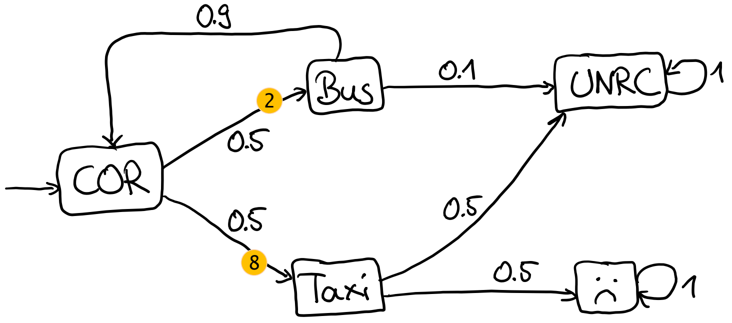
Formal Approaches to Decision-Making under Uncertainty

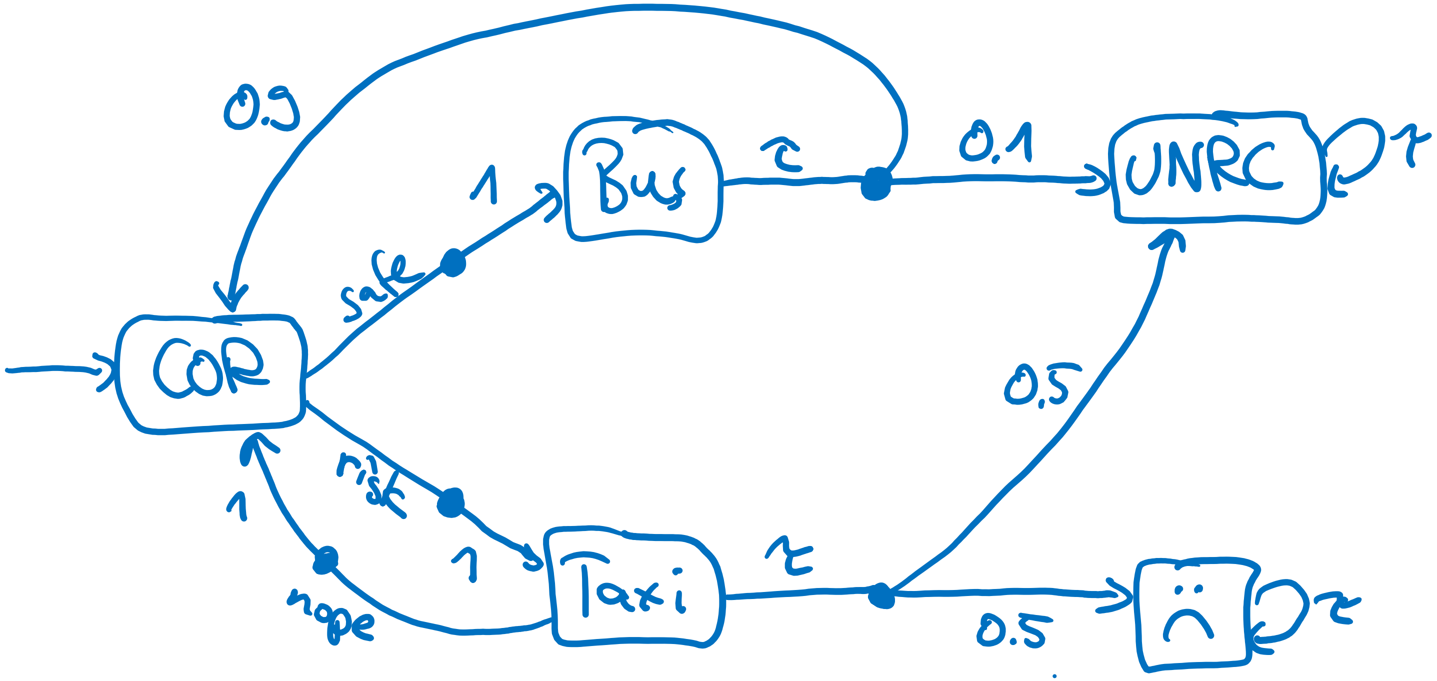
Lecture 2-2: Markov Decision Processes

Arnd Hartmanns
Formal Methods and Tools
UNIVERSITY OF TWENTE

Markov Decision Processes

A Markov decision process **MDP**
from Córdoba airport to Rio Cuarto:



