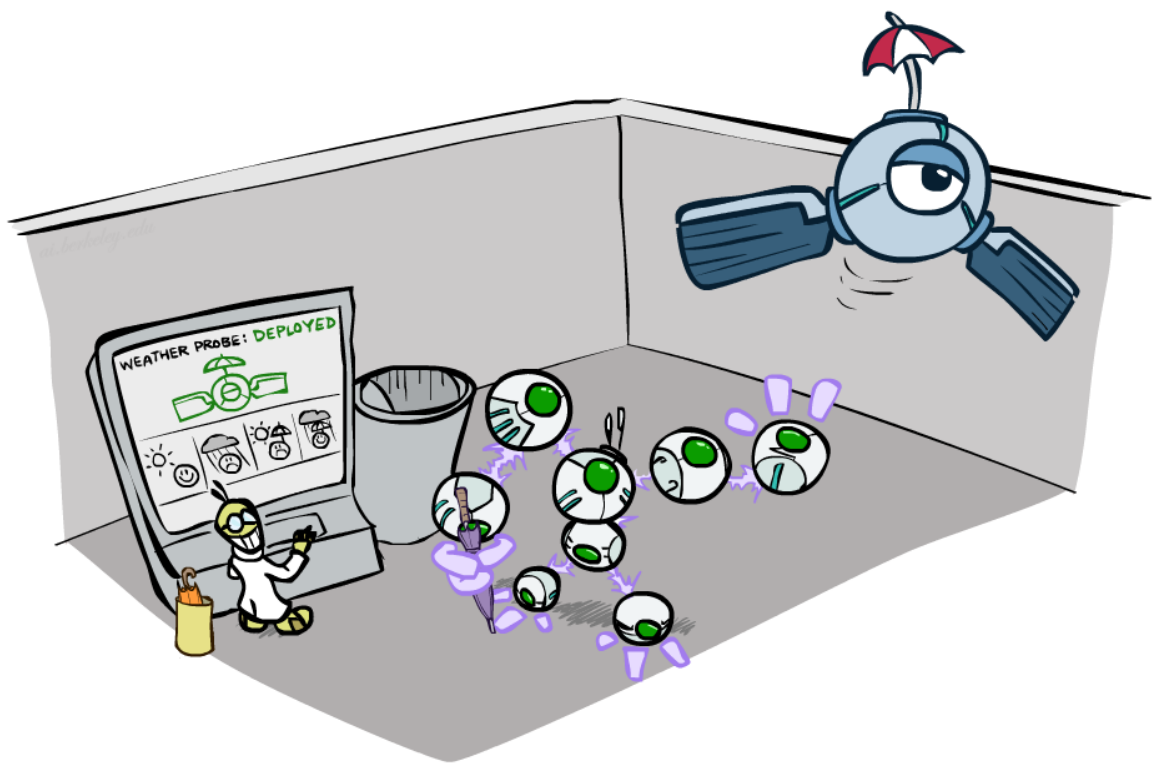
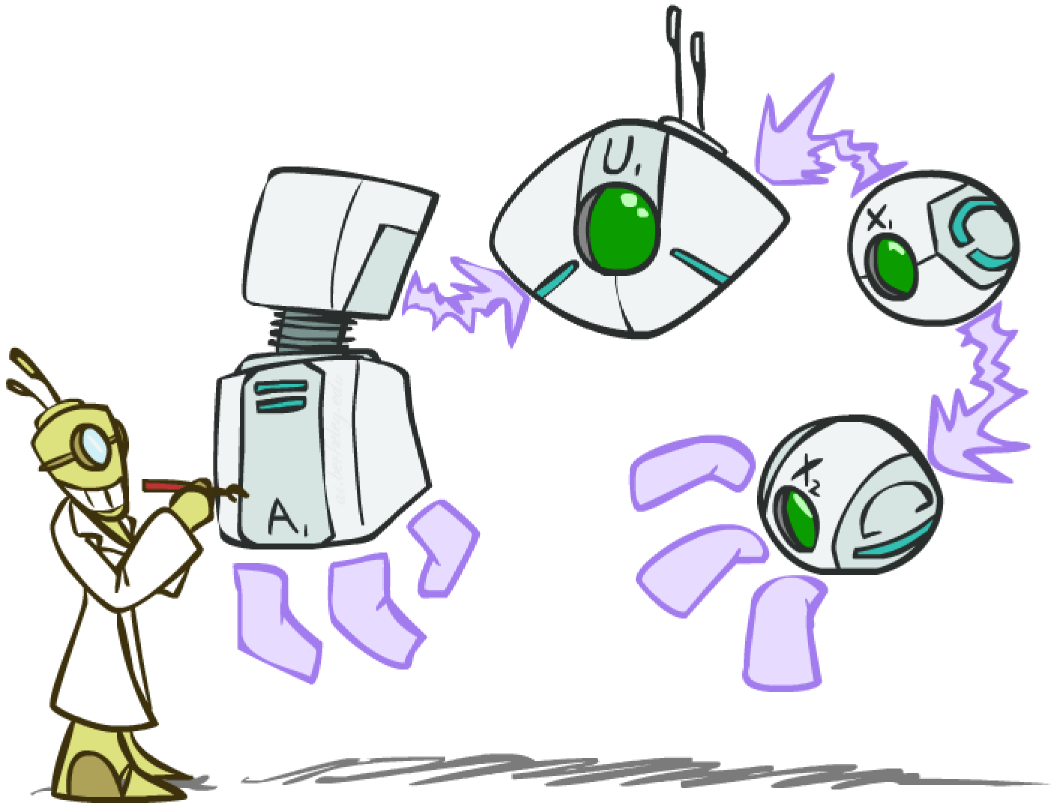
决策网络和完全知情的价值

