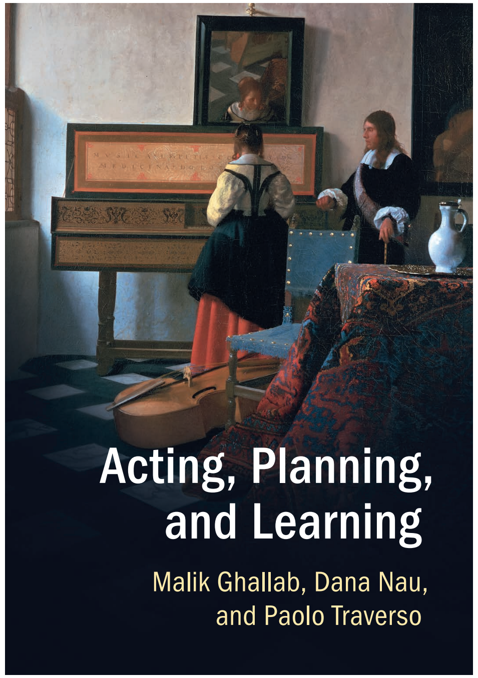
Last update: 11:09 PM, March 7, 2025




# Chapter 8
# Probabilistic Representation and Acting

Dana S. Nau
University of Maryland

with contributions from
Mark “mak” Roberts

# Motivation

- Situations where actions have multiple possible outcomes and each outcome has a *known* probability distribution of occurring
  - *Part IV: Non-deterministic Models* addresses multiple actions outcomes with *unknown* probability distributions

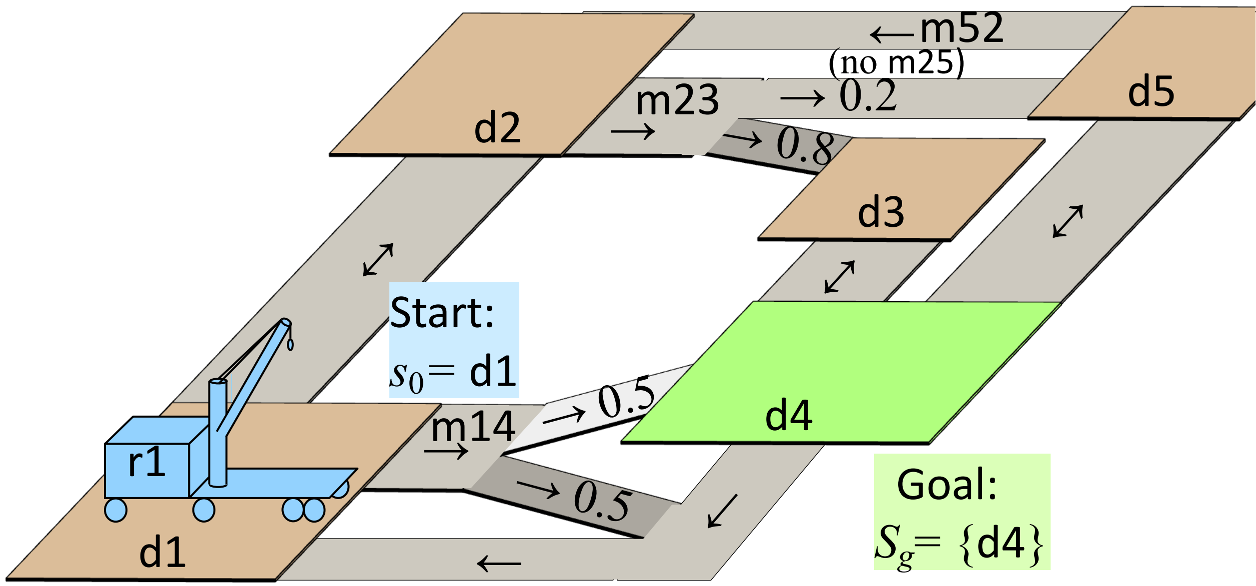
- Several possible action representations
  - Bayes nets, probabilistic actions, …
- Book doesn't commit to any representation
  - Mainly concentrates on the underlying semantics

Credit: Dennis Hill, CC BY 2.0

roll-die($d$)
 pre: holding($d$) = true
 eff:
  1/6: top($d$) $\leftarrow$ 1
  1/6: top($d$) $\leftarrow$ 2
  1/6: top($d$) $\leftarrow$ 3
  1/6: top($d$) $\leftarrow$ 4
  1/6: top($d$) $\leftarrow$ 5
  1/6: top($d$) $\leftarrow$ 6

# Definitions and Example

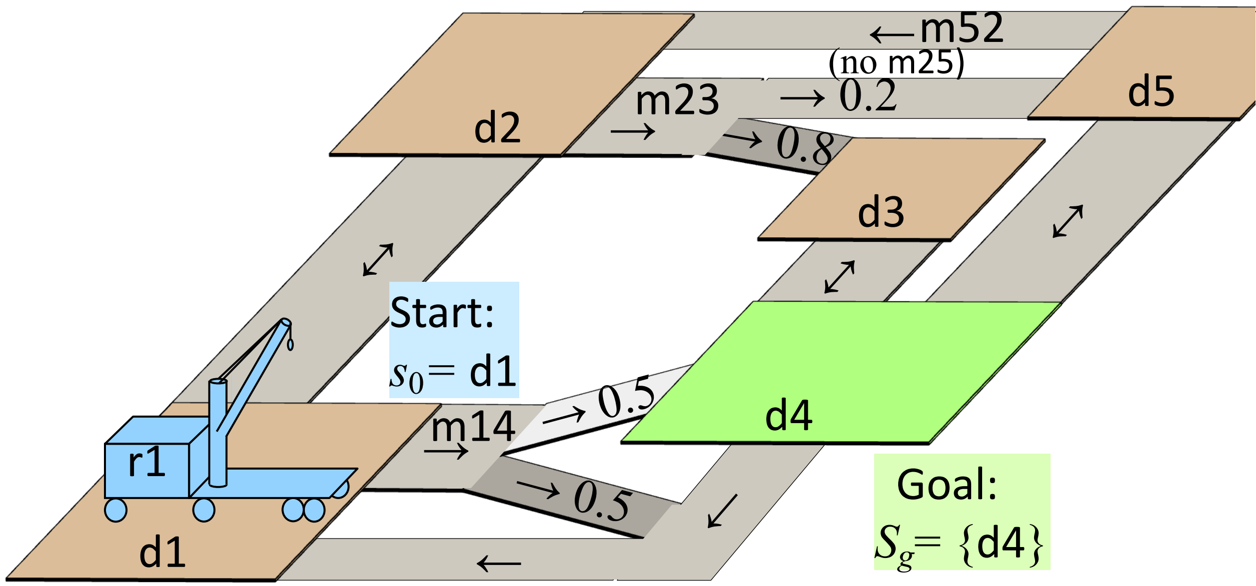
- Probabilistic domain model: $\Sigma = (S, A, \gamma, \Pr, \text{cost})$
  - $S$ and $A$ – finite sets of states and actions
  - $\gamma: S \times A \to 2^S$
- $\gamma(s,a)$ = {all possible "next states" after applying action $a$ in state $s$}
  - $a$ is applicable in state $s$ iff $\gamma(s,a) \neq \emptyset$
- $\Pr(s' \mid s, a)$ = probability that $a$ will take us to $s'$ from $s$
  - $\Pr(s' \mid s, a) \neq 0$ iff $s' \in \gamma(s,a)$
- cost: $S \times A \times S \to \mathbb{R}$
  - cost$(s,a,s')$ = cost if $a$ takes us to $s'$ from $s$
  - may omit, default is cost$(s,a,s') = 1$
- Applicable$(s)$ = {all actions applicable in $s$}
  $= \{a \in A \mid \gamma(s,a) \neq \emptyset\}$




- Start at d1, want to get to d4
- Some roads are one-way, some are two-way
- Unreliable steering when the road forks
  - may take the wrong fork
- Simplified state and action names:
  - write {loc(r1)=d2} as d2
  - write move(r1,d2,d3) as m23