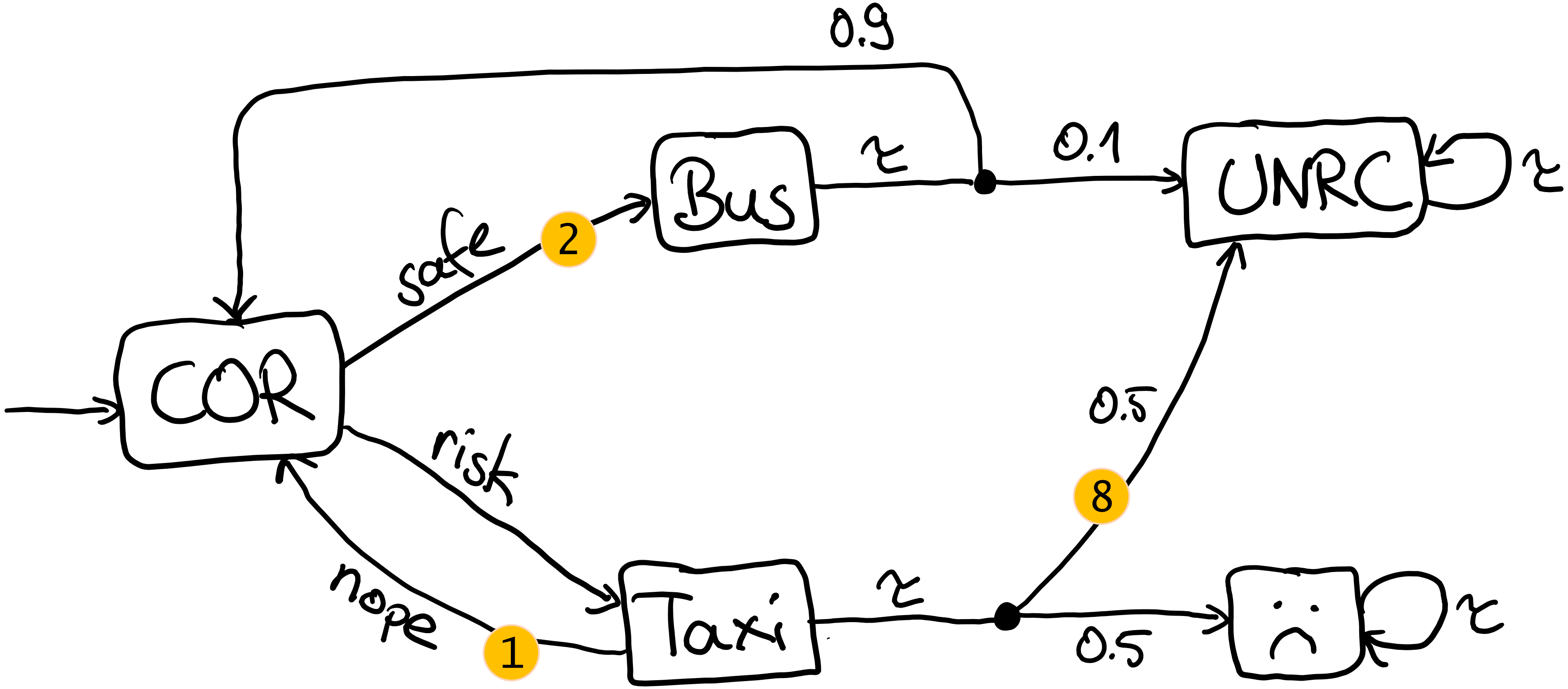


Markov Decision Processes

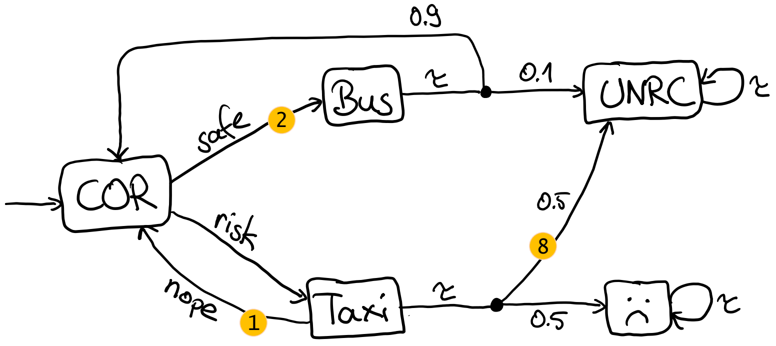
A Markov decision process **MDP**
from Córdoba airport to Rio Cuarto:

DTMC: $P(\Diamond\{UNRC\}) = ?$

here: $P_{max}(\Diamond\{UNRC\}) = ?$



Markov Decision Processes



An **MDP** is a 4-tuple $M = \langle S, A, T, s_I \rangle$ where

S are the states with initial state s_I,

A is the finite set of action names,

$T\colon S \to 2^{A \times \mathrm{Dist}(S)}$ is the nondeterministic transition function.