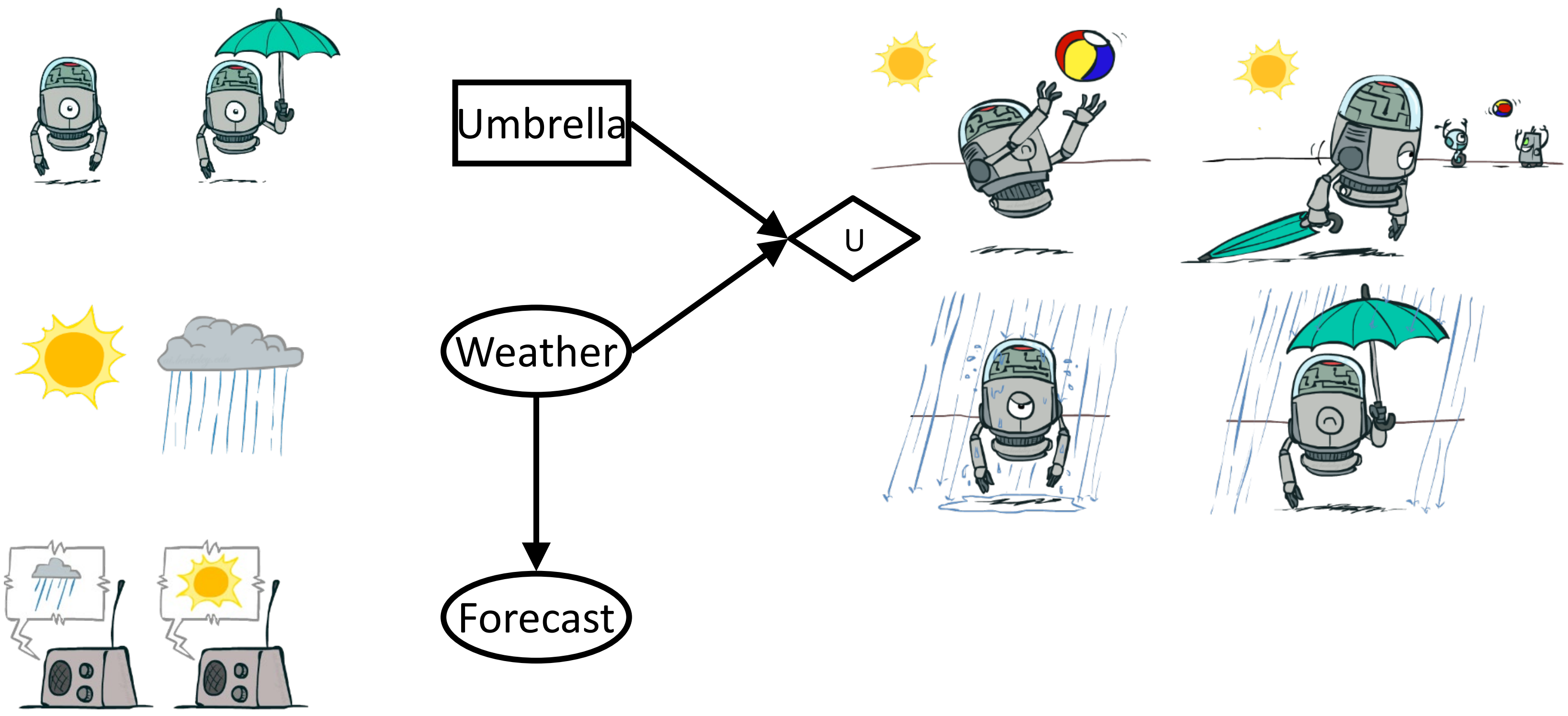
Decision Networks and Value of Perfect Information



决策网络（Decision Networks）

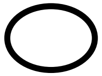
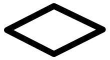


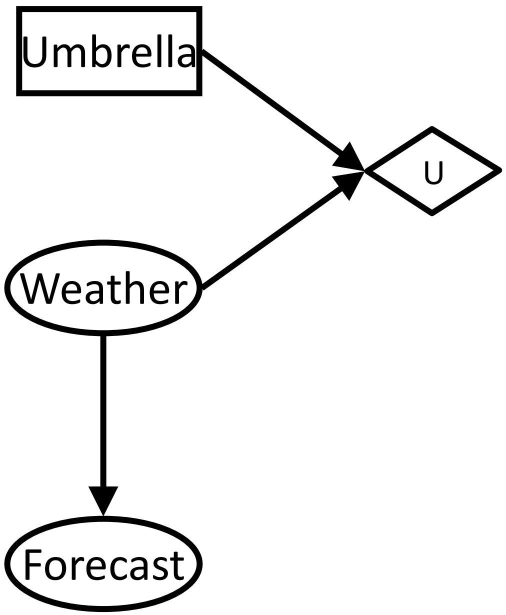
决策网络



决策网络

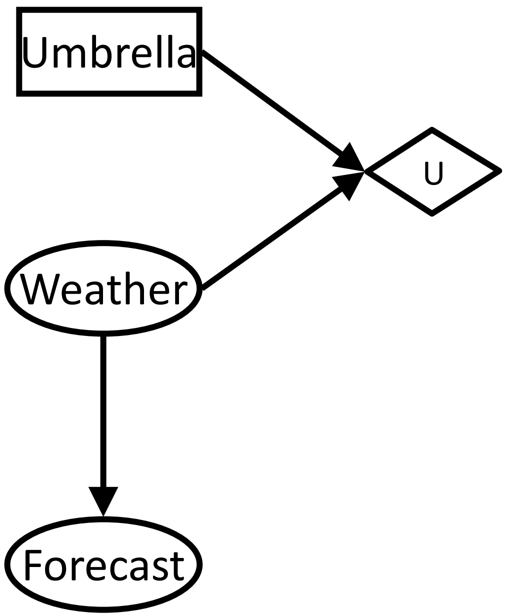
- MEU： 在给定观察值的情况下，选择能够最大化功效期望值的行动

- 直接的操作
  - 贝叶斯网络加上新的节点：功效和行动
  - 对每一个行动计算功效期望值
- 新节点的类型:
  -  机遇节点（就像随机网络中的随机变量节点）
  -  行动节点（矩形，不能有父节点,作为一种观察到的事实）
  -  功效节点(依赖于行动和机遇节点的取值)



决策网络

- 行动选择
  - 根据观察值实例化相应变量
  - 给行动节点赋值每一种可能的行动
  - 当在给定某种观察情况下，计算功效节点的所有父节点（随机变量）的后验概率
  - 计算每一个行动的期望功效值
  - 选择最大化期望功效值的行动