# Example

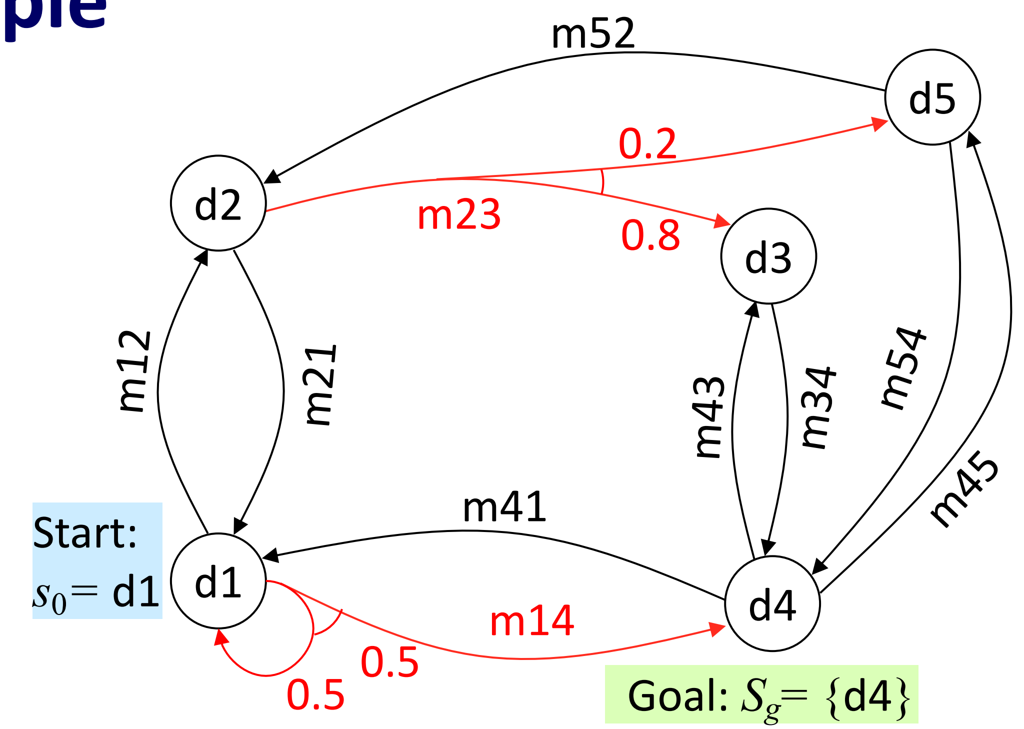
- γ(d1,m12) = {d2}
  - Pr(d2 | d1, m12) = 1
- m21, m34, m41, m43, m45, m52, m54:
  - deterministic like m12
- γ(d1,m14) = {d1,d4}
  - Pr(d4 | d1, m14) = 0.5
  - Pr(d1 | d1, m14) = 0.5
- γ(d2,m23) = {d3,d5}
  - Pr(d3 | d2, m23) = 0.8
  - Pr(d5 | d2, m23) = 0.2
- there's no m25




- Start at d1, want to get to d4
- Some roads are one-way, some are two-way
- Unreliable steering when the road forks
  - may take the wrong fork
- Simplified state and action names:
  - write {loc(r1)=d2} as d2
  - write move(r1,d2,d3) as m23

# Example

- $\gamma$(d1,m12) = {d2}
  - Pr(d2 | d1, m12) = 1
- m21, m34, m41, m43, m45, m52, m54:
  - deterministic like m12
- $\gamma$(d1,m14) = {d1,d4}
  - Pr(d4 | d1, m14) = 0.5
  - Pr(d1 | d1, m14) = 0.5
- $\gamma$(d2,m23) = {d3,d5}
  - Pr(d3 | d2, m23) = 0.8
  - Pr(d5 | d2, m23) = 0.2
- there's no m25




- We will represent these problems as a graph
  - Nodes are assignments to variables (i.e., states)
  - Weighted edges change the assignment (i.e, actions)
    - Label is action instance; value indicates Pr(s'| s,a)
- Simplified state and action names:
  - write {loc(r1)=d2} as d2
  - write move(r1,d2,d3) as m23

# Policies

- *Policy*: function $\pi : S' \rightarrow A$ where $S' \subseteq S$
  - require $\pi(s) \in \text{Applicable}(s)$ for every $s \in S'$
  - $Domain(\pi) = S'$
- *Transitive closure*
  - $\hat{\gamma}(s_0,\pi) =$ {all states reachable from $s_0$ using $\pi$}
    = union of the following sets

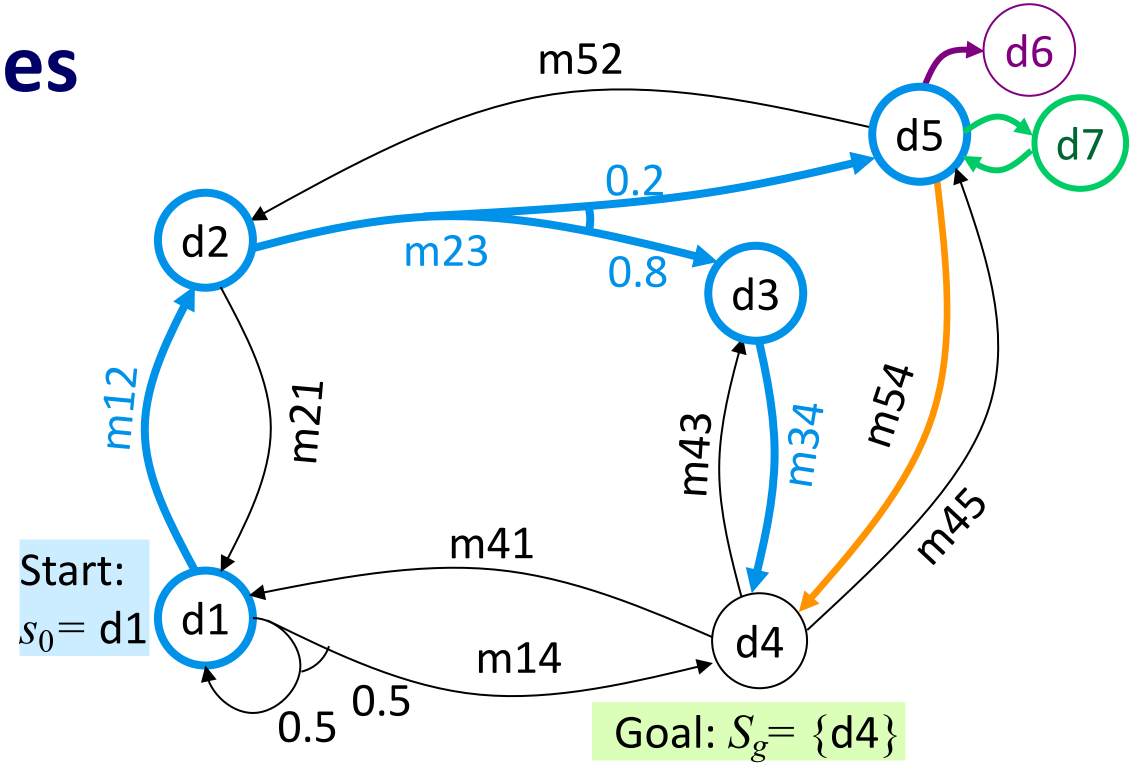
    $S_0 = \{s_0\}$

    $S_1 =$ {states reachable from $S_0$} $= \bigcup\{\gamma(s,\pi(s)) \mid s \in S_0\}$

    $S_2 =$ {states reachable from $S_1$} $= \bigcup\{\gamma(s,\pi(s)) \mid s \in S_1\}$

    …
- *Reachability graph*: $Graph(s,\pi) = (V,E)$
  - $V = \hat{\gamma}(s,\pi)$
  - $E = \{(s,s') \mid s \in V,\ s' \in \gamma(s,\pi(s))\}$
- $leaves(s,\pi) = \hat{\gamma}(s,\pi) \setminus \text{Domain}(\pi)$
  - may be empty

