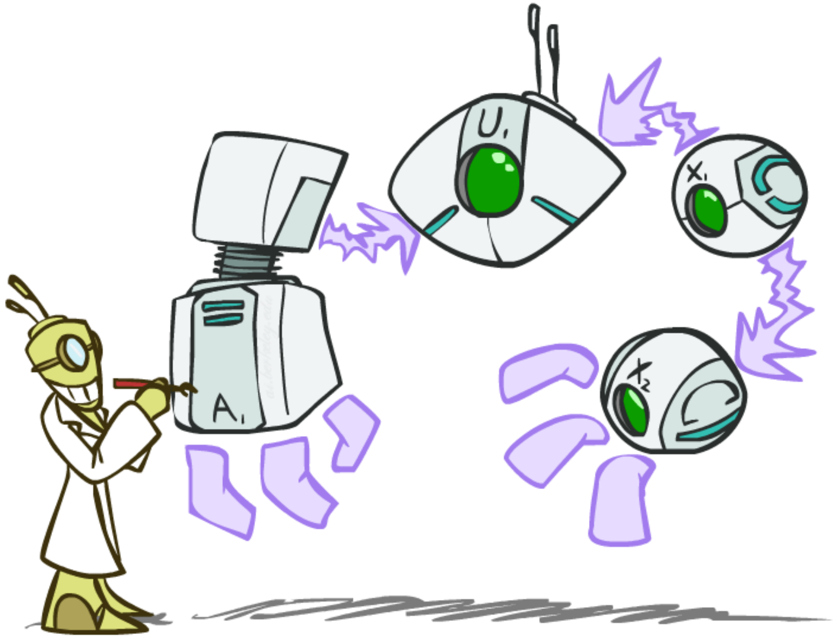
Artificial Intelligence - INFOF311

Decision networks and VPI

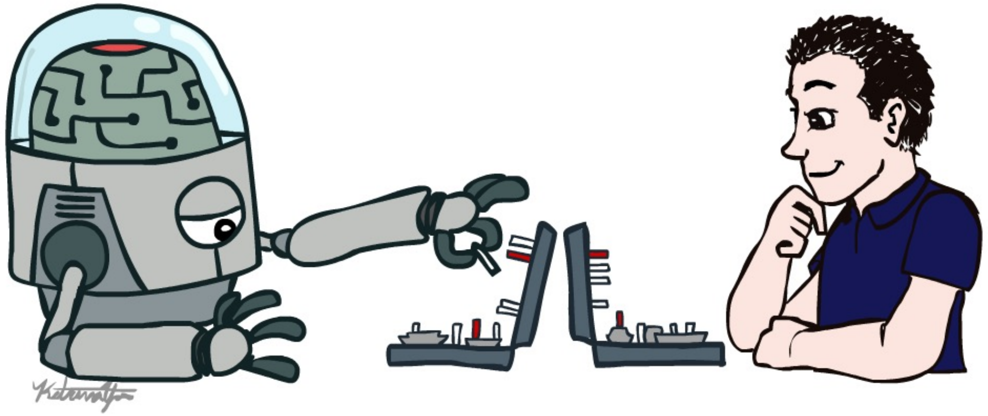


Instructor : Tom Lenaerts

Acknowledgement

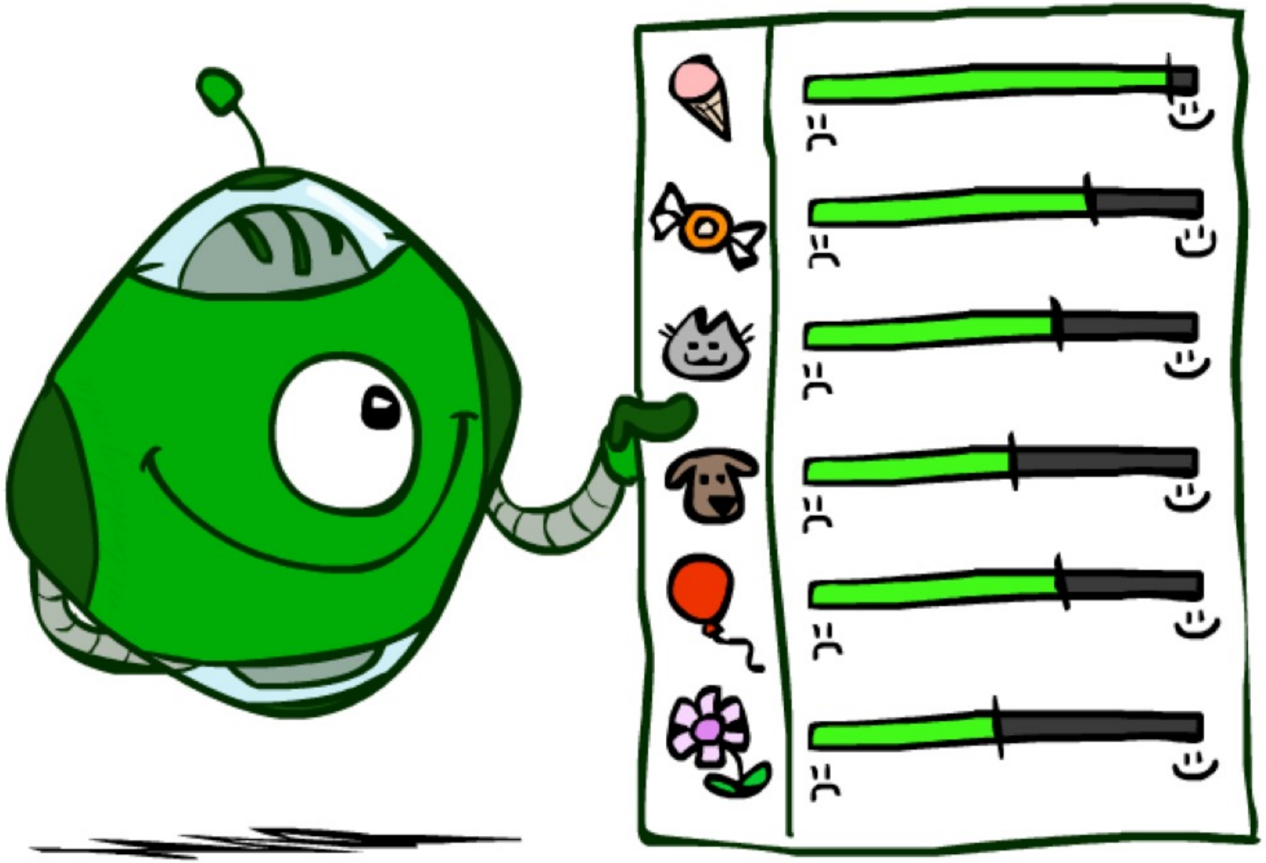
We thank Stuart Russell for his generosity in allowing us to use the slide set of the UC Berkeley Course CS188, Introduction to Artificial Intelligence. These slides were created by Dan Klein, Pieter Abbeel and Anca Dragan for CS188 Intro to AI at UC Berkeley. All CS188 materials are available at http://ai.berkeley.edu.