决策网络

Umbrella = leave

$$\mathrm{EU}(\mathrm{leave}) = \sum_{w} P(w)U(\mathrm{leave}, w)$$

$$= 0.7 \cdot 100 + 0.3 \cdot 0 = 70$$

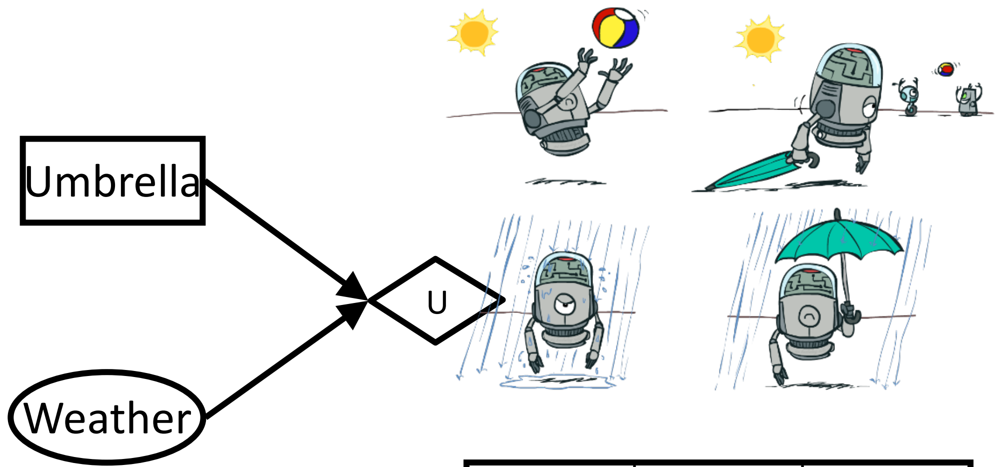
Umbrella = take

$$\mathrm{EU}(\mathrm{take}) = \sum_{w} P(w)U(\mathrm{take}, w)$$

$$= 0.7 \cdot 20 + 0.3 \cdot 70 = 35$$

最优决策= leave

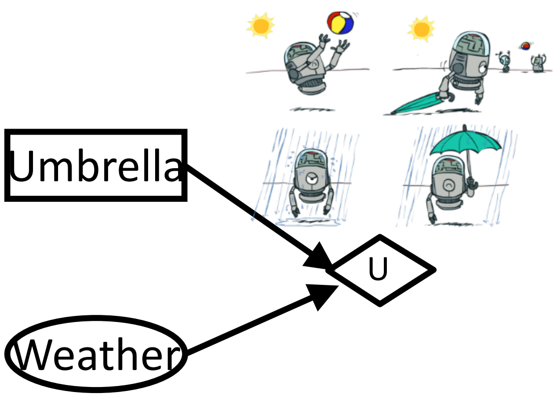
$$\mathrm{MEU}(\emptyset) = \max_{a} \mathrm{EU}(a) = 70$$

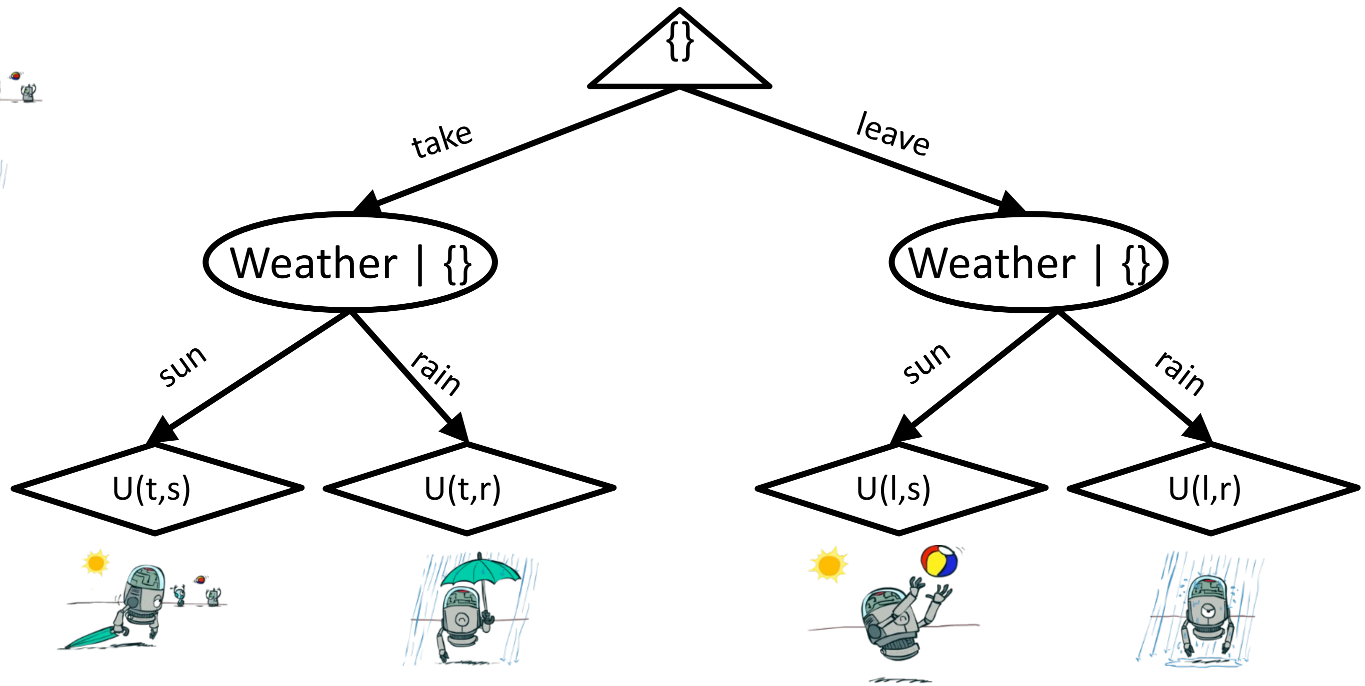


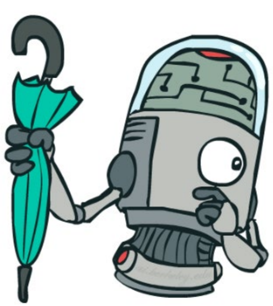
W	P(W)
sun	0.7
rain	0.3

A	W	U(A,W)
leave	sun	100
leave	rain	0
take	sun	20
take	rain	70

决策过程树







- 很像是 期望最大值(expectimax / MDPs)
- 区别在什么地方?