Set minus




**Poll**: Can we use a plan (sequence of actions) instead?
A. yes B. no
C. don't know

$\pi_1 = \{$(d1,m12), (d2,m23), (d3,m34)$\}$

$\pi_2 = \{$(d1,m12), (d2,m23), (d3,m34), (d5,m54)$\}$

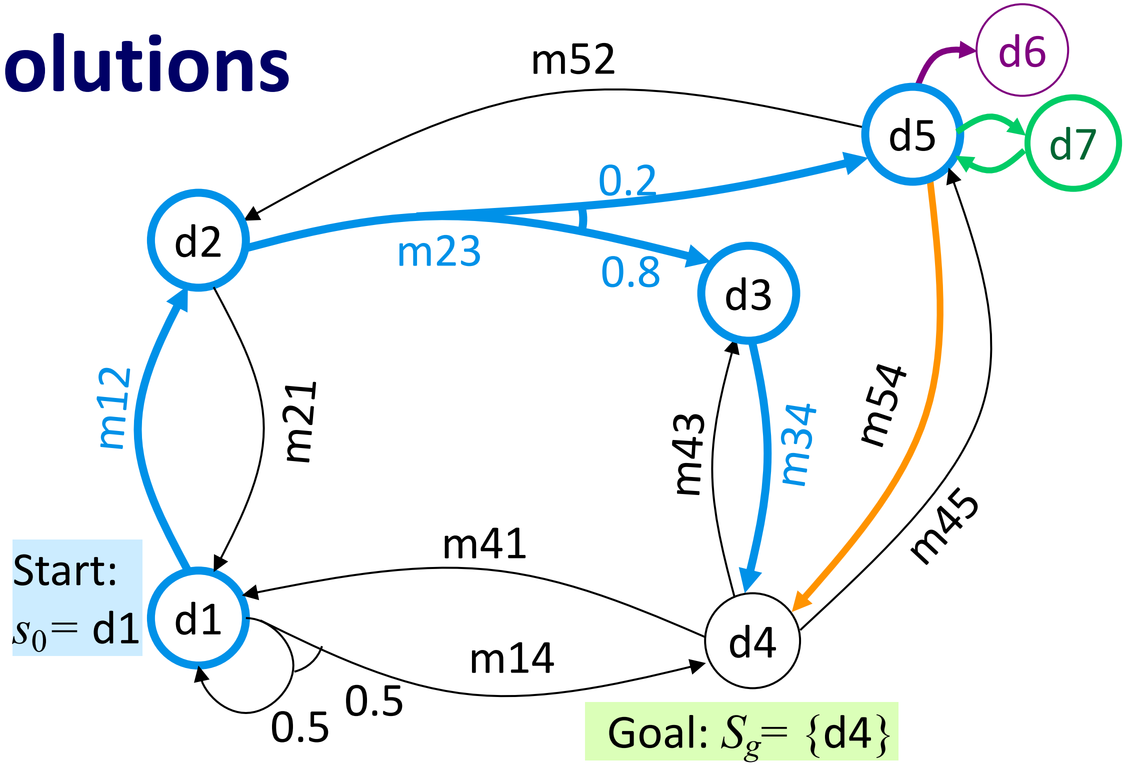
$\pi_3 = \{$(d1,m12), (d2,m23), (d3,m34), (d5,m56)$\}$

$\pi_4 = \{$(d1,m12), (d2,m23), (d3,m34), (d5,m57), (d7,m75)$\}$

**Poll**: What are the leaves of $\pi_3$?

# Problems, Solutions

- MDP problem: $P = (\Sigma, s_0, S_g)$, require $s_0 \notin S_g$
  - This is a specific type of MDP problem called a *goal reachability* problem
  - More generally, MDPs specify a set of terminal states
- *Solution* for $(\Sigma, s_0, S_g)$:
  - A policy π such that $leaves(s_0,\pi) \cap S_g \neq \emptyset$
- A solution policy π is *closed* if it doesn't stop at non-goal states unless there's no way to continue
  - for every state $s$ in $\hat{\gamma}(s_0,\pi)$, either
    $s \in Domain(\pi)$ (i.e., $\pi(s)$ is defined)
    or $s \in S_g$
    or Applicable$(s) = \emptyset$




**Poll**. Suppose d3 was the goal instead. Which policies are closed wrt. d3?

$\pi_1$ = {(d1,m12), (d2,m23), (d3,m34)}

$\pi_2$ = {(d1,m12), (d2,m23), (d3,m34), (d5,m54)}

$\pi_3$ = {(d1,m12), (d2,m23), (d3,m34), (d5,m56)}

$\pi_4$ = {(d1,m12), (d2,m23), (d3,m34), (d5,m57), (d7,m75)}

**Poll**. Is $\pi_1$ a solution?
A. yes B. no C. don't know

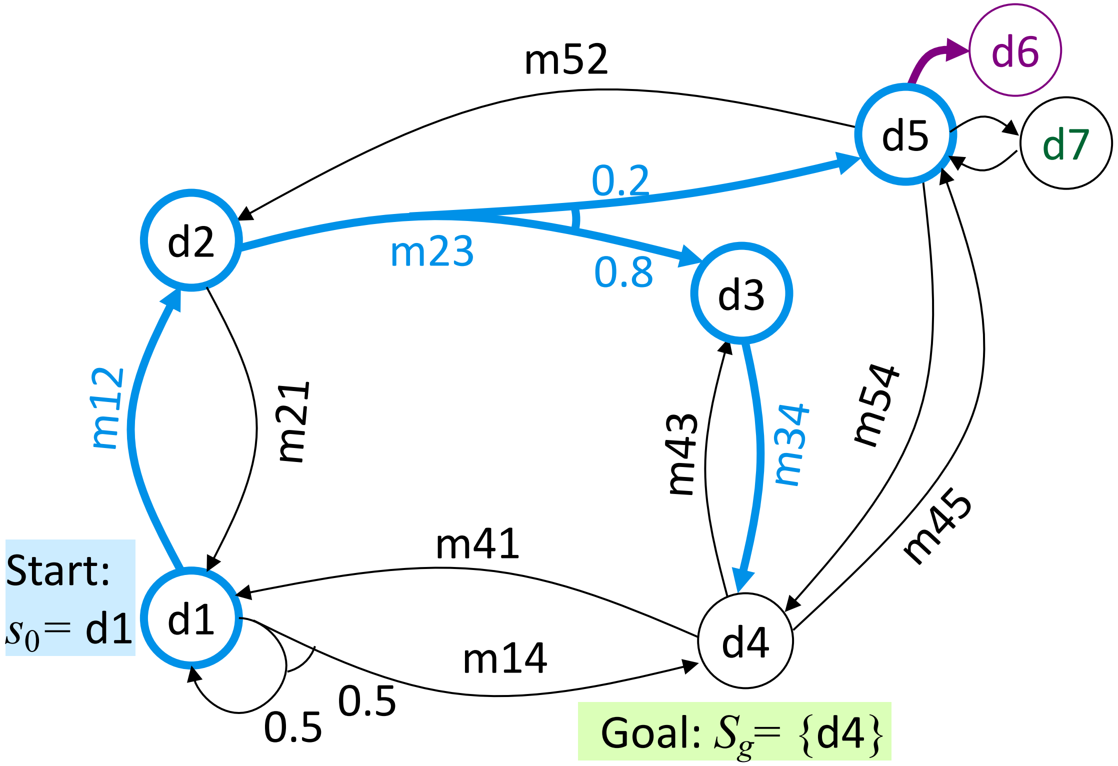
**Poll.** Is $\pi_1$ a closed solution?