CHAI Center for Human-Compatible Artificial Intelligence



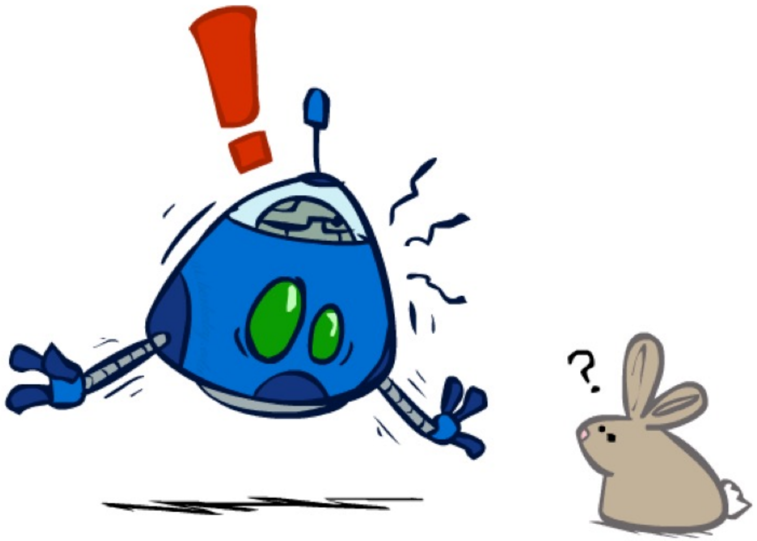
The slides for INFOF311 are slightly modified versions of the slides of the spring and summer CS188 sessions in 2021 and 2022

Utilities



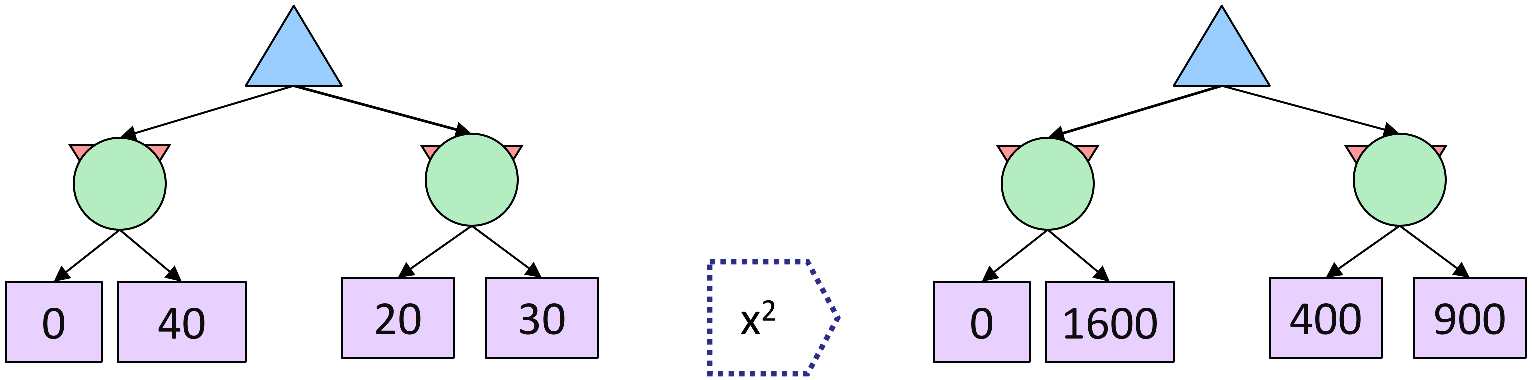
Maximum Expected Utility

- Principle of maximum expected utility:
  - A **rational agent** should choose the action that maximizes its expected utility, given its knowledge
  - $action = \underset{a}{\operatorname{argmax}} EU(a)$
    - With $EU(a) = \sum_{s'} P(RESULT(a) = s')\boldsymbol{U(s')}$



- Questions:
  - Where does $U(s')$ come from?
  - How do we know such utilities even exist?
  - How do we know that averaging (EU) even makes sense?
  - What if our behavior (preferences) can't be described by utilities?