DTMC: $T\colon S \to \mathrm{Dist}(S)$

Shorthand for transitions: write $s \xrightarrow{a} \mu$ if $\langle a, \mu \rangle \in T(s)$.

Paths: $\overbrace{s_0}^{\text{state}}\ \overbrace{a_0\mu_0}^{\text{transition}}\ \overbrace{s_1}^{\text{next state}}\ a_1\mu_1 \ldots$ ...

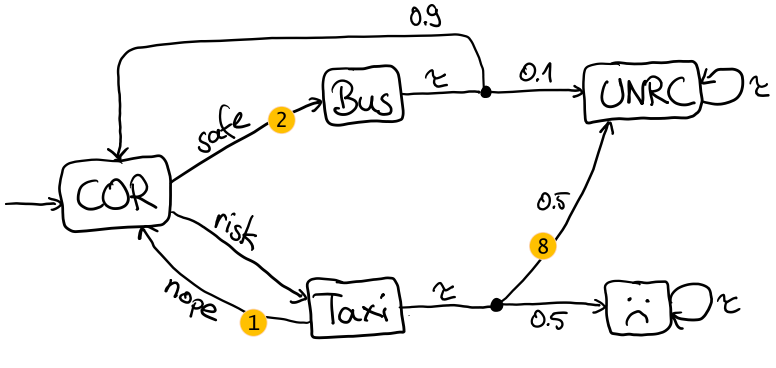
$T(\text{Taxi}) = \{ \langle \text{nope}, \{\text{COR} \mapsto 1\} \rangle, \langle \tau, \{\text{UNRC} \mapsto 0.5, ☹ \mapsto 0.5\} \rangle \}$

A **reward structure** maps *branches* of transitions to reward values:

$$R\colon S \times \big(A \times \mathrm{Dist}(S)\big) \times S \to \mathbb{R}$$

DTMC: $R\colon S \times S \to \mathbb{R}$

Markov Decision Processes



Typical restrictions:

– no deadlocks: $\forall s \in S\colon |T(s)| \geq 1$

– all branch probabilities in $\mathbb{Q}$

– action determinism (in ML and planning):

$$s \xrightarrow{a} \mu_1 \land s \xrightarrow{a} \mu_2 \;\Rightarrow \mu_1 = \mu_2$$

– discrete-time Markov chain (DTMC):

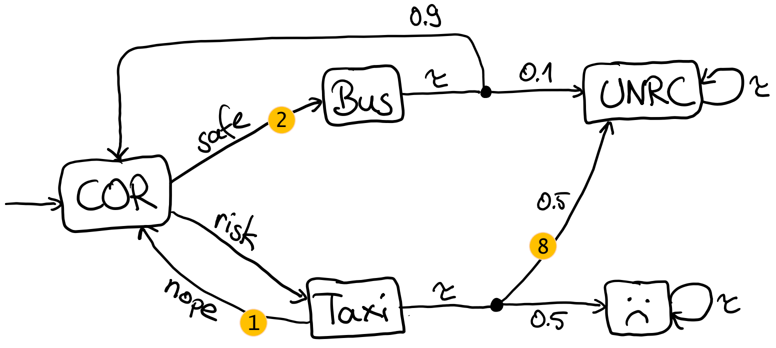
no nondeterminism $\rightarrow$ $\forall s \in S\colon |T(s)| = 1$

– rewards in $[0, \infty)$ or $\mathbb{N}$

– state rewards:

same reward on every incoming branch to a state

Markov Decision Makers



strategies, policies, adversaries

Schedulers resolve all nondeterminism; e.g.