# Histories

Run-Policy($\Sigma$, $s_0$, $S_g$, $\pi$)
  $s \leftarrow s_0$
  **while** $s \notin S_g$ and $s \in Domain(\pi)$ **do**
    perform action $\pi(s)$
    $s \leftarrow$ observe resulting state

- *History*: sequence of states $\sigma = \langle s_0, s_1, s_2, \ldots \rangle$ produced by Run-Policy
  - May be finite or infinite
- Let $H(s,\pi)$ = {all possible histories from $s$ using $\pi$}
- If $\sigma \in H(s,\pi)$ then
  - $\Pr(\sigma \mid s, \pi) = \prod_{s_i, s_{i+1} \in \sigma} \Pr(s_{i+1} \mid s_i, \pi(s_i))$
    = product of probabilities of state transitions
- $\sum_{\sigma \in H(s,\pi)} \Pr(\sigma \mid s, \pi) = 1$

**Poll.** If $s \notin Domain(\pi)$ then what is $H(s,\pi)$?
A. undefined  B. $\emptyset$  C. $\{\langle\rangle\}$  D. $\{s\}$
E. $\{\langle s \rangle\}$  F. other  G. unsure




- $\pi_3$ = {(d1,m12), (d2,m23), (d3,m34), (d5,m56)}
- $H(s_0, \pi_3) = \{\sigma_1, \sigma_2\}$, where:
  - $\sigma_1$ = ⟨d1,d2,d3,d4⟩
  - $\sigma_2$ = ⟨d1,d2,d5,d6⟩
- $\Pr(\sigma_1 \mid s_0, \pi_3) = 1 \times 0.8 \times 1 = 0.8$
- $\Pr(\sigma_2 \mid s_0, \pi_3) = 1 \times 0.2 \times 1 = 0.2$

# Unsafe Solutions

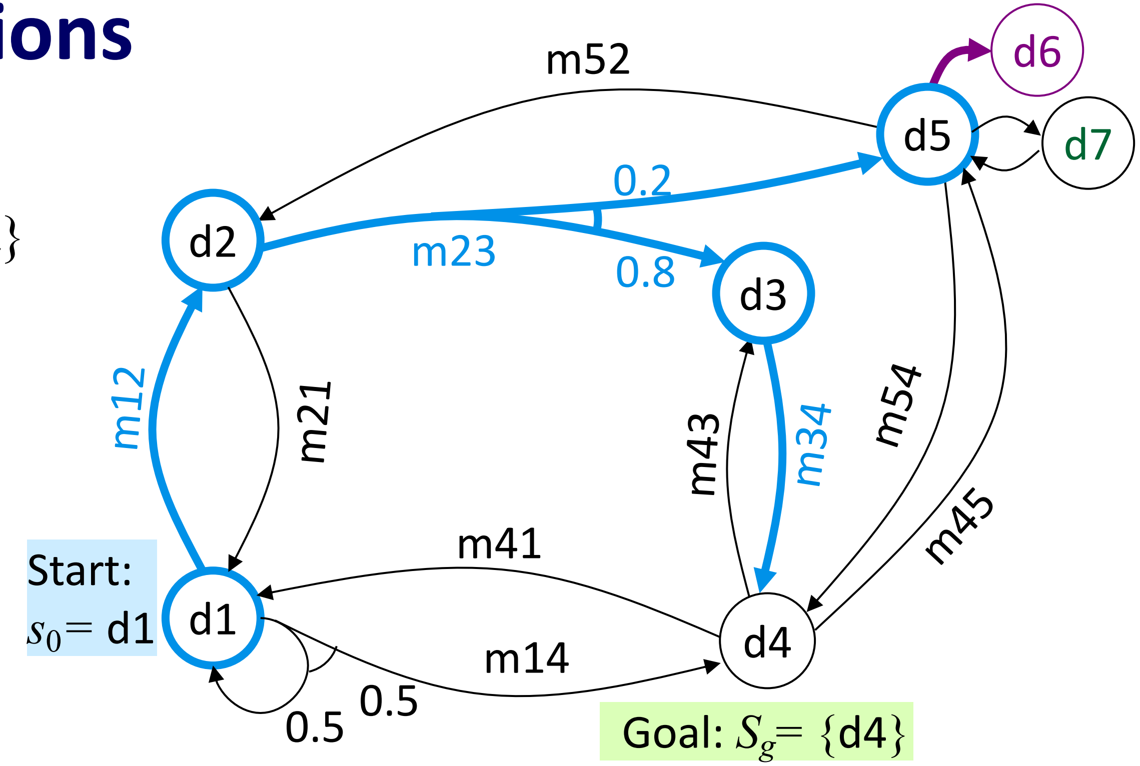
- Probability of reaching a goal state:

  $\Pr(S_g \mid s, \pi) = \sum_{\sigma \in H(s,\pi)} \{\Pr(\sigma \mid s, \pi) \mid \sigma \text{ ends at a state in } S_g\}$

- Equivalently:

$$\Pr(S_g \mid s, \pi) = \begin{cases} 1, & \text{if } s \in S_g \\ \sum_{s' \in \gamma(s,\pi(s))} \Pr(S_g \mid s', \pi), & \text{otherwise} \end{cases}$$

- A solution is *unsafe* if $0 < \Pr(S_g \mid s_0, \pi) < 1$




- $\pi_3 = \{$(d1,m12), (d2,m23), (d3,m34), (d5,m56)$\}$

- $H(s_0, \pi_3) = \{\sigma_1, \sigma_2\}$:
  - $\sigma_1 = \langle$d1,d2,d3,d4$\rangle$ ends at a goal state; $\Pr(\sigma_1 \mid s_0, \pi_3) = 1 \times 0.8 \times 1 = 0.8$
  - $\sigma_2 = \langle$d1,d2,d5,d6$\rangle$ doesn't; $\Pr(\sigma_2 \mid s_0, \pi_3) = 1 \times 0.2 \times 1 = 0.2$

- $\Pr(S_g \mid s_0, \pi_3) = \Pr(\sigma_1 \mid s_0, \pi_3) = 0.8$

# Unsafe Solutions

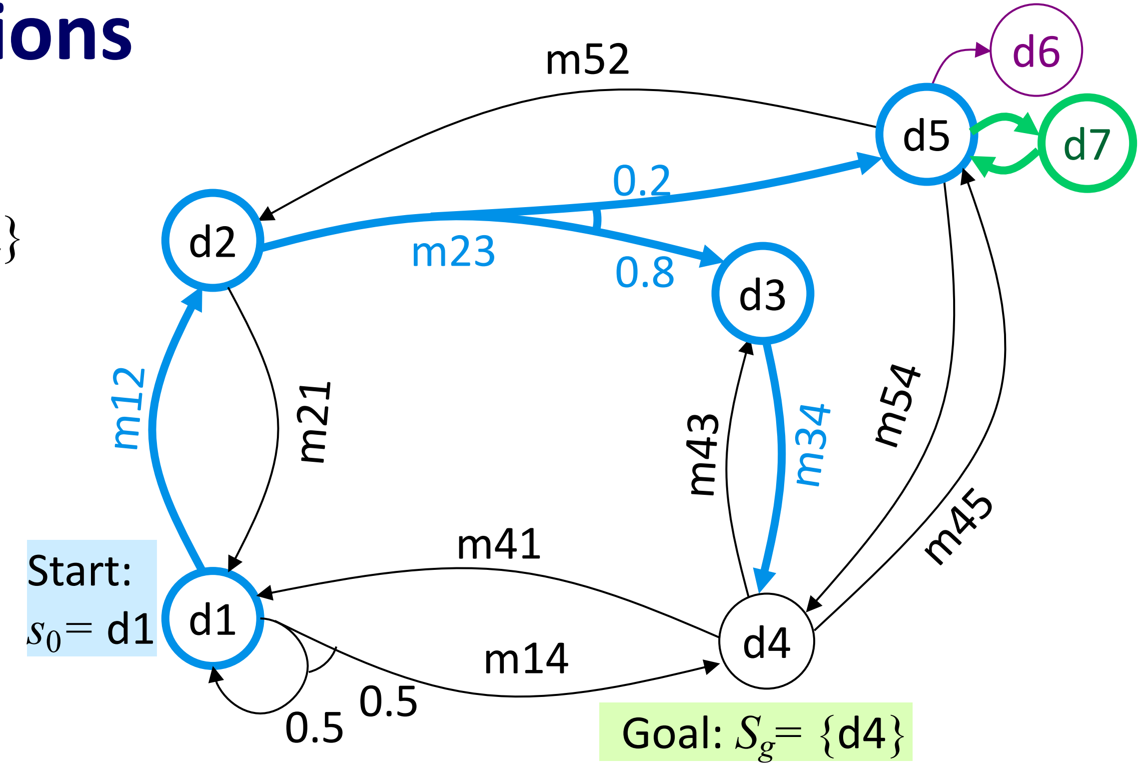
- Probability of reaching a goal state:

  $\Pr(S_g \mid s, \pi) = \sum_{\sigma \in H(s,\pi)} \{\Pr(\sigma \mid s, \pi) \mid \sigma \text{ ends at a state in } S_g\}$

- Equivalently:

$$\Pr(S_g \mid s, \pi) = \begin{cases} 1, & \text{if } s \in S_g \\ \sum_{s' \in \gamma(s,\pi(s))} \Pr(S_g \mid s', \pi), & \text{otherwise} \end{cases}$$

- A solution is *unsafe* if $0 < \Pr(S_g \mid s_0, \pi) < 1$




  - $\pi_4$ = {(d1,m12), (d2,m23), (d3,m34), (d5,m57), (d7,m75)}

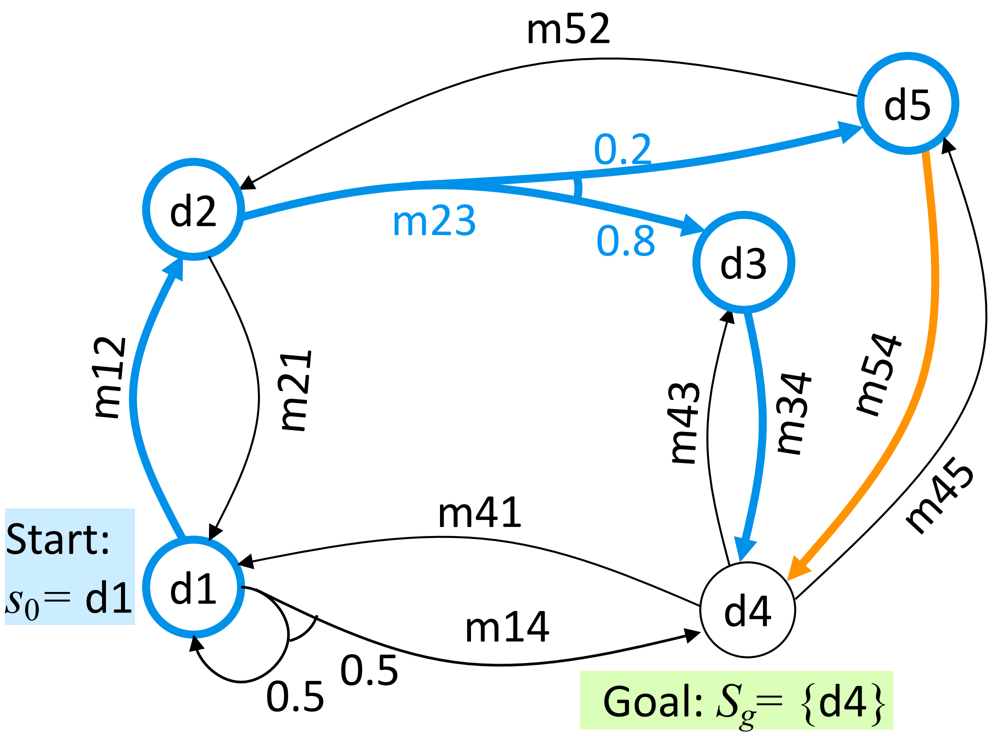
  - $H(s_0, \pi_4) = \{\sigma_1, \sigma_2\}$:
    - $\sigma_1$ = ⟨d1,d2,d3,d4⟩ ends at a goal state; $\Pr(\sigma_1 \mid s_0, \pi_4) = 1 \times .8 \times 1 = 0.8$
    - $\sigma_3$ = ⟨d1,d2,d5,d7,d5,d7,…⟩ doesn't; $\Pr(\sigma_3 \mid s_0, \pi_4) = 1 \times .2 \times 1 \times 1 \times 1 \times \ldots = 0.2$

  - $\Pr(S_g \mid s_0, \pi_4) = \Pr(\sigma_1 \mid s_0, \pi_4) = 0.8$

# Safe Solutions

- A solution is *safe* if $\Pr(S_g \mid s_0, \pi) = 1$

- An *acyclic* safe solution:
  - $\pi_2 = \{$(d1, m12), (d2, m23), (d3, m34), (d5, m54)$\}$




- $H(s_0, \pi_2) = \{\sigma_1, \sigma_2\}$, where:
  - $\sigma_1 = \langle$d1, d2, d3, d4$\rangle$ $\quad \Pr(\sigma_1 \mid s_0, \pi_2) = 1 \times .8 \times 1 = .8$
  - $\sigma_4 = \langle$d1, d2, d5, d4$\rangle$ $\quad \Pr(\sigma_4 \mid s_0, \pi_2) = 1 \times .2 \times 1 = .2$

- $\Pr(S_g \mid s_0, \pi_2) = .8 + .2 = 1$

# Safe Solutions

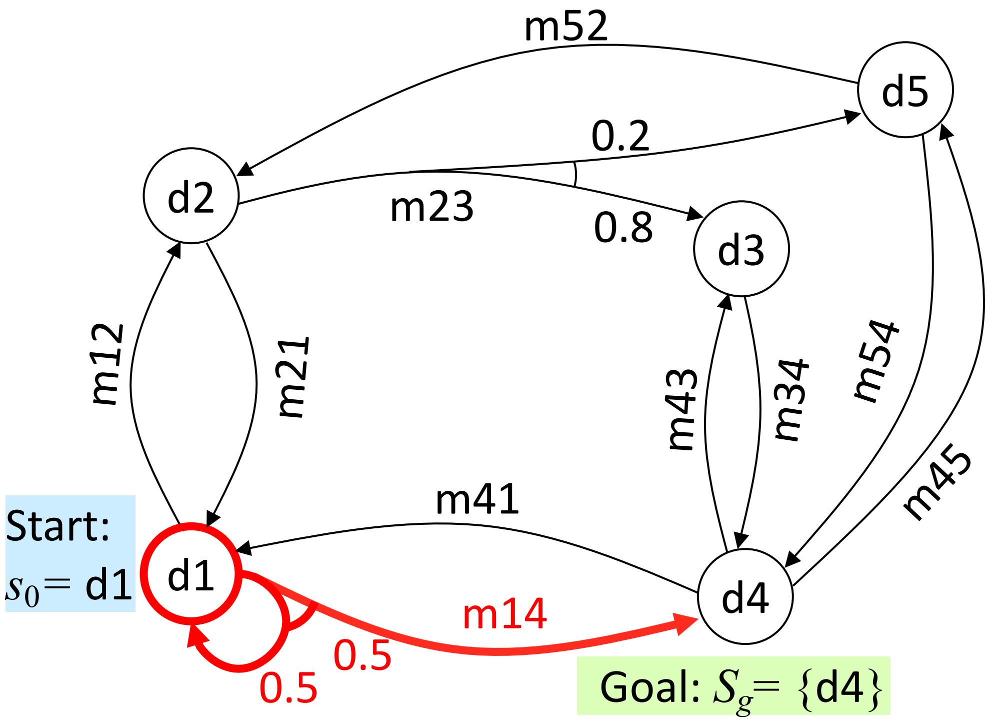
- A solution is *safe* if $\Pr(S_g \mid s_0, \pi) = 1$

- A *cyclic* safe solution:
  - $\pi_5 = \{(\mathrm{d1}, \mathrm{m14})\}$

- $H(s_0, \pi_5)$ contains infinitely many histories:
  - $\sigma_5 = \langle \mathrm{d1}, \mathrm{d4} \rangle$ $\quad \Pr(\sigma_5 \mid s_0, \pi_5) = ½$
  - $\sigma_6 = \langle \mathrm{d1}, \mathrm{d1}, \mathrm{d4} \rangle$ $\quad \Pr(\sigma_6 \mid s_0, \pi_5) = (½)^2 = ¼$
  - $\sigma_7 = \langle \mathrm{d1}, \mathrm{d1}, \mathrm{d1}, \mathrm{d4} \rangle$ $\quad \Pr(\sigma_7 \mid s_0, \pi_5) = (½)^3 = {}^1/_8$

    …
  - $\sigma_\infty = \langle \mathrm{d1}, \mathrm{d1}, \mathrm{d1}, \mathrm{d1}, \mathrm{d1}, \ldots \rangle$

- $\Pr(S_g \mid s_0, \pi_5) = ½ + ¼ + {}^1/_8 + \ldots = 1$




Poll: what is $\Pr(\sigma_\infty \mid s_0, \pi_5)$?