The need for numbers



- For worst-case minimax reasoning, terminal value scale doesn't matter
  - We just want better states to have higher evaluations (get the ordering right)
  - The optimal decision is invariant under any ***monotonic transformation***
- For average-case expectimax reasoning, we need ***magnitudes*** to be meaningful
  - Only positive linear transformation preserve optimal policy.

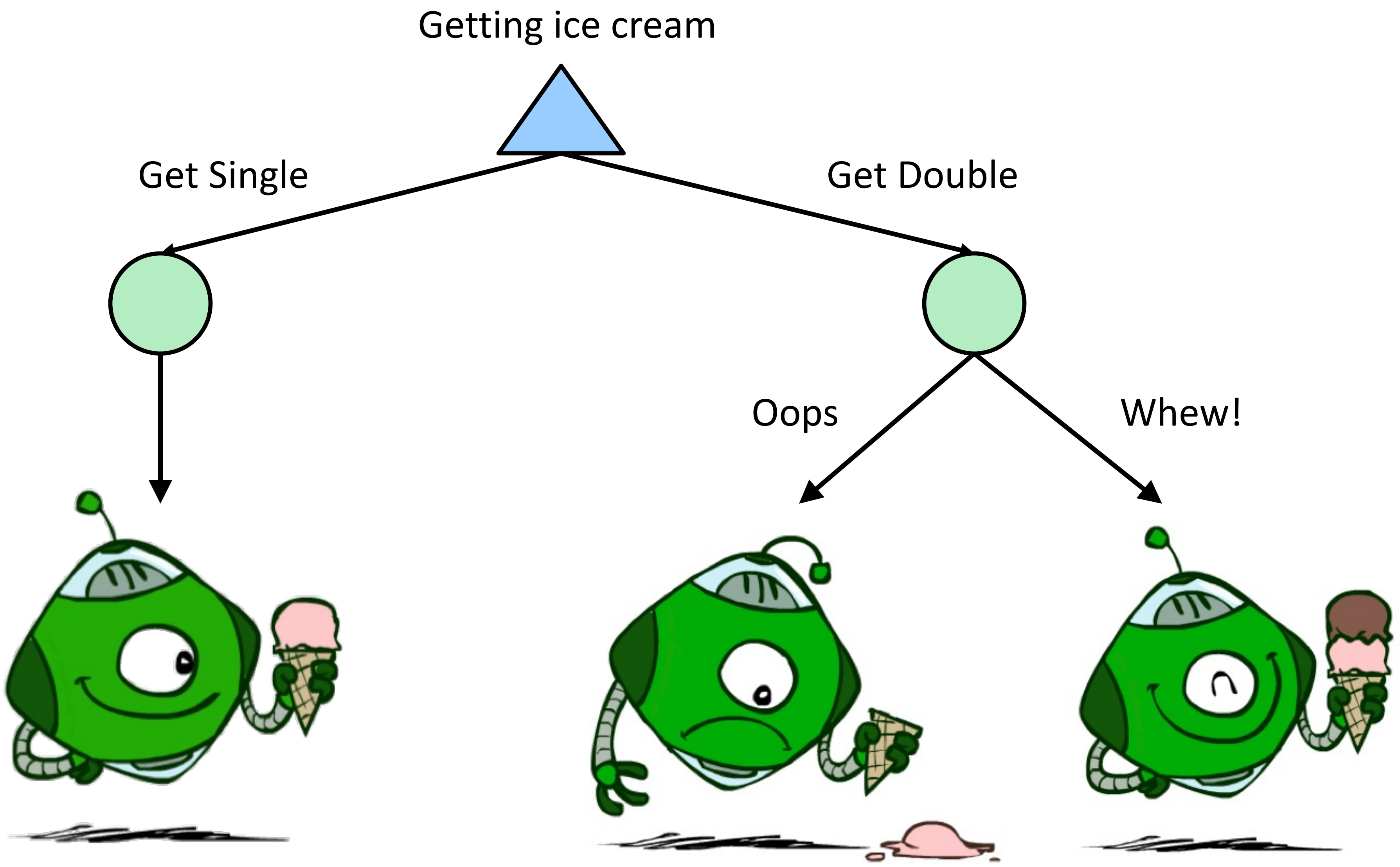
Utilities

- Utilities are functions from outcomes (states of the world) to real numbers that describe an agent's preferences

- Where do utilities come from?
  - In a game, may be simple (+1/-1)
  - Utilities summarize the agent's goals
  - Theorem: any "rational" preferences can be summarized as a utility function

- We hard-wire utilities and let behaviors emerge
  - Why don't we let agents pick utilities?
  - Why don't we prescribe behaviors?

Utilities: Uncertain Outcomes



Preferences

- An agent must have preferences among:
  - Prizes: P_1, P_2, etc.
  - Lotteries: situations with uncertain prizes