举例：加入Forecast变量后的决策树

Umbrella = leave

$$\mathrm{EU}(\mathrm{leave}|\mathrm{bad}) = \sum_w P(w|\mathrm{bad})U(\mathrm{leave}, w)$$

$$= 0.34 \cdot 100 + 0.66 \cdot 0 = 34$$

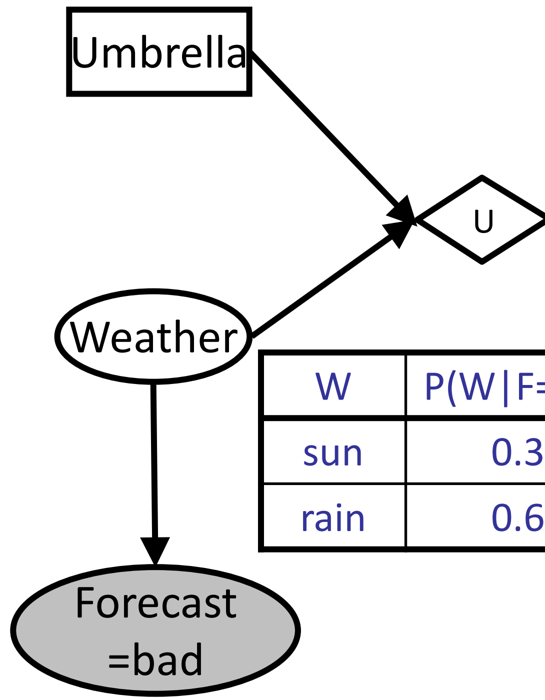
Umbrella = take

$$\mathrm{EU}(\mathrm{take}|\mathrm{bad}) = \sum_w P(w|\mathrm{bad})U(\mathrm{take}, w)$$

$$= 0.34 \cdot 20 + 0.66 \cdot 70 = 53$$

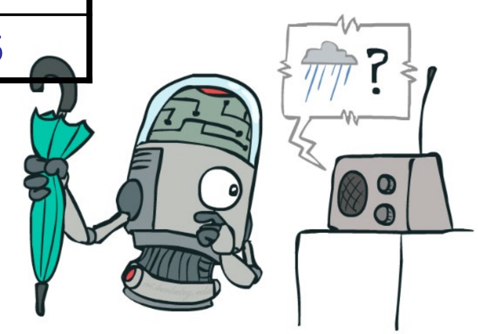
最优决策= take

$$\mathrm{MEU}(F = \mathrm{bad}) = \max_a \mathrm{EU}(a|\mathrm{bad}) = 53$$

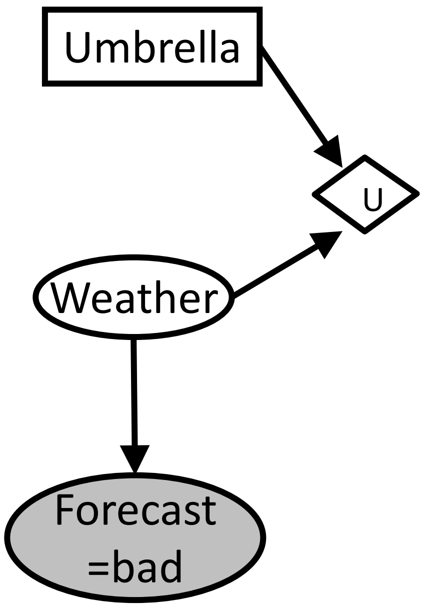


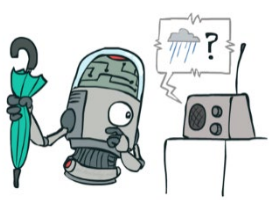
A	W	U(A,W)
leave	sun	100
leave	rain	0
take	sun	20
take	rain	70