A. 1

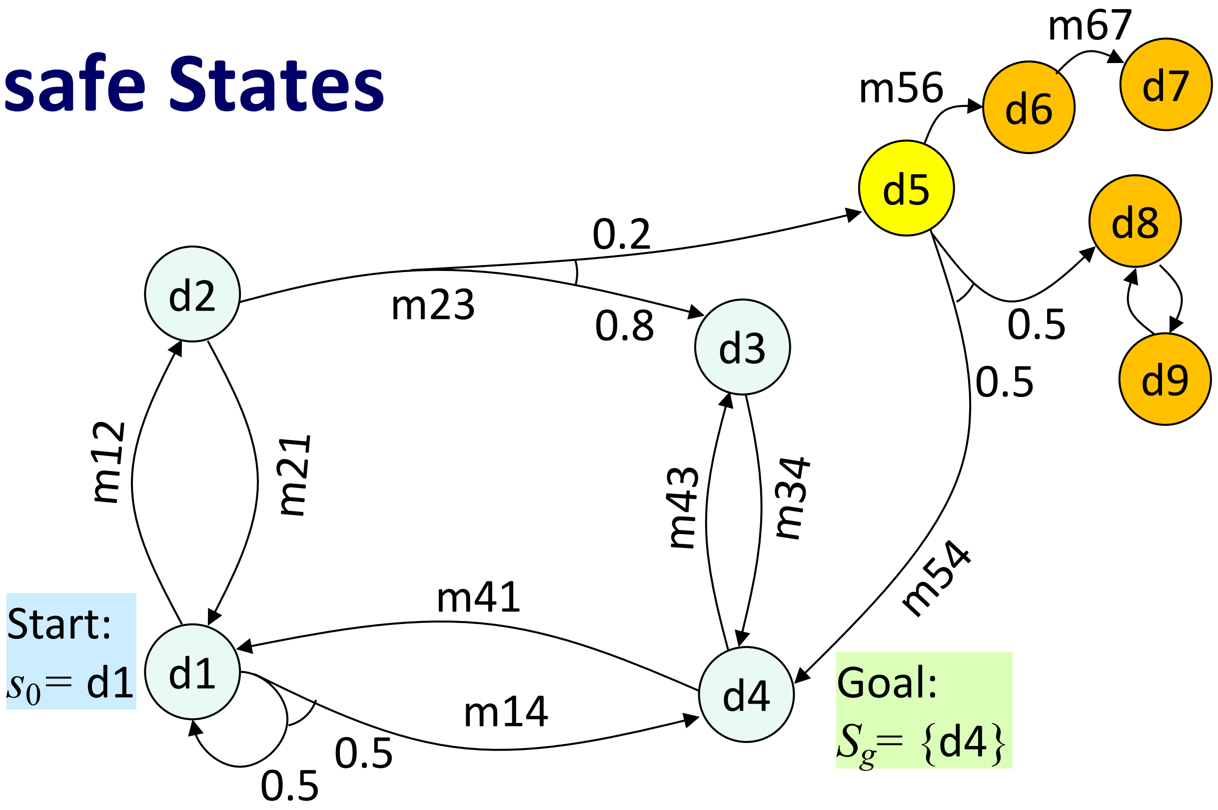
B. 0

C. a number between 0 and 1

D. undefined

# Safe and Unsafe States

- $s$ is *safe* if $\exists\pi$ such that $\Pr(S_g \mid s, \pi) = 1$
  - same as saying $(\Sigma, s, S_g)$ has a safe solution
  - d1, d2, d3, d4
- $s$ is *unsafe* if $\exists\pi$ s.t. $\Pr(S_g \mid s, \pi) > 0$
  and $\forall\pi$, $\Pr(S_g \mid s, \pi) < 1$
  - same as saying $(\Sigma, s, S_g)$ has an unsafe solution but no safe solution
    - d5
- $s$ is a *dead end* if $\forall\pi$, $\Pr(S_g \mid s, \pi) = 0$
  - same as saying $(\Sigma, s, S_g)$ has no solution
    - d6, d7, d8, d9
- An MDP is *safe* if all of its states are safe

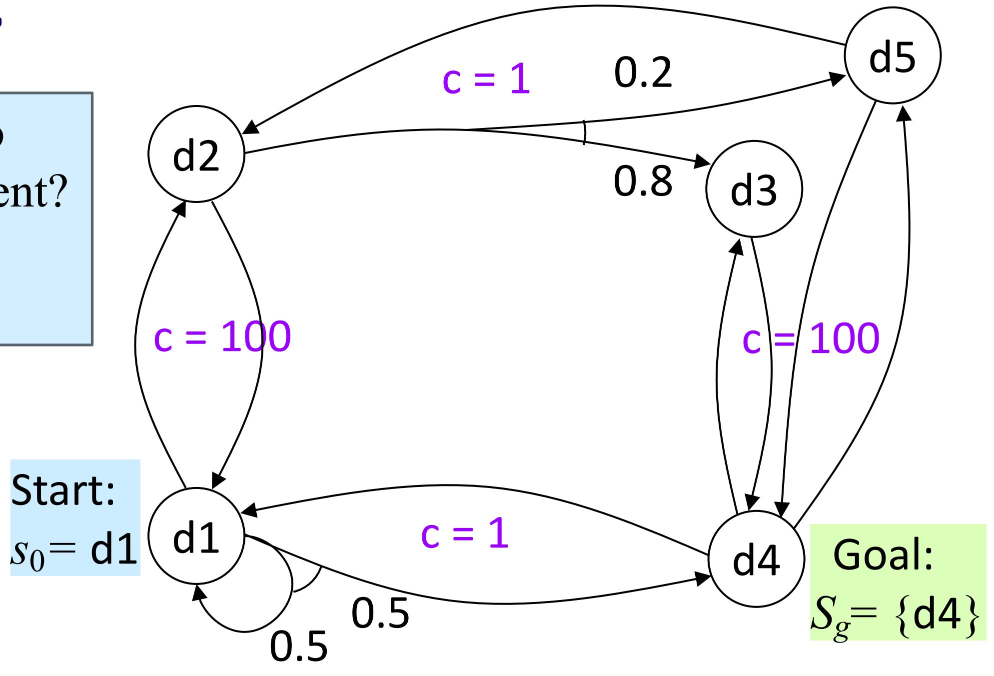



- d7 is an *immediate* dead end
  - No applicable actions
- d6, d8, d9 are *deep* dead ends
  - Applicable actions, but can't reach $S_g$

# Expected Cost

- cost($s,a,s'$) = cost of using $a$ in $s$
- Extend example so that:
  - each "horizontal" action costs 1
  - each "vertical" action costs 100

**Poll**: Are the two versions equivalent?
A. yes
B. no




- Let $\sigma = \langle s_0, s_1, s_2, \ldots \rangle \in H(s_0,\pi)$
  - i.e., starting at $s_0$, π can produce history $\sigma$
- Then cost($\sigma$) = $\sum_i$ cost($s_i$,π($s_i$))
- Let $\pi$ be a safe solution, i.e., $\Pr(S_g | s_0,\pi) = 1$
- At each state $s \in Domain(\pi)$, expected cost of following π to goal:
  - Weighted sum of history costs:
    - $V^\pi(s) = \sum_{\sigma \in H(s,\pi)} \Pr(\sigma \mid s, \pi)\ \mathrm{cost}(\sigma)$

My version

  - Recursive equation

From the book

$$V^\pi(s) = \begin{cases} 0, & \text{if } s \in S_g \\ \sum_{s' \in \gamma(s,\pi(s))} \Pr(s' \mid s, \pi(s))[\mathrm{cost}(s,\pi(s),s') + V^\pi(s')], & \text{otherwise} \end{cases}$$