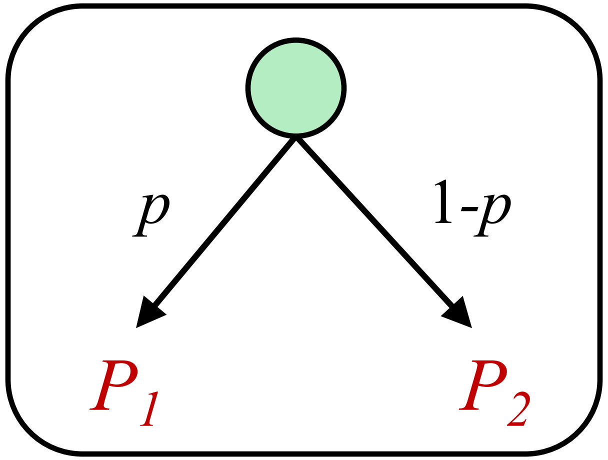
    $L = [p, P_1;\ (1-p), P_2]$

- Notation:
  - Preference: $A \succ B$
  - Indifference: $A \sim B$

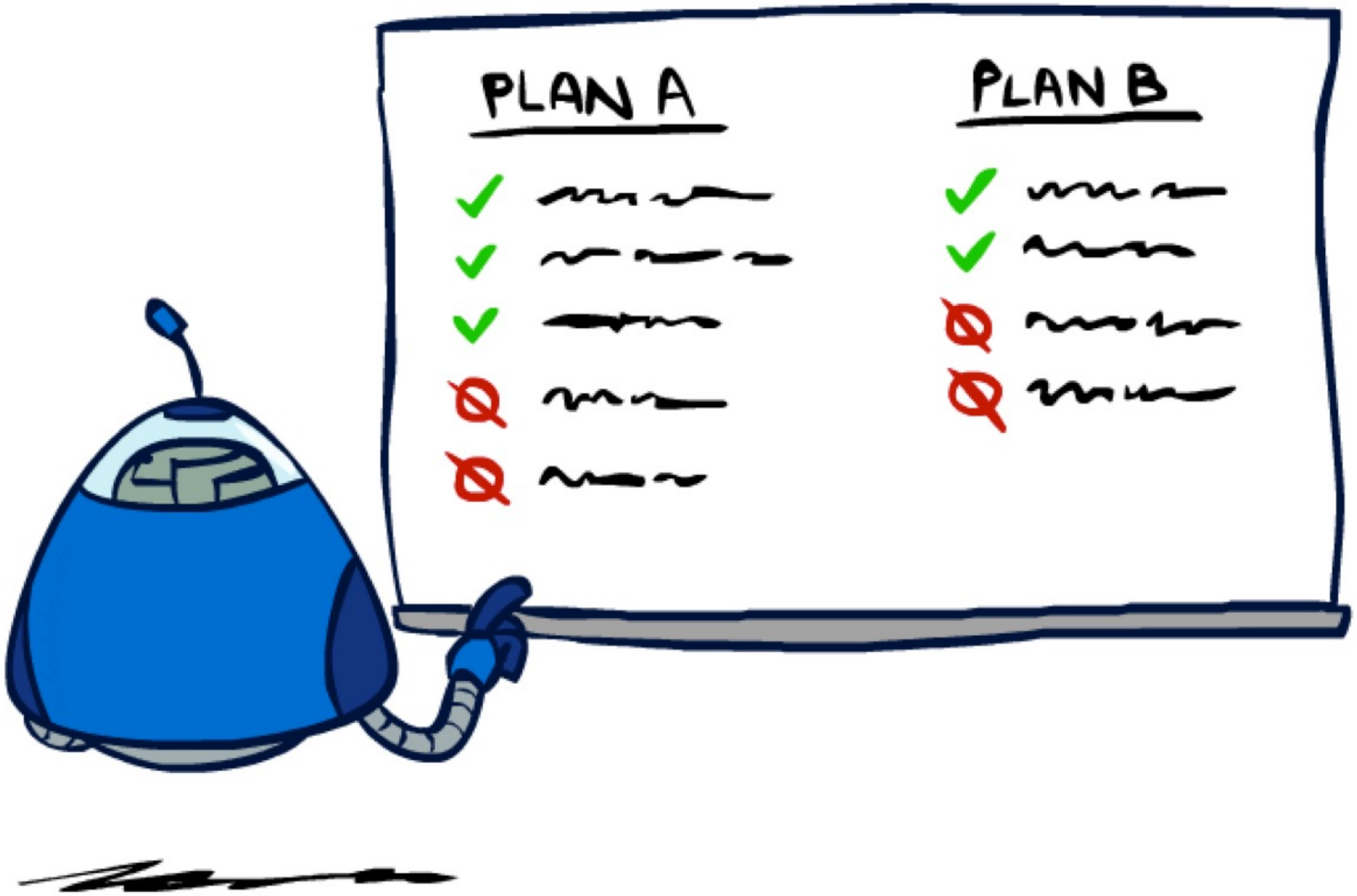
A Prize

P_1

A Lottery



Rationality

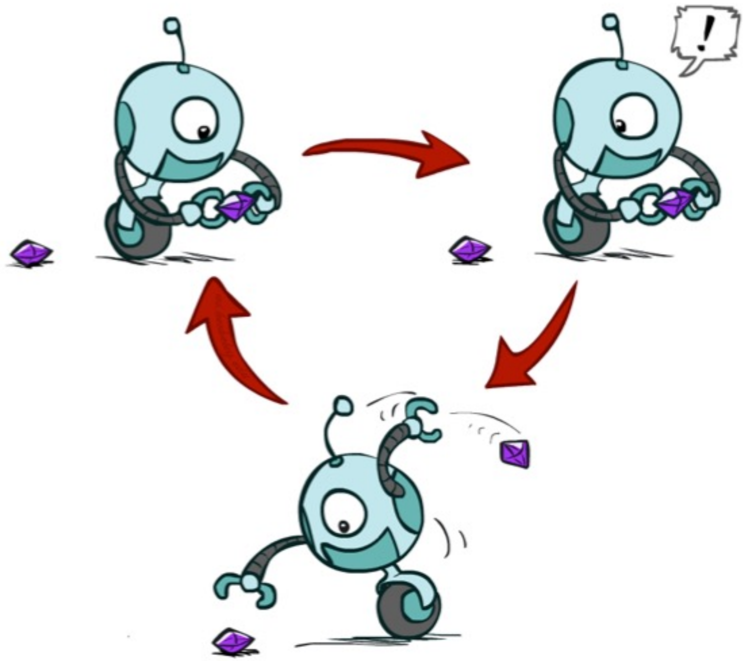


Rational Preferences

- We want some constraints on preferences before we call them rational, such as:

Axiom of Transitivity: $(A \succ B) \wedge (B \succ C) \Rightarrow (A \succ C)$

- For example: an agent with intransitive preferences can be induced to give away all of its money
  - If B > C, then an agent with C would pay (say) 1 cent to get B
  - If A > B, then an agent with B would pay (say) 1 cent to get A
  - If C > A, then an agent with A would pay (say) 1 cent to get C



Rational Preferences

The Axioms of Rationality

Orderability:

$$(A \succ B) \vee (B \succ A) \vee (A \sim B)$$

Transitivity:

$$(A \succ B) \wedge (B \succ C) \Rightarrow (A \succ C)$$

Continuity:

$$(A \succ B \succ C) \Rightarrow \exists p\ [p, A;\ 1\text{-}p, C] \sim B$$

Substitutability:

$$(A \sim B) \Rightarrow [p, A;\ 1\text{-}p, C] \sim [p, B;\ 1\text{-}p, C]$$

Monotonicity:

$$(A \succ B) \Rightarrow$$
$$(p \geq q) \Leftrightarrow [p, A;\ 1\text{-}p, B] \succ [q, A;\ 1\text{-}q, B]$$

Theorem: Rational preferences imply behavior describable as maximization of expected utility

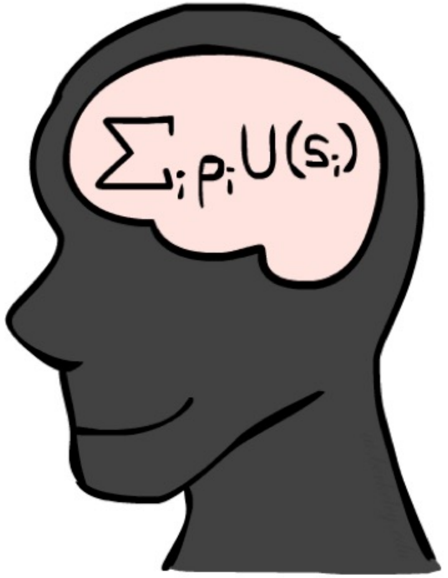
MEU Principle

- Theorem [Ramsey, 1931; von Neumann & Morgenstern, 1944]
  - Given any preferences satisfying these constraints, there exists a real-valued function U such that:

$$U(A) > U(B) \Leftrightarrow A \succ B; \; U(A) = U(B) \Leftrightarrow A \sim B$$

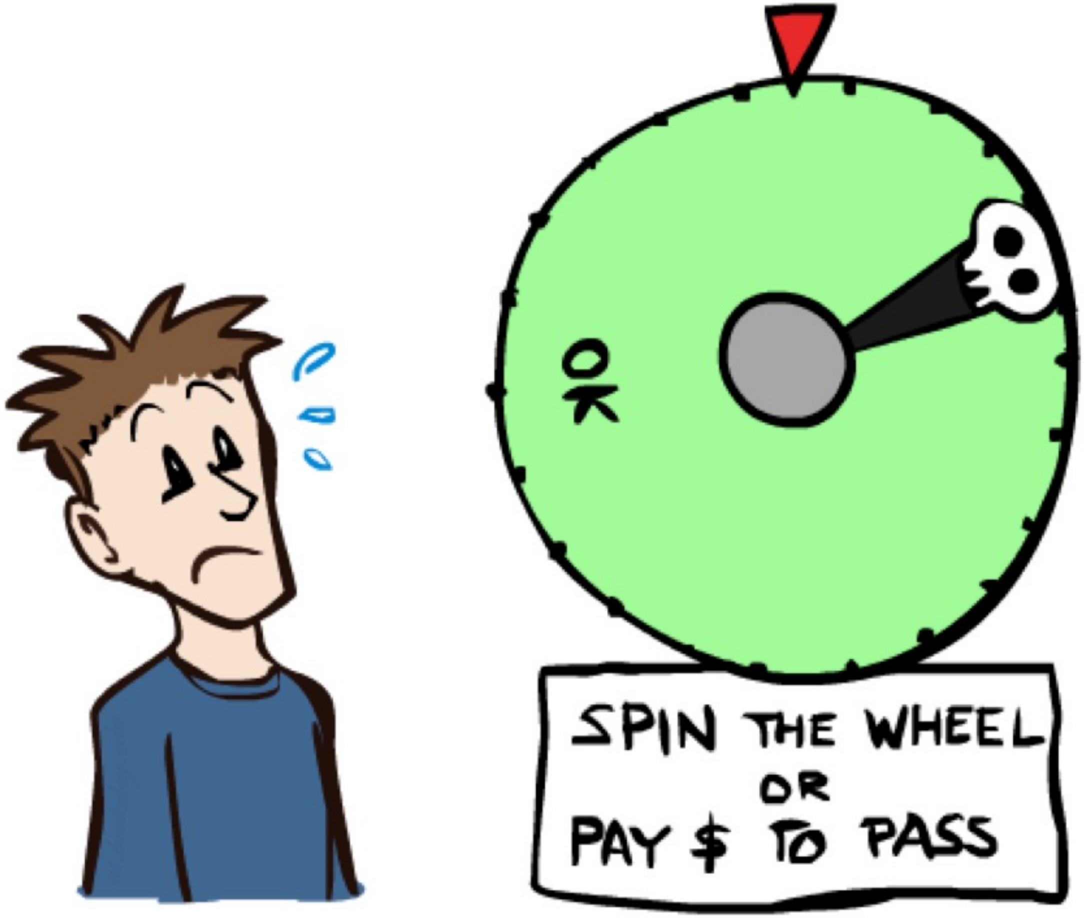
$$U([p_1,S_1; \ldots ; p_n,S_n]) = p_1 U(S_1) + \ldots + p_n U(S_n)$$