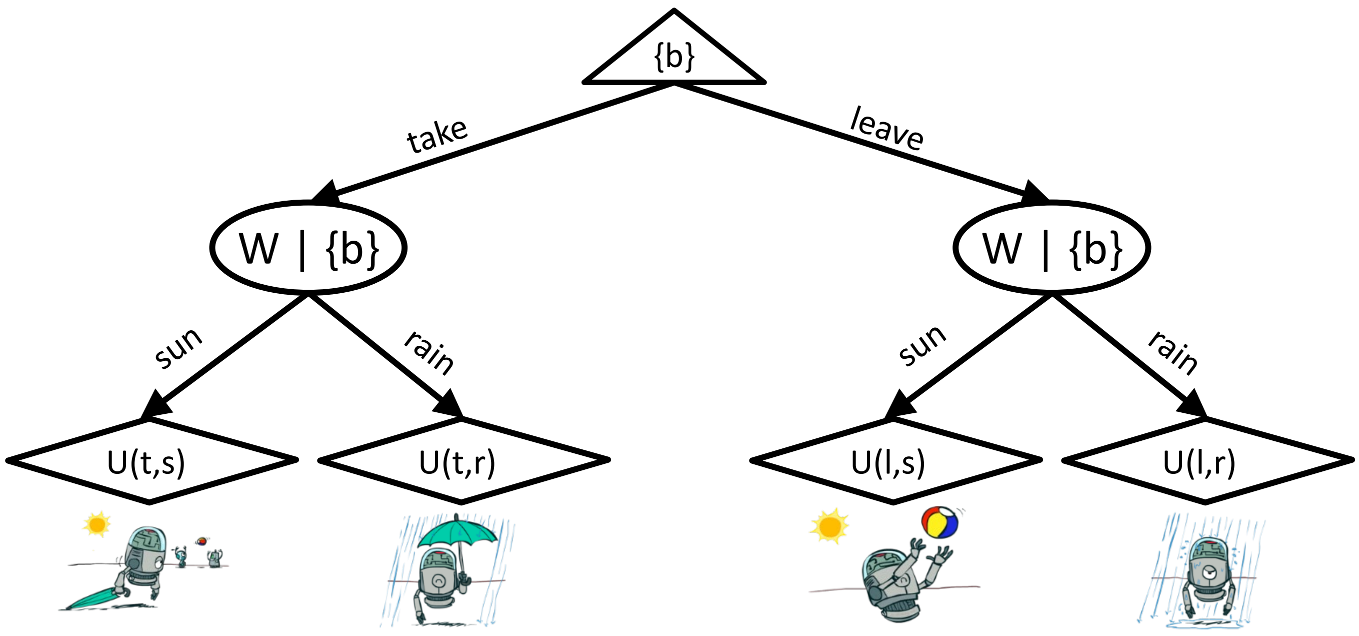
W	P(W|F=bad)
sun	0.34
rain	0.66



决策过程树

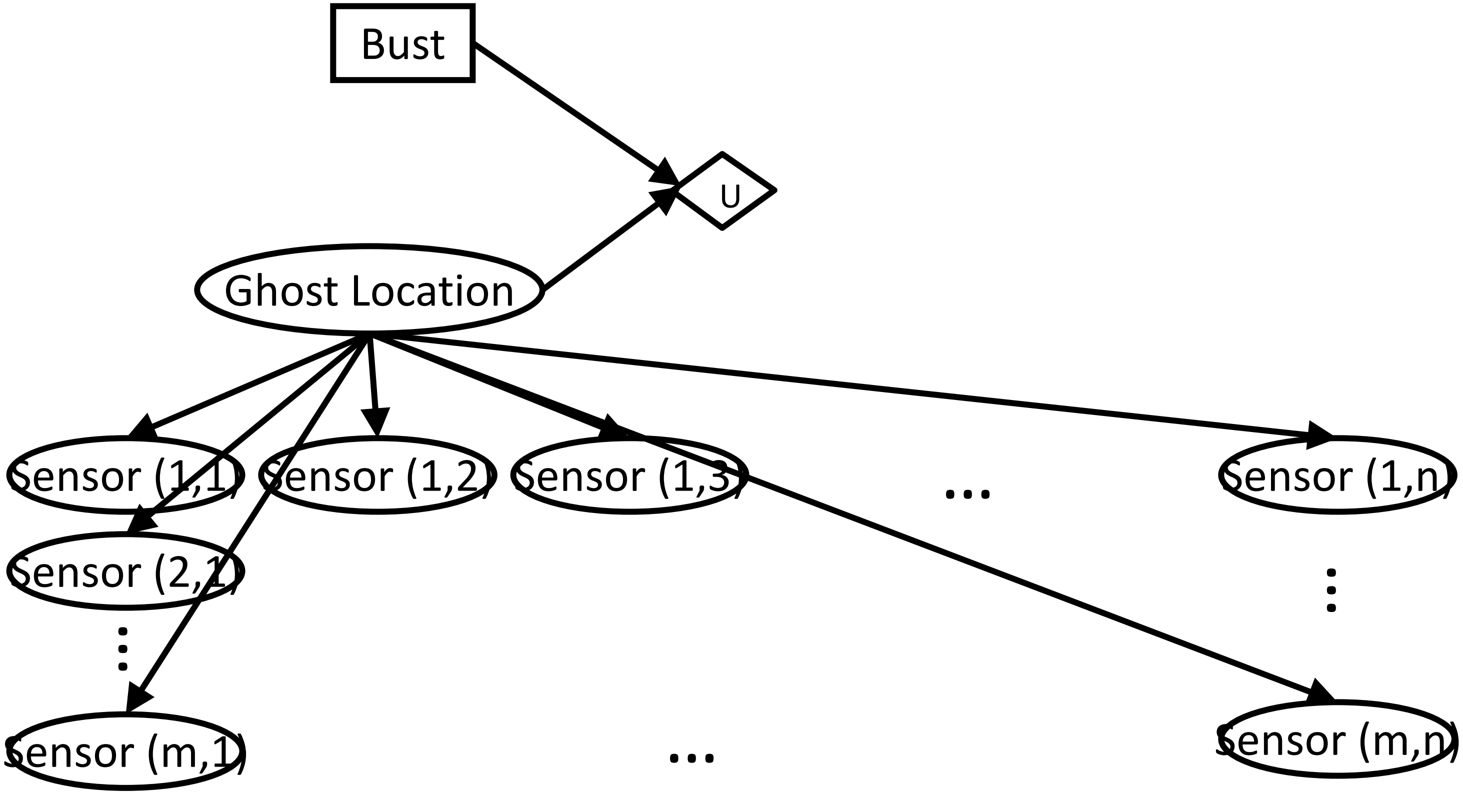

Umbrella
U
Weather
Forecast
=bad




{b}
take
leave
W | {b}
W | {b}
sun
rain
sun
rain
U(t,s)
U(t,r)
U(l,s)
U(l,r)