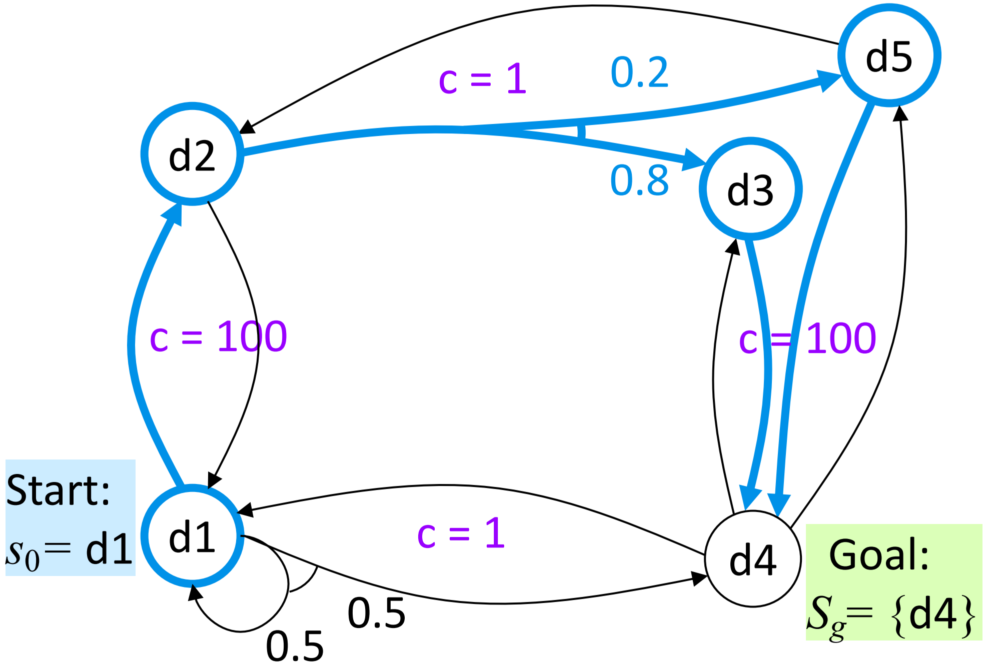
Run-Policy($\Sigma$, $s_0$, $S_g$, π)
  $s \leftarrow s_0$
  **while** $s \notin S_g$ and $s \in Domain(\pi)$ **do**
    perform action $\pi(s)$
    $s \leftarrow$ observe resulting state

# Example

- $\pi_3 = \{$(d1, m12), (d2, m23), (d3, m34), (d5, m54)$\}$

- Weighted sum of history costs:
  - $\sigma_1 = \langle$d1, d2, d3, d4$\rangle$
    - $\Pr(\sigma_1 \,|\, s_0, \pi_3) = 0.8$
    - $\text{cost}(\sigma_1) = 100 + 1 + 100 = 201$
  - $\sigma_2 = \langle$d1, d2, d5, d4$\rangle$
    - $\Pr(\sigma_2 \,|\, s_0, \pi_3) = 0.2$
    - $\text{cost}(\sigma_2) = 100 + 1 + 100 = 201$

- $V^{\pi_3}(\text{d1}) = .8(201) + .2(201) = 201$




- Recursive equation ⇒ 4 equations, 4 unknowns

$$
\begin{aligned}
V^{\pi_3}(\text{d1}) &= 100 + V^{\pi_3}(\text{d2})\\
V^{\pi_3}(\text{d2}) &= 1 + .8(V^{\pi_3}(\text{d3})) + .2(V^{\pi_3}(\text{d5}))\\
V^{\pi_3}(\text{d3}) &= 100 + V^{\pi_3}(\text{d4})\\
V^{\pi_3}(\text{d5}) &= 100 + V^{\pi_3}(\text{d4})\\
V^{\pi_3}(\text{d4}) &= 0
\end{aligned}
$$

- So $V^{\pi_3}(\text{d1}) = 100 + 1 + .8(100) + .2(100) = 201$

# Example

- $\pi_7$ = {(d1, m14), (d2, m23), (d3, m34), (d5, m54)}
- Weighted sum of history costs:
  - $\sigma_5 = \langle$d1, d4$\rangle$

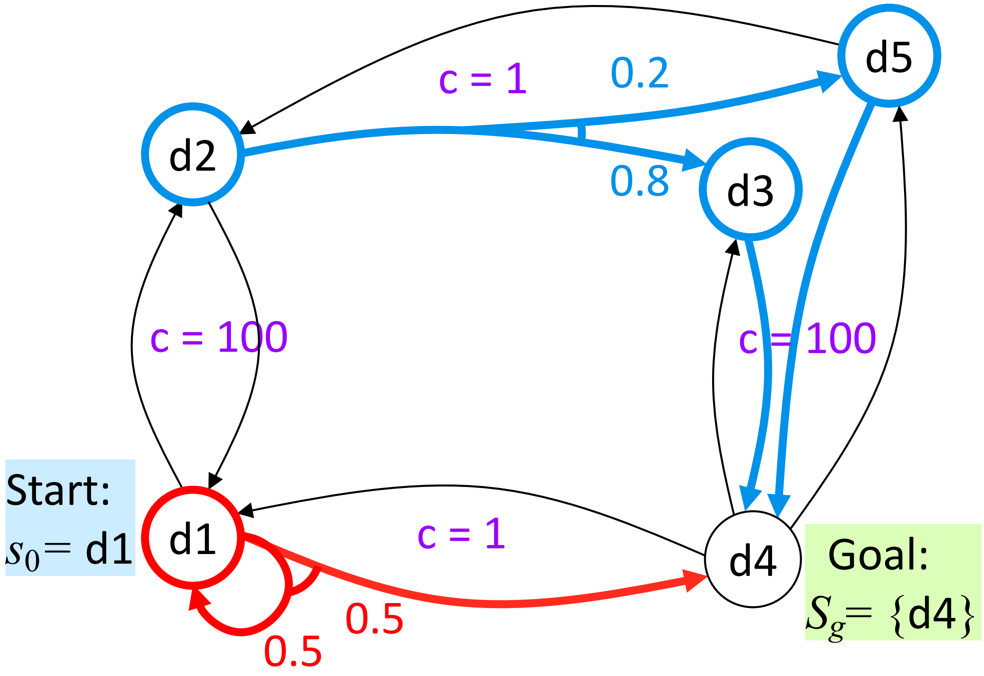
    $\Pr(\sigma_5 \mid \pi_7) = ½$,  cost $(\sigma_5) = 1$
  - $\sigma_6 = \langle$d1, d1, d4$\rangle$

    $\Pr(\sigma_6 \mid \pi_7) = (½)^2$,  cost $(\sigma_6) = 2$
  - $\sigma_7 = \langle$d1, d1, d1, d4$\rangle$

    $\Pr(\sigma_7 \mid \pi_7) = (½)^3$,  cost $(\sigma_7) = 3$

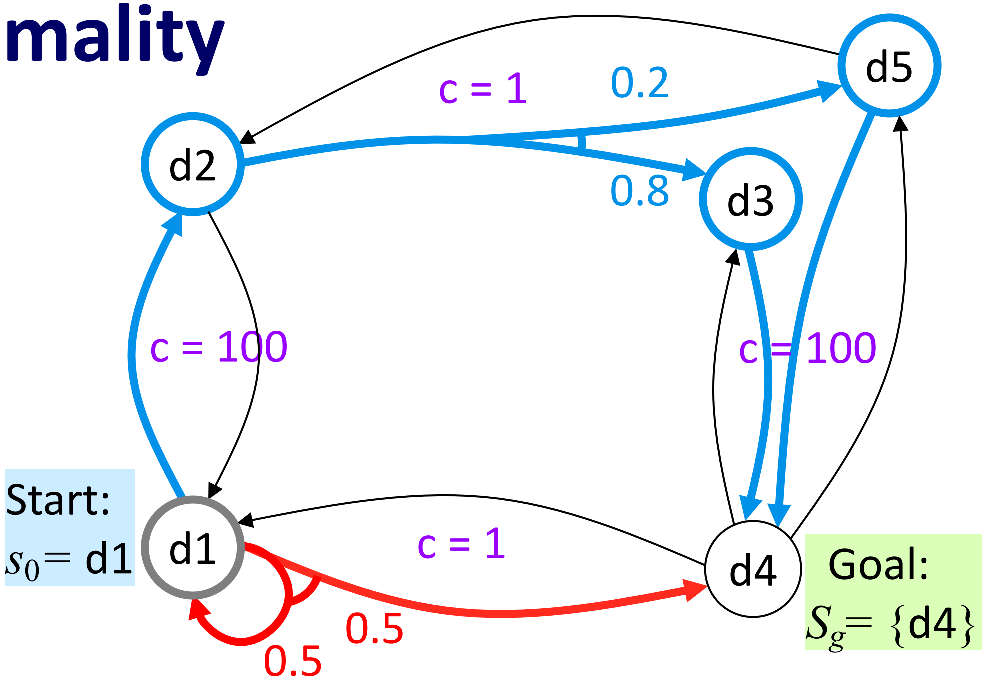
    …
- $V^{\pi_7}(\text{d1}) = (½)1 + (½)^2\, 2 + (½)^3\, 3 + \ldots = 2$
- Recursive equation:

$$V^{\pi_7}(\text{d1}) = 1 + ½(0) + ½(V^{\pi_7}(\text{d1}))$$
$$½V^{\pi_7}(\text{d1}) = 1$$
$$V^{\pi_7}(\text{d1}) = 2$$