  - I.e. values assigned by U preserve preferences of both prizes and lotteries!
  - Optimal policy invariant under ***positive affine transformation*** $U' = aU+b$, $a>0$



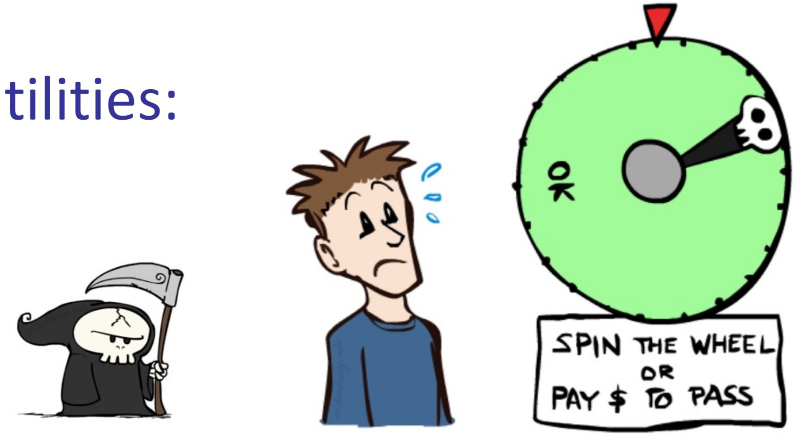
- Maximum expected utility (MEU) principle:
  - Choose the action that maximizes expected utility
  - Note: rationality does ***not*** require representing or manipulating utilities and probabilities
    - E.g., a lookup table for perfect tic-tac-toe

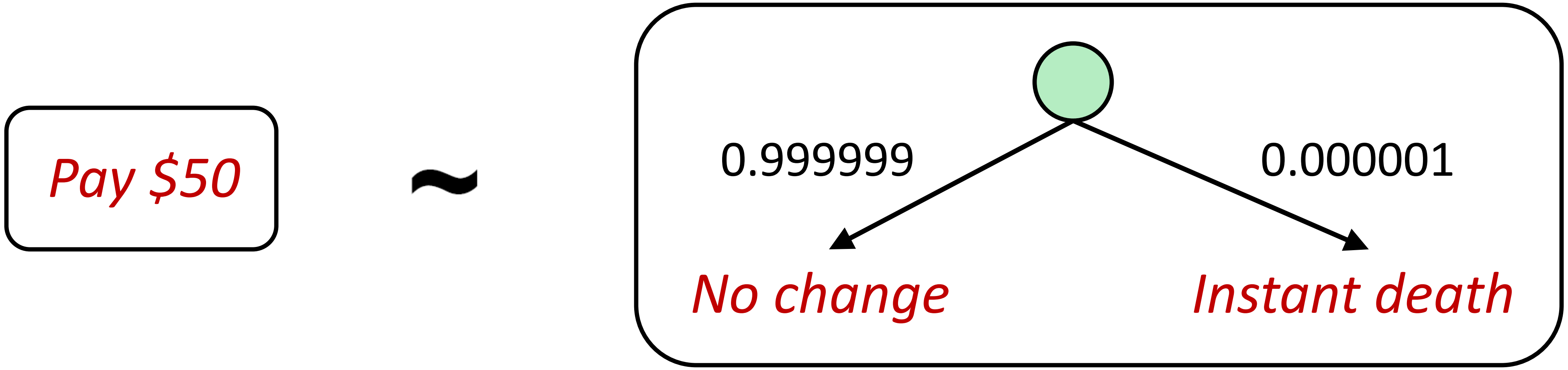
Human Utilities



Human Utilities

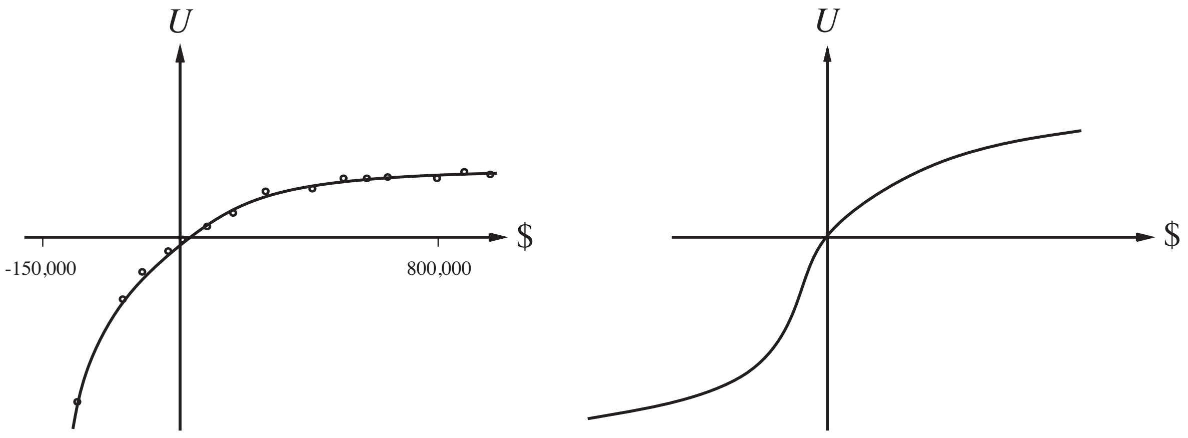
- Utilities map states to real numbers. Which numbers?
- Standard approach to assessment (elicitation) of human utilities:
  - Compare a prize A to a ***standard lottery*** L_p between
    - "best possible prize" $u_\top$ with probability p
    - "worst possible catastrophe" $u_\bot$ with probability $1\text{-}p$
  - Adjust lottery probability p until indifference: $A \sim L_p$
  - Resulting p is a utility in $[0,1]$