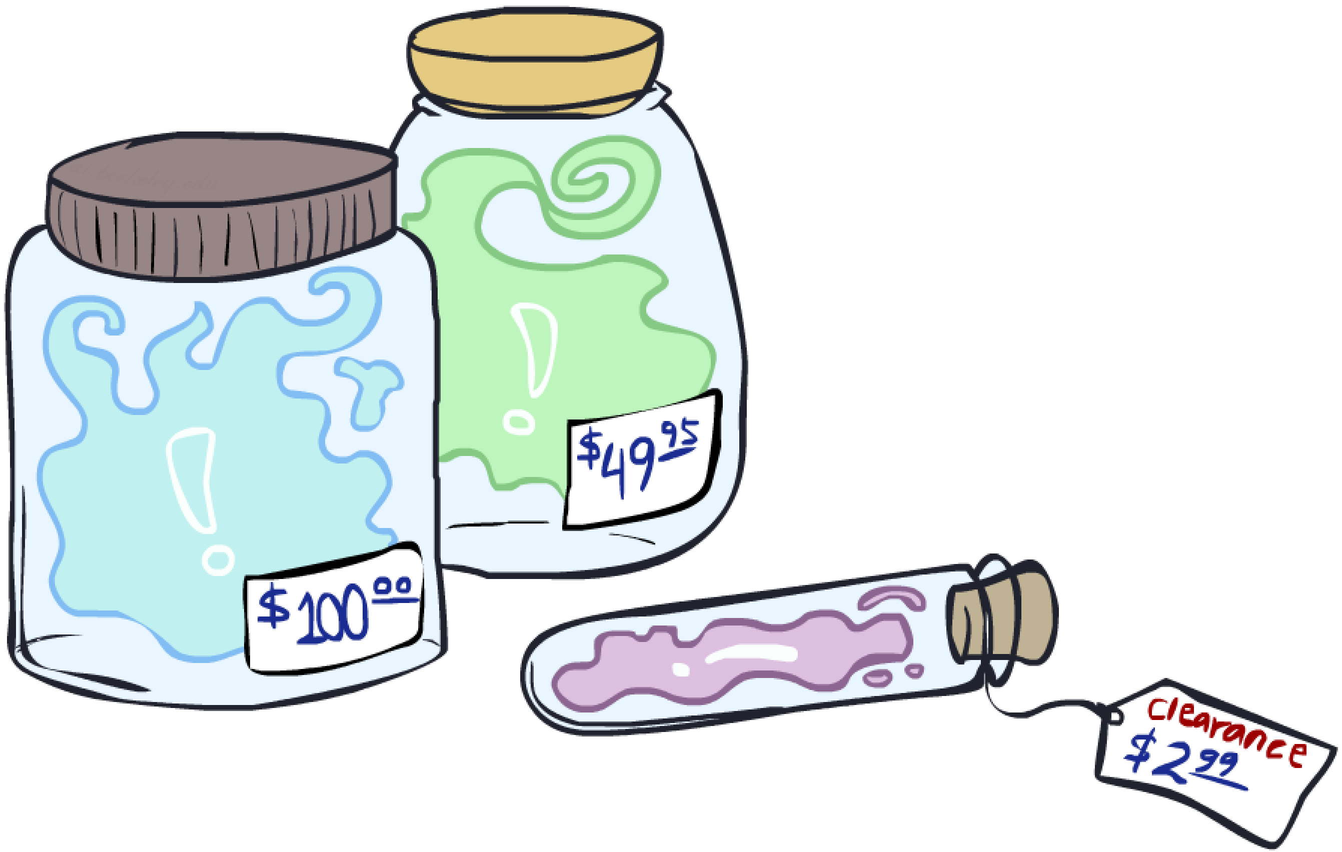
幽灵追捕者的决策

Demo: Ghostbusters with probability



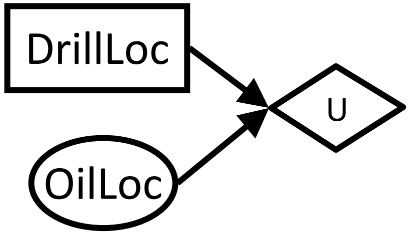
幽灵追捕者演示with Probability

信息的价值 (Value of Information)



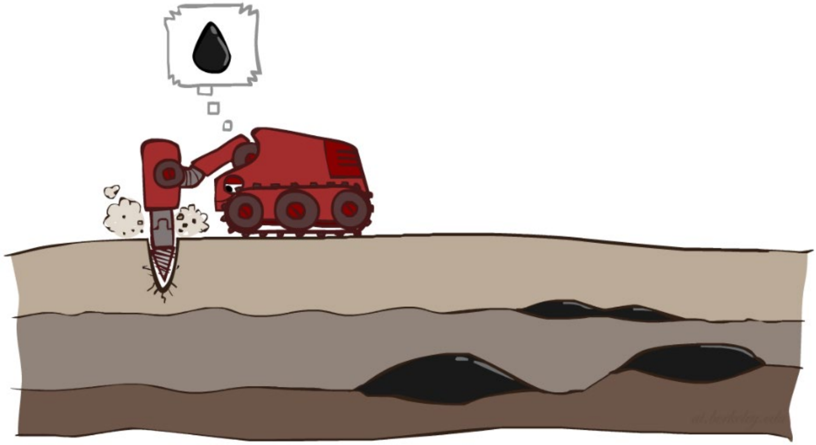
信息的价值

- 想法：计算获取某个观察变量的价值
  - 可以直接在决策网络上计算
- 举例：油井开采
  - 两个地方 A 和 B，只有一个地方有石油，价值为 k
  - 你只能在一个地方钻井
  - 先验概率为每个地方 0.5，并且互斥
  - 在任何一个地方钻井的 EU = k/2，MEU = k/2
- 问题：OilLoc 的 信息价值 是多少？
  - 即知道 A or B 哪一个地方有石油的信息的价值
  - 价值是在获取这个新信息后在 MEU 上的期望增值
  - 勘探的结果可能会说“oil in a” or “oil in b,” 的概率各为 0.5
  - 如果我们知道 OilLoc，MEU is k（无论是在a或b）
  - 那么在知道了OilLoc 这个信息后在 MEU上的增值是多少？
  - VPI(OilLoc) = k/2
  - 这条信息的合理价格：k/2



D	O	U
a	a	k
a	b	0
b	a	0
b	b	k

O	P
a	1/2
b	1/2



VPI ： Weather举例

MEU with no evidence

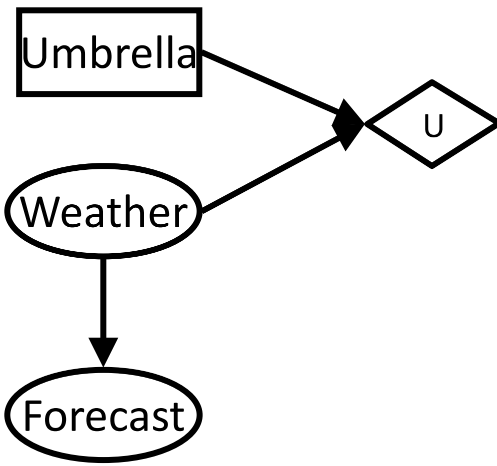
$$\mathrm{MEU}(\phi) = \max_a \mathrm{EU}(a) = 70$$

MEU if forecast is bad

$$\mathrm{MEU}(F = \mathrm{bad}) = \max_a \mathrm{EU}(a|\mathrm{bad}) = 53$$

MEU if forecast is good

$$\mathrm{MEU}(F = \mathrm{good}) = \max_a \mathrm{EU}(a|\mathrm{good}) = 95$$

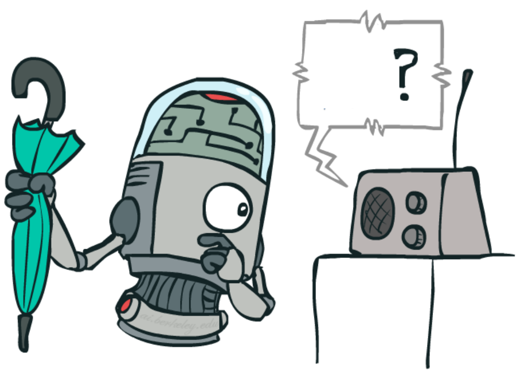


A	W	U
leave	sun	100
leave	rain	0
take	sun	20
take	rain	70

Forecast distribution

F	P(F)
good	0.59
bad	0.41

$$0.59 \cdot (95) + 0.41 \cdot (53) - 70$$
$$77.8 - 70 = 7.8$$

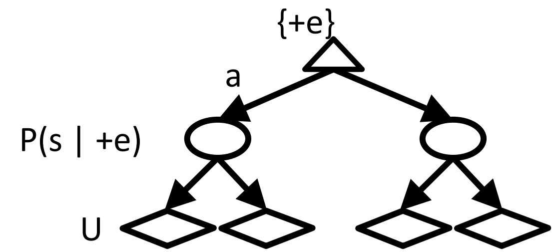


$$\mathrm{VPI}(E'|e) = \left(\sum_{e'} P(e'|e)\mathrm{MEU}(e, e') \right) - \mathrm{MEU}(e)$$