- Given safe solution π,
  - Compute $V^\pi$ by solving $n$ linear equations, $n$ unknowns
  - $n$ = number of states reachable from $s_0$ using π $= |\hat{\gamma}(s_0,\pi)|$

# Dominance and Optimality




- Let π and $\pi'$ be safe solutions
  - π *dominates* π′ if $V^\pi(s) \le V^{\pi'}(s)$ at every state *s* where they're both defined
    - i.e., every state $s \in Domain(\pi) \cap Domain(\pi')$
- On the previous two slides
  - $\pi_3$ = {(d1, m12), (d2, m23), (d3, m34), (d5, m54)}
  - $\pi_7$ = {(d1, m14), (d2, m23), (d3, m34), (d5, m54)}
  - They differ only at d1
    - $V^{\pi_3}$(d1) = 201;   $V^{\pi_7}$(d1) = 2
  - $\pi_7$ dominates $\pi_3$
- Compare $\pi_3$ with $\pi_5$ = {(d1,m14)}
  - the only state in the domain of both policies is d1
    - $V^{\pi_3}$(d1) = 201;   $V^{\pi_5}$(d1) = 2
  - $\pi_5$ dominates $\pi_3$

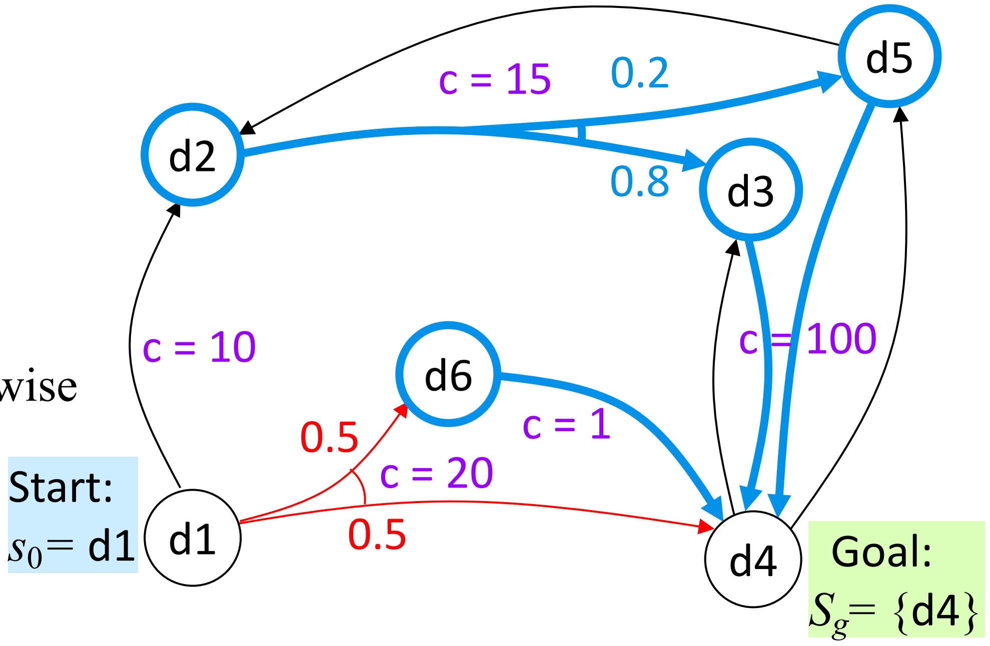
- π is *optimal* if π dominates *every* safe solution
- If π and π′ are both optimal, then $V^\pi(s) = V^{\pi'}(s)$ at every state where they're both defined
- Example: compare $\pi_5$ and $\pi_7$
  - the only state where both are defined is d1
  - $V^{\pi_5}$(d1) = $V^{\pi_7}$(d1) = 2

# Optimality

- Let $V^*(s)$ = expected cost of an optimal safe solution
- *Optimality principle* (Bellman's theorem):

$$V^*(s) = \begin{cases} 0, & \text{if } s \text{ is a goal} \\ \min_{a \in \text{Applicable}(s)} \sum_{s' \in \gamma(s,a)} \Pr(s' \mid s,a)[\text{cost}(s,a,s') + V^*(s')], & \text{otherwise} \end{cases}$$

- Example:
  - $V^*(\text{d4}) = 0$
  - $V^*(\text{d3}) = 100$
  - $V^*(\text{d5}) = \min\{100,\ 15 + V^*(\text{d2})\}$
  - $V^*(\text{d2}) = 0.8[15 + V^*(\text{d3})] + 0.2[15 + V^*(\text{d5})]$
    $= 15 + 0.8V^*(\text{d3}) + 0.2\ V^*(\text{d5}) = 95 + 0.2V^*(\text{d5})$
  - $V^*(\text{d6}) = 1$
  - $V^*(\text{d1}) = \min\{10 + V^*(\text{d2}),\ 0.5[20 + V^*(\text{d6})] + 0.5[20]\}$
    $= \min\{10 + V^*(\text{d2}),\ 20 + 0.5V^*(\text{d6})\}$
    $= \min\{10 + V^*(\text{d2}),\ 20.5\}$




**Poll.** What is $V^*(\text{d5})$?
A. 100  B. $15 + V^*(\text{d2})$  C. other  D. don't know

**Poll.** What is $V^*(\text{d1})$?
A. $10 + V^*(\text{d2})$  B. 20.5  C. other  D. don't know

# Summary

- Actions with probabilistic outcomes
- $\gamma(s,a)$ = a set of states, $\Pr(s' \mid s, a)$
- $\text{cost}(s,a,s') \in \mathbb{R}$
- Policies
  - Transitive closure
  - Reachability graph, leaves
- MDP problem: $P = (\Sigma, s_0, S_g)$, require $s_0 \notin S_g$
  - This is a *goal reachability* problem
- Solutions, closed solutions
- *History*: sequence of states
  $\sigma = \langle s_0, s_1, s_2, \ldots \rangle$ produced by Run-Policy
- $H(s,\pi)$ = {all possible histories from $s$ using $\pi$}

- Probability of reaching a goal state:

$$\Pr(S_g \mid s, \pi) = \sum_{\sigma \in H(s,\pi)} \{\Pr(\sigma \mid s,\pi) \mid \sigma \text{ ends in } S_g\}$$

or equivalently:

$$\Pr(S_g \mid s, \pi) = \begin{cases} 1, & \text{if } s \in S_g \\ \sum_{s' \in \gamma(s,\pi(s))} \Pr(S_g \mid s', \pi), & \text{otherwise} \end{cases}$$

- Unsafe and safe solutions
  - Acyclic and cyclic safe solutions
- Expected cost

$$V^\pi(s) = \sum_{\sigma \in H(s,\pi)} \Pr(\sigma \mid s, \pi)\, \text{cost}(\sigma)$$

or equivalently:

$$V^\pi(s) = 0, \text{ if } s \in S_g$$

$$= \sum_{s' \in \gamma(s,\pi(s))} \Pr(s' \mid s, \pi(s))[\text{cost}(s,\pi(s),s') + V^\pi(s')], \text{ otherwise}$$

- Planning as optimization