Money

- Money ***does not*** behave as a utility function, but we can talk about the utility of having money (or being in debt)
- Given a lottery $L = [p, \$X; (1-p), \$Y]$
  - The ***expected monetary value*** $\text{EMV}(L) = pX + (1-p)Y$
  - The utility is $U(L) = pU(\$X) + (1-p)U(\$Y)$
  - Typically, $U(L) < U(\text{ EMV}(L))$
  - In this sense, people are ***risk-averse***
  - E.g., how much would you pay for a lottery ticket L=[0.5, \$10,000; 0.5, \$0]?
  - The ***certainty equivalent*** of a lottery $\text{CE}(L)$ is the cash amount such that $\text{CE}(L) \sim L$
  - The ***insurance premium*** is $\text{EMV}(L) - \text{CE}(L)$
  - If people were risk-neutral, this would be zero!

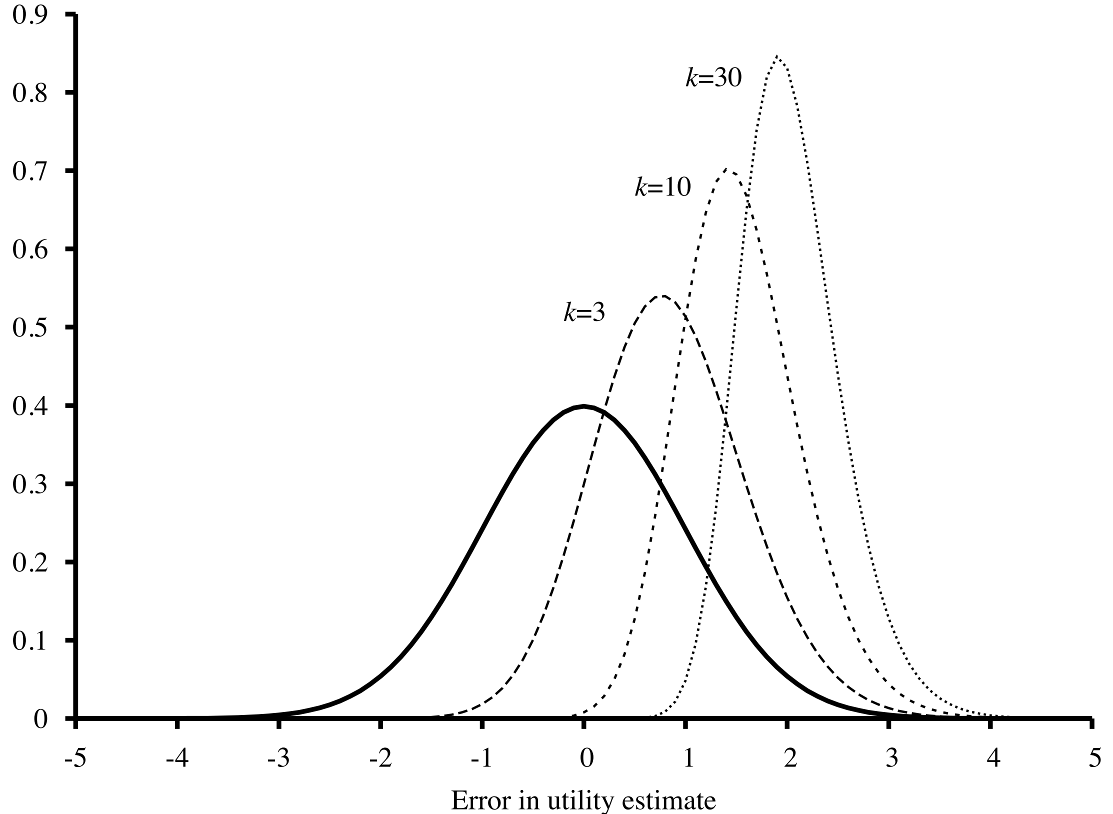


(a) (b)

