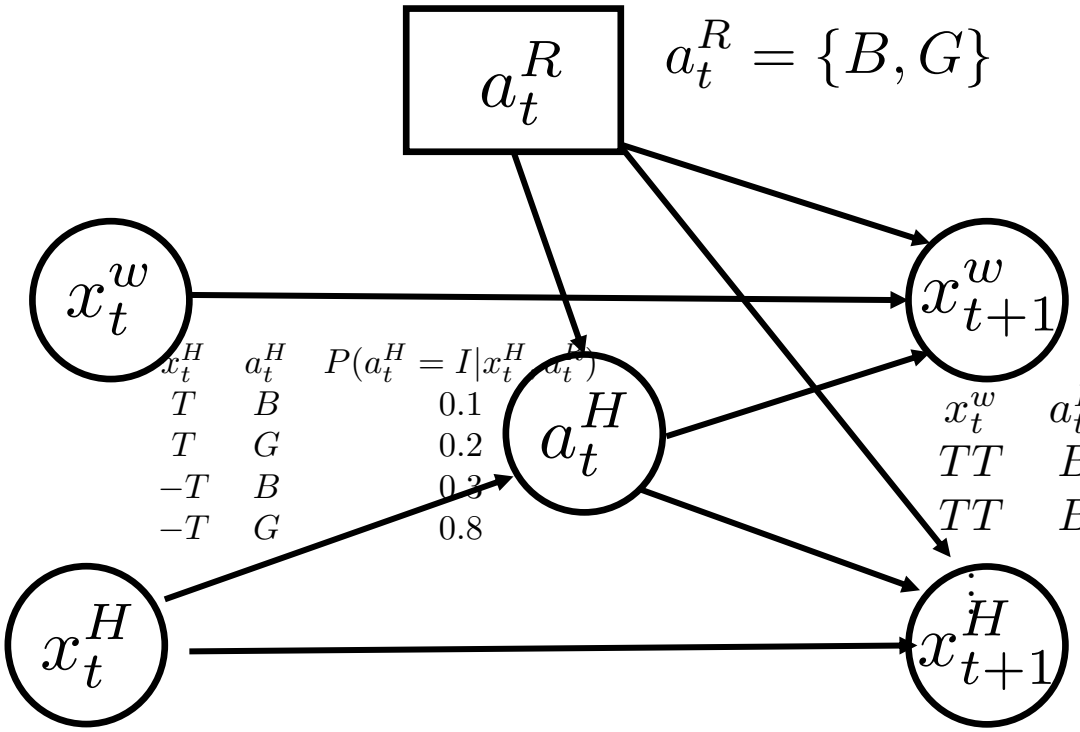
$$
\begin{aligned}
T_s &= T(x_t, a_t^R, x_{t+1}) \\
&= P(x_{t+1}|x_t, a_t^R) \\
&= P(x_{t+1}^w, x_{t+1}^H|x_t^w, x_t^H, a_t^R) \\
&= \sum_{a_t^H} P(x_{t+1}^w, x_{t+1}^H, a_t^H|x_t^w, x_t^H, a_t^R) \\
&= \sum_{a_t^H} P(^w_{t+1}, x_{t+1}^H|x_t^w, x_t^H, a_t^R, a_t^H) P(a_t^H|x_t^w, x_t^H, a_t^R) \\
&= \sum_{a_t^H} P(x_{t+1}^w, x_{t+1}^H|x_t^w, x_t^H, a_t^R, a_t^H) P(a_t^H|x_t^H, a_t^R) \\
&= \sum_{a_t^H} P(x_{t+1}^w|x_t^w, a_t^R, a_t^H) P(x_{t+1}^H|x_t^w, x_t^H, a_t^R, a_t^H) P(a_t^H|x_t^H, a_t^R)
\end{aligned}
$$