$$\mathcal{S}: S \to A \times \mathrm{Dist}(S)$$

= memoryless deterministic scheduler

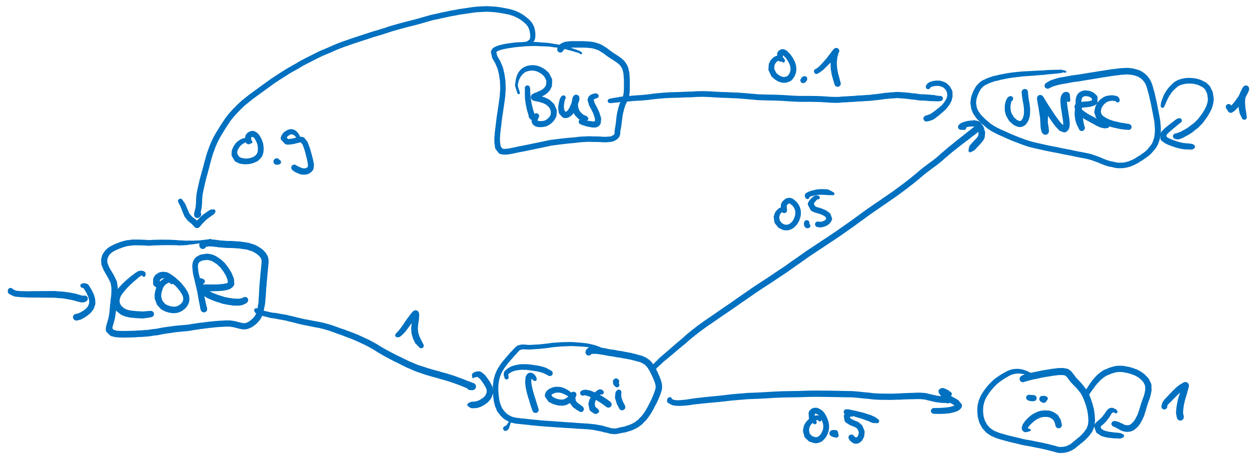
Example:

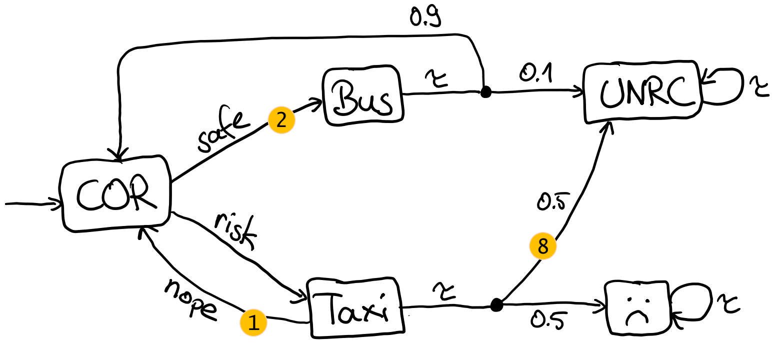
$\mathcal{S}(COR) = \langle risk, \{Taxi \mapsto 1\} \rangle$ simplify: risk

$\mathcal{S}(Bus) = \tau$ $\mathcal{S}(Taxi) = \tau$

definition of one of many possible schedulers

Induced DTMC $M|_{\mathcal{S}}$:





Unbounded probabilistic reachability:

$$\mathrm{P}_{opt}(\Diamond\, G)$$

for $opt \in \{\max, \min\}$ and $G \subseteq S$:

induced DTMC of S

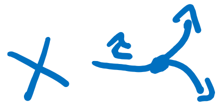
$$\mathrm{P}_{\max}(\Diamond\, G) = \sup_{\mathcal{S}} \mathbb{P}_{M|_{\mathcal{S}}}(\{s_0 \ldots \in \mathrm{Paths}(M) \mid \exists i\colon s_i \in G\})$$

$$\mathrm{P}_{\min}(\Diamond\, G) = \inf_{\mathcal{S}} \mathbb{P}_{M|_{\mathcal{S}}}(\{s_0 \ldots \in \mathrm{Paths}(M) \mid \exists i\colon s_i \in G\})$$

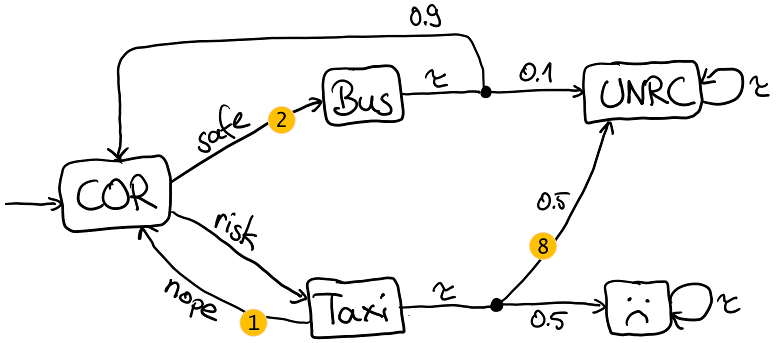
Examples:

$P_{max}(\Diamond \{☹\}) = 0.5$

$P_{min}(\Diamond \{☹\}) = 0$



Reachability Probabilities



Step-bounded probabilistic reachability:

$$\mathrm{P}_{opt}(\lozenge^{\leq b}\ G)$$

for $opt \in \{\max, \min\}$ and $G \subseteq S$:

$$\mathrm{P}_{\max}(\lozenge^{\leq b} G) = \sup_{\mathfrak{S}} \mathbb{P}_{M|_{\mathfrak{S}}}(\{s_0 \ldots \in \mathrm{Paths}(M) \mid \exists i < b : s_i \in G\})$$

$$\mathrm{P}_{\min}(\lozenge^{\leq b} G) = \inf_{\mathfrak{S}} \mathbb{P}_{M|_{\mathfrak{S}}}(\{s_0 \ldots \in \mathrm{Paths}(M) \mid \exists i < b : s_i \in G\})$$

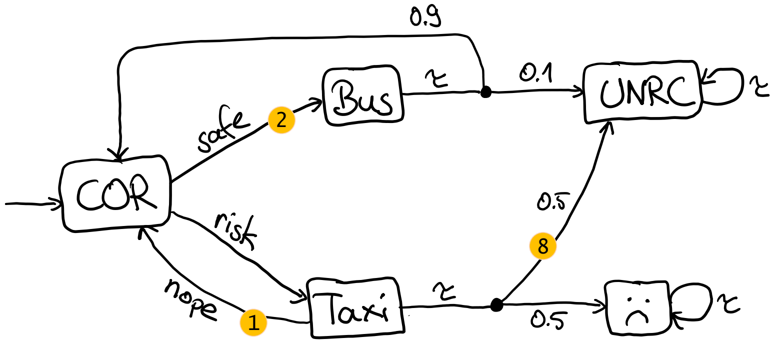
Example:

$$\mathrm{P}_{\max}(\lozenge^{\leq 2} \{UNRC\}) = 0.5$$

$$\text{—} \quad {}^{\leq 3} \quad \text{—} \quad = 0.5$$

$$\text{—} \quad {}^{\leq 4} \quad \text{—} \quad = 0.1 + 0.9 \cdot 0.5 = 0.55$$

Reachability Probabilities



Step-bounded probabilistic reachability:

$$\mathrm{P}_{opt}(\Diamond^{\leq b}\ G)$$

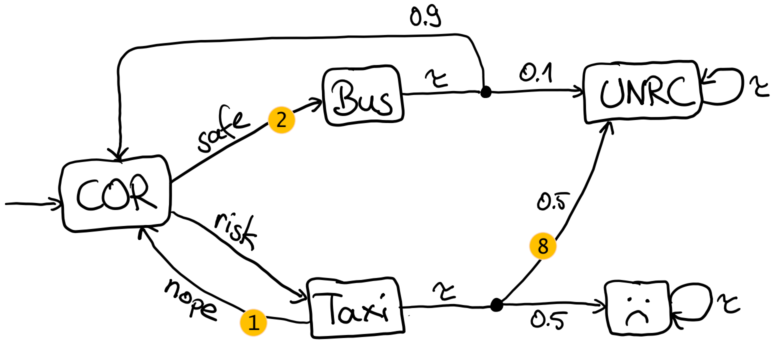
for $opt \in \{\max, \min\}$ and $G \subseteq S$:

$$\mathrm{P}_{\max}(\Diamond\ G) = \sup_{\mathcal{S}} \mathbb{P}_{M|_{\mathcal{S}}}(\{s_0 \ldots \in \mathrm{Paths}(M) \mid \exists i < b \colon s_i \in G\})$$

$$\mathrm{P}_{\min}(\Diamond\ G) = \ \inf_{\mathcal{S}} \mathbb{P}_{M|_{\mathcal{S}}}(\{s_0 \ldots \in \mathrm{Paths}(M) \mid \exists i < b \colon s_i \in G\})$$

Complication: need step-positional schedulers

$$\mathcal{S}_{st} \colon S \times \mathbb{N} \to A \times \mathrm{Dist}(S)$$



Step-bounded probabilistic reachability:

$$\mathrm{P}_{opt}(\diamond^{\leq b}\ G)$$

for $opt \in \{\max, \min\}$ and $G \subseteq S$:

$$\mathrm{P}_{\max}(\diamond\ G) = \sup_{\mathcal{S}_{st}} \mathbb{P}_{M|_{\mathcal{S}_{st}}}(\{s_0 \ldots \in \mathrm{Paths}(M) \mid \exists i < b\colon s_i \in G\})$$

$$\mathrm{P}_{\min}(\diamond\ G) = \inf_{\mathcal{S}_{st}} \mathbb{P}_{M|_{\mathcal{S}_{st}}}(\{s_0 \ldots \in \mathrm{Paths}(M) \mid \exists i < b\colon s_i \in G\})$$

Complication: need step-positional schedulers

$$\mathcal{S}_{st}\colon S \times \mathbb{N} \to A \times \mathrm{Dist}(S)$$

Expected Rewards

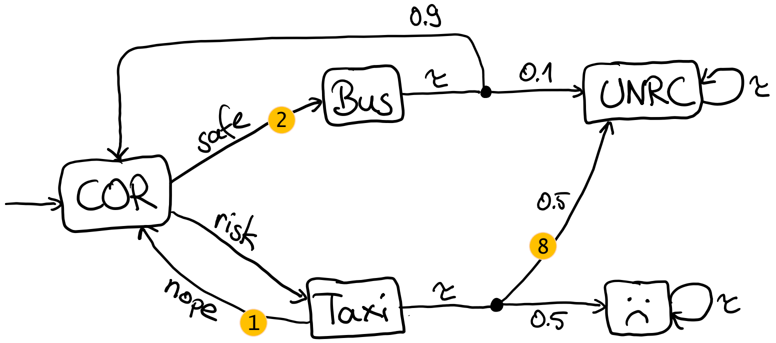
Expected accumulated **reward** to reach a goal:

$$\mathrm{R}_{opt}(\diamond\, G)$$

for $opt \in \{\max, \min\}$ and $G \subseteq S$:

$$\mathrm{R}_{\max}(\diamond\, G) = \sup_{\mathcal{S}} \mathbb{E}\left(rew_{\diamond G}^{M|_{\mathcal{S}}}\right)$$

$$\mathrm{R}_{\min}(\diamond\, G) = \inf_{\mathcal{S}} \mathbb{E}\left(rew_{\diamond G}^{M|_{\mathcal{S}}}\right)$$

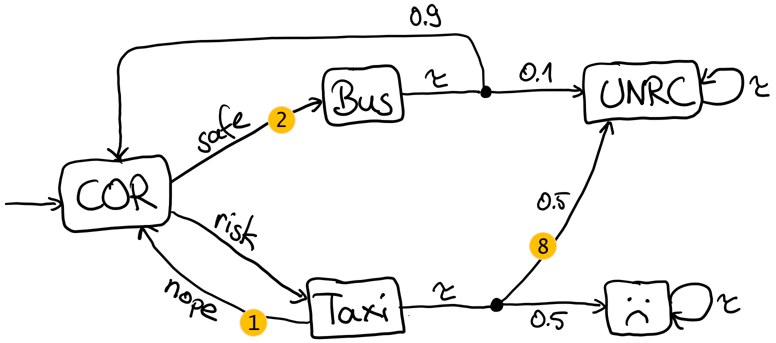


Example:

$$\mathrm{R}_{\min}(\diamond\{UNRC\}) = \min\{\infty, \infty, \underline{2 + 0.9\cdot 2 + 0.9^2\cdot 2 + \ldots}\}$$

$$\mathrm{R}_{\max}(\diamond\{UNRC\}) = \max\{ \text{———} \text{ " } \text{———} \} = \infty$$

Expected Rewards



Expected accumulated **reward** to reach a goal:

$$\mathrm{R}_{opt}(\diamond G)$$

for $opt \in \{\max, \min\}$ and $G \subseteq S$:

$$\mathrm{R}_{\max}(\diamond G) = \sup_{\mathcal{S}} \mathbb{E}\left(rew_{\diamond G}^{M|_{\mathcal{S}}}\right)$$

$$\mathrm{R}_{\min}(\diamond G) = \inf_{\mathcal{S}} \mathbb{E}\left(rew_{\diamond G}^{M|_{\mathcal{S}}}\right)$$

→ $\mathrm{R}_{\max}(\diamond G) = \infty$ if $P_{\min}(\diamond G) < 1$

$\mathrm{R}_{\min}(\diamond G) = \infty$ if $P_{\max}(\diamond G) < 1$