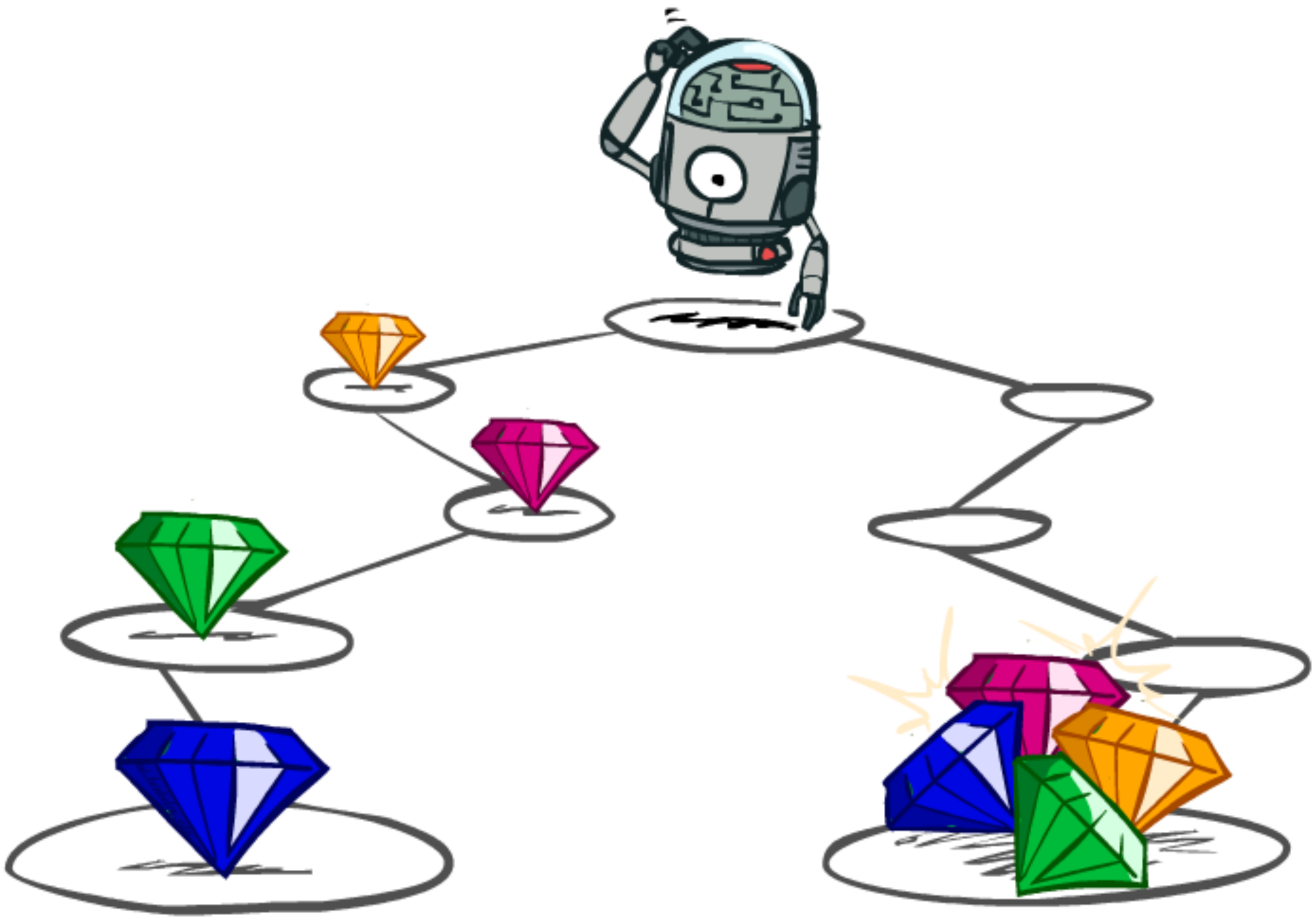
Post-decision Disappointment: the Optimizer's Curse

- Usually we don't have direct access to exact utilities, only ***estimates***
  - E.g., you could make one of k investments
  - An unbiased expert assesses their expected net profit $V_1,\ldots,V_k$
  - You choose the best one V^*
  - With high probability, ***its actual value is considerably less*** than V^*
- This is a serious problem in many areas:
  - Future performance of mutual funds
  - Efficacy of drugs measured by trials
  - Statistical significance in scientific papers
  - Winning an auction

Suppose true net profit is 0
and estimate ~ N(0,1);
Max of k estimates:



Utilities of Sequences

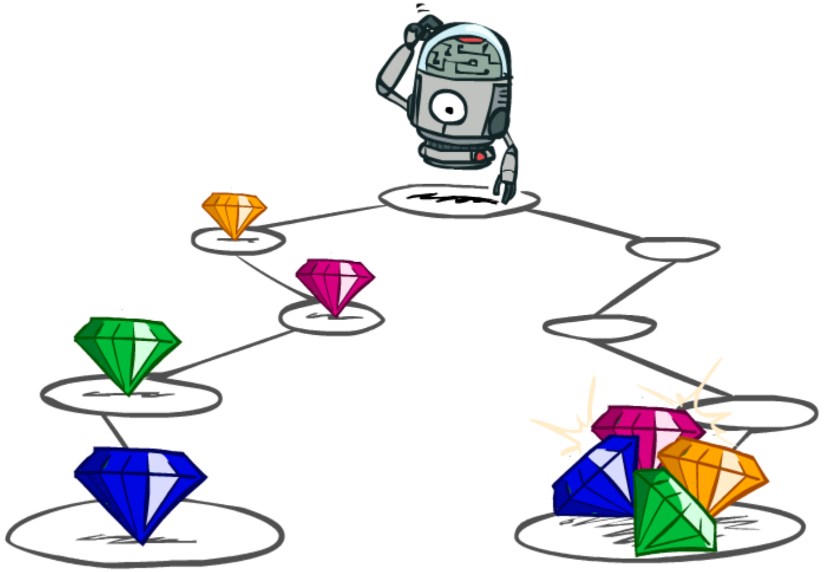


Utilities of Sequences

- What preferences should an agent have over prize sequences?

- More or less? [1, 2, 2] or [2, 3, 4]

- Now or later? [0, 0, 1] or [1, 0, 0]