Transition Function

$$
\begin{aligned}
T_s &= T(x_t, a_t^R, x_{t+1}) \\
&= P(x_{t+1}|x_t, a_t^R) \\
&= P(x_{t+1}^w, x_{t+1}^H|x_t^w, x_t^H, a_t^R) \\
&= \sum_{a_t^H} P(x_{t+1}^w, x_{t+1}^H, a_t^H|x_t^w, x_t^H, a_t^R) \\
&= \sum_{a_t^H} P(^w_{t+1}, x_{t+1}^H|x_t^w, x_t^H, a_t^R, a_t^H) P(a_t^H|x_t^w, x_t^H, a_t^R) \\
&= \sum_{a_t^H} P(x_{t+1}^w, x_{t+1}^H|x_t^w, x_t^H, a_t^R, a_t^H) P(a_t^H|x_t^H, a_t^R) \\
&= \sum_{a_t^H} P(x_{t+1}^w|x_t^w, a_t^R, a_t^H) P(x_{t+1}^H|x_t^w, x_t^H, a_t^R, a_t^H) P(a_t^H|x_t^H, a_t^R) \\
&= \sum_{a_t^H} P(x_{t+1}^w|x_t^w, a_t^R, a_t^H) P(x_{t+1}^H|x_t^H, a_t^R, a_t^H) P(a_t^H|x_t^H, a_t^R)
\end{aligned}
$$

Reward

$$R(*) = 0$$
$$R(RR) = 10$$
$$R(TR) = 5$$



x_t^H	a_t^H	$P(a_t^H = I \mid x_t^H, a_t^H)$
T	B	0.1
T	G	0.2
$-T$	B	0.3
$-T$	G	0.8

x_t^w	a_t^R	a_t^H	x_{t+1}^w
TT	B	I	HT
TT	B	$\neg I$	RT

$$x_t^H = \{T, \neg T\}$$

x_t^H	a_t^R	a_t^H	$P(x_t^H = T \mid x_t^H, a_t^H, a_t^R)$
T	B	I	1.0
T	B	$\neg I$	1.0
$\vdots$			
$-T$	B	$\neg I$	0.8
$-T$	G	$\neg I$	0.9

State S		$V_0(S)$	$V_1(S)$	$V_2(S)$
TT	¬Trust			
TR	″			
TH	″			
RT	″			
RR	″			
RH	″			
HT	″			
HR	″			
HH	″			
TT	Trust			
TR	″			
TH	″			
RT	″			
RR	″			
RH	″			
HT	″			
HR	″			
HH	″			

	State S	$V_0(S)$	$V_1(S)$	$V_2(S)$
TT	¬Trust	0		
TR	"	**5**		
TH	"	0		
RT	"	0		
RR	"	**10**		
RH	"	0		
HT	"	0		
HR	"	0		
HH	"	0		
TT	Trust	0		
TR	"	**5**		
TH	"	0		
RT	"	0		
RR	"	**10**		
RH	"	0		
HT	"	0		
HR	"	0		
HH	"	0		

Value Iteration

$$U_1(s_1) = max_{a_1^R}[R(s_1) + \gamma * \sum_{s_2} P(s_2|s_1, a_1^R) * V_0(s_2)]$$

$$\begin{aligned}
U_1(TT, -trust >) = max_{a_1^R}[&0 + P(RT, trust|TT, -trust, B)V(RT, trust) + ..., \\
&0 + P(TR, trust|TT, -trust, G)V(TR, trust) + P(TH, trust|TT, -trust, G)V(TH, trust) \\
&+ P(TR, -trust|TT, -trust, G)V(TR, -trust) \\
&+ P(TH, -trust|TT, -trust, G)V(TH, -trust)
\end{aligned}$$

Value Iteration

$$U_1(s_1) = max_{a_1^R}[R(s_1) + \gamma * \sum_{s_2} P(s_2|s_1, a_1^R) * V_0(s_2)]$$

$$\begin{aligned}
U_1(TT, -trust >) = max_{a_1^R}[&0 + P(RT, trust|TT, -trust, B)V(RT, trust) + ..., \\
&0 + P(TR, trust|TT, -trust, G)V(TR, trust) + P(TH, trust|TT, -trust, G)V(TH, trust) \\
&+ P(TR, -trust|TT, -trust, G)V(TR, -trust) \\
&+ P(TH, -trust|TT, -trust, G)V(TH, -trust)
\end{aligned}$$

$$\begin{aligned}
P(TR, trust|TT, -trust, G) = &P(TR|TT, G, intervene) * P(trust| - trust, G, intervene)P(intervene|trust, G) \\
&+ P(TR|TT, G, -intervene)* \\
&P(trust| - trust, G, -intervene)P(-intervene| - trust, G) \\
&= 0 + 0.9 * 0.2 = 0.18
\end{aligned}$$