公式解释：信息的价值

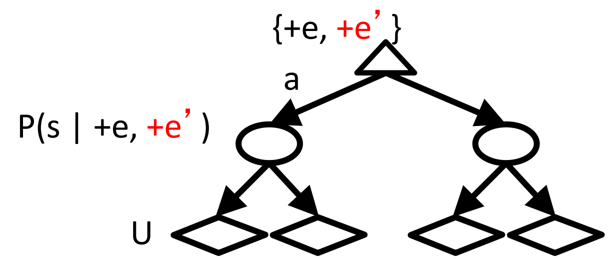
- 假设我们当前已知 E=e. 现在行动的最大期望功效值为:

$$\text{MEU}(e) = \max_a \sum_s P(s|e)\, U(s,a)$$

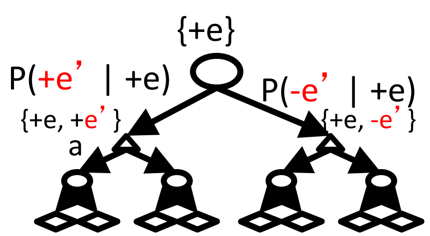


- 假设现在我们有获悉了 E’ = e’. 则现在行动的最大期望值为:

$$\text{MEU}(e,e') = \max_a \sum_s P(s|e,e')\, U(s,a)$$



- 但，E’ 是一个随机变量，它的值是未知的，所以我们不知道 e’ 将是何值，
- 期望值，如果 E’ 的值被揭示后，我们再行动的期望值: $\text{MEU}(e,E') = \sum_{e'} P(e'|e)\text{MEU}(e,e')$



- 信息的价值: 在 E’ 的值揭示出来后再行动比 现在就行动的 MEU 的增值 :

$$\text{VPI}(E'|e) = \text{MEU}(e,E') - \text{MEU}(e)$$

VPI 属性

- 非负性

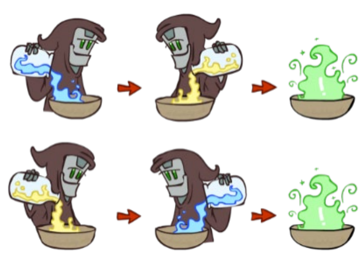
$$\forall E', e : \text{VPI}(E'|e) \geq 0$$

- 非加和性（不一定）

$$\text{VPI}(E_j, E_k|e) \neq \text{VPI}(E_j|e) + \text{VPI}(E_k|e)$$

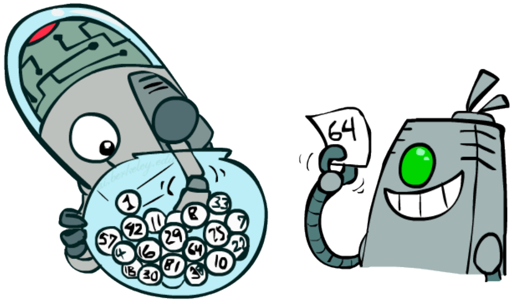
- 顺序-独立

$$\begin{aligned}\text{VPI}(E_j, E_k|e) &= \text{VPI}(E_j|e) + \text{VPI}(E_k|e, E_j) \\ &= \text{VPI}(E_k|e) + \text{VPI}(E_j|e, E_k)\end{aligned}$$

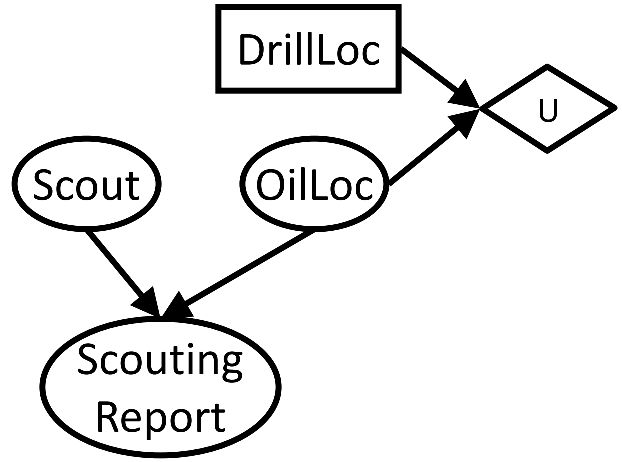


VPI 小问题

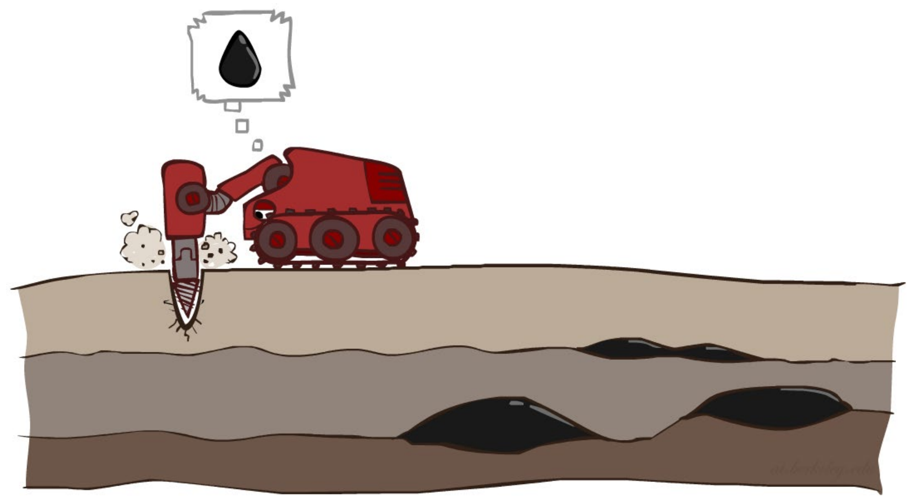
- 假设你在购买彩票． 奖金是 0 或 100元． 你可以购买从1到100之间的任何一个数，（中奖的几率是 1%）． 那么，知道中奖号码这条信息的价值是多少?



VPI 小问题



- VPI(OilLoc) ?
- VPI(ScoutingReport) ?
- VPI(Scout) ?
- VPI(Scout | ScoutingReport) ?



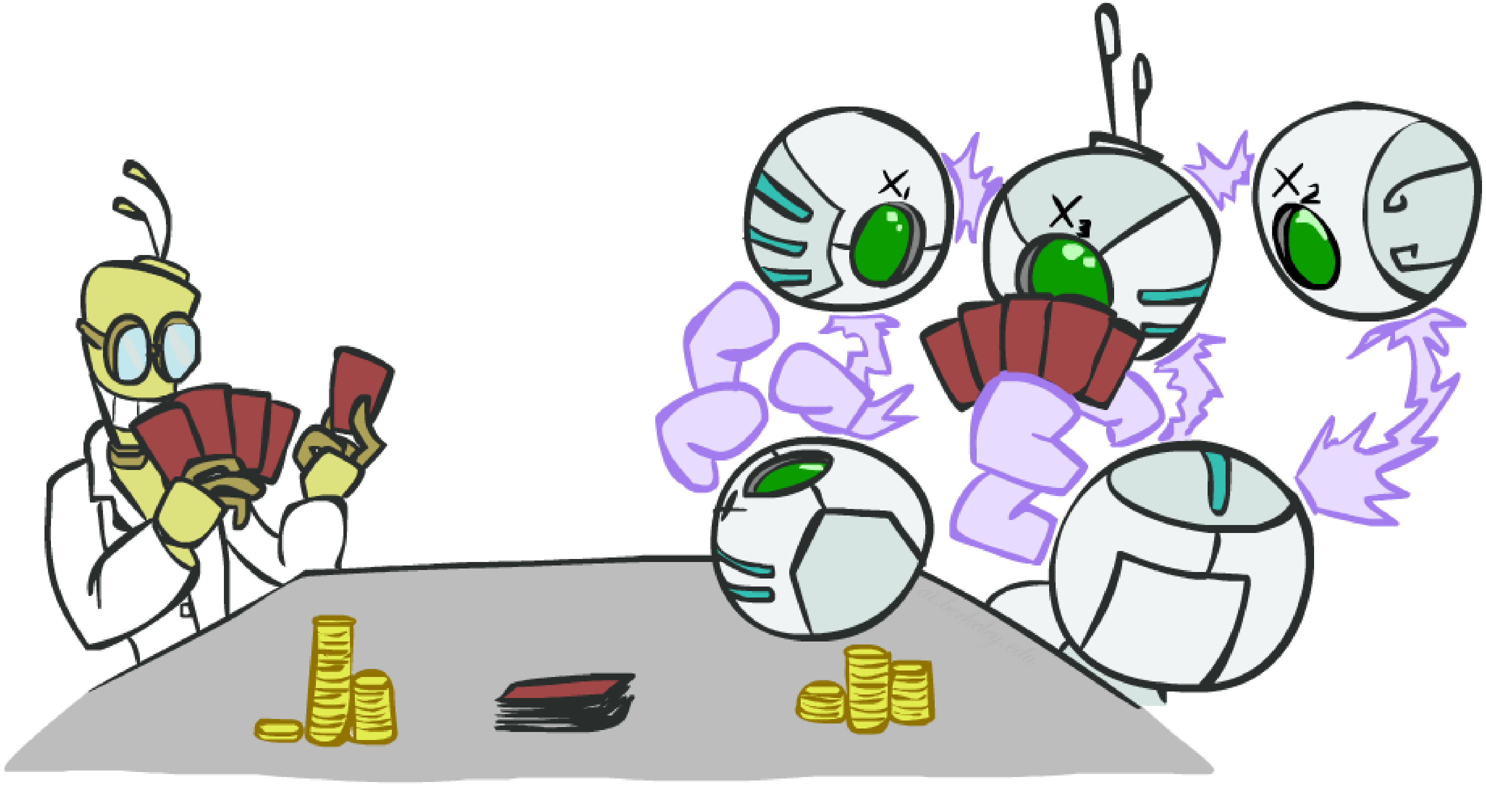
- 通常情况下：

  If Parents(U) $\perp\!\!\!\perp$ Z | CurrentEvidence

  Then VPI(Z | CurrentEvidence) = 0

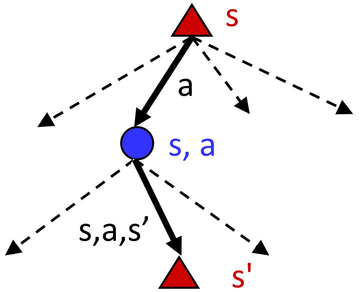
  (比如第3种情况里)

POMDPs

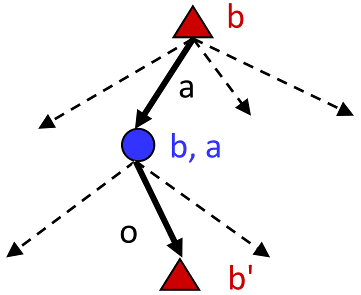


POMDPs

- MDPs have:
  - States S
  - Actions A
  - Transition function P(s'|s,a) (or T(s,a,s'))
  - Rewards R(s,a,s')



- POMDPs add:
  - Observations O
  - Observation function P(o|s) (or O(s,o))



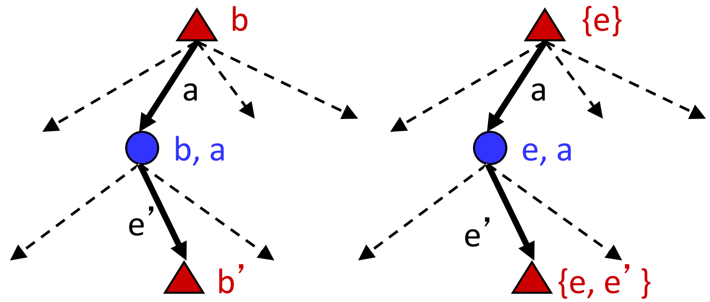
- POMDPs are MDPs over belief states b (distributions over S)

- We'll be able to say more in a few lectures

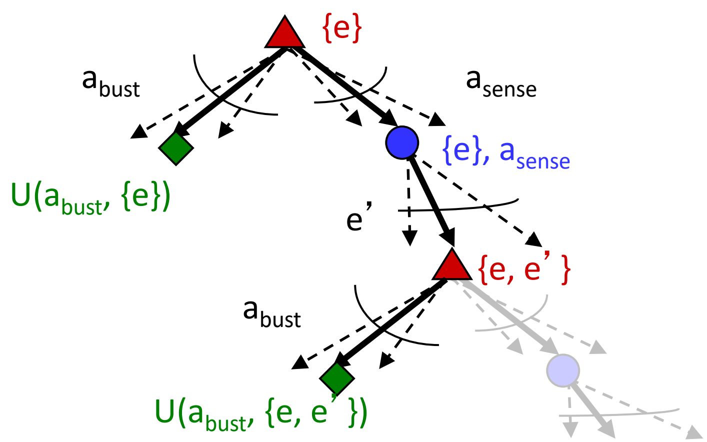
Example: Ghostbusters

Demo: Ghostbusters with VPI

- In (static) Ghostbusters:
  - Belief state determined by evidence to date {e}
  - Tree really over evidence sets
  - Probabilistic reasoning needed to predict new evidence given past evidence



- Solving POMDPs
  - One way: use truncated expectimax to compute approximate value of actions
  - What if you only considered busting or one sense followed by a bust?
  - You get a VPI-based agent!



Video of Demo Ghostbusters with VPI