Value Iteration

$$U_1(s_1) = max_{a_1^R}[R(s_1) + \gamma * \sum_{s_2} P(s_2|s_1, a_1^R) * V_0(s_2)]$$

$$\begin{aligned}
U_1(TT, -trust >) = max_{a_1^R}[&0 + P(RT, trust|TT, -trust, B)V(RT, trust) + ..., \\
&0 + P(TR, trust|TT, -trust, G)V(TR, trust) + P(TH, trust|TT, -trust, G)V(TH, trust) \\
&+ P(TR, -trust|TT, -trust, G)V(TR, -trust) \\
&+ P(TH, -trust|TT, -trust, G)V(TH, -trust)
\end{aligned}$$

$$\begin{aligned}
P(TR, trust|TT, -trust, G) = &P(TR|TT, G, intervene) * P(trust| - trust, G, intervene)P(intervene|trust, G) \\
&+ P(TR|TT, G, -intervene)* \\
&P(trust| - trust, G, -intervene)P(-intervene| - trust, G) \\
&= 0 + 0.9 * 0.2 = 0.18
\end{aligned}$$

$$\begin{aligned}
P(TR, -trust|TT, -trust, G) = &P(TR|TT, G, intervene) * P(-trust| - trust, G, intervene) \\
&P(intervene| - trust, G) + P(TR|TT, G, -intervene)* \\
&P(-trust| - trust, G, -intervene)P(-intervene| - trust, G) \\
&= 0 + 1 * 0.1 * 0.2 = 0.02
\end{aligned}$$

Value Iteration

$$U_1(s_1) = max_{a_1^R}[R(s_1) + \gamma * \sum_{s_2} P(s_2|s_1, a_1^R) * V_0(s_2)]$$

$$\begin{aligned}
U_1(TT, -trust >) = max_{a_1^R}[&0 + P(RT, trust|TT, -trust, B)V(RT, trust) + ..., \\
&0 + P(TR, trust|TT, -trust, G)V(TR, trust) + P(TH, trust|TT, -trust, G)V(TH, trust) \\
&+ P(TR, -trust|TT, -trust, G)V(TR, -trust) \\
&+ P(TH, -trust|TT, -trust, G)V(TH, -trust) = max_{a_1^R}[0, 0.18 * 5 + 0.02 * 5] = 1
\end{aligned}$$

$$\begin{aligned}
P(TR, trust|TT, -trust, G) = &P(TR|TT, G, intervene) * P(trust| - trust, G, intervene)P(intervene|trust, G) \\
&+ P(TR|TT, G, -intervene)* \\
&P(trust| - trust, G, -intervene)P(-intervene| - trust, G) \\
&= 0 + 0.9 * 0.2 = 0.18
\end{aligned}$$

$$\begin{aligned}
P(TR, -trust|TT, -trust, G) = &P(TR|TT, G, intervene) * P(-trust| - trust, G, intervene) \\
&P(intervene| - trust, G) + P(TR|TT, G, -intervene)* \\
&P(-trust| - trust, G, -intervene)P(-intervene| - trust, G) \\
&= 0 + 1 * 0.1 * 0.2 = 0.02
\end{aligned}$$

	State S	$V_0(S)$	$V_1(S)$	$V_2(S)$
TT	¬Trust	0	1	
TR	″	5	12	
TH	″	0	0	
RT	″	0	2	
RR	″	10	10	
RH	″	0	0	
HT	″	0	0	
HR	″	0	0	
HH	″	0	0	
TT	Trust	0	4	
TR	″	5	14	
TH	″	0	0	
RT	″	0	8	
RR	″	10	10	
RH	″	0	0	
HT	″	0	0	
HR	″	0	0	
HH	″	0	0	

	State S	$V_0(S)$	$V_1(S)$	$V_2(S)$
TT	¬Trust	0	1	4.76
TR	″	5	12	12
TH	″	0	0	0
RT	″	0	2	2
RR	″	10	10	10
RH	″	0	0	0
HT	″	0	0	0
HR	″	0	0	0
HH	″	0	0	0
TT	Trust	0	4	11.2
TR	″	5	14	14
TH	″	0	0	0
RT	″	0	8	8
RR	″	10	10	10
RH	″	0	0	0
HT	″	0	0	0
HR	″	0	0	0
HH	″	0	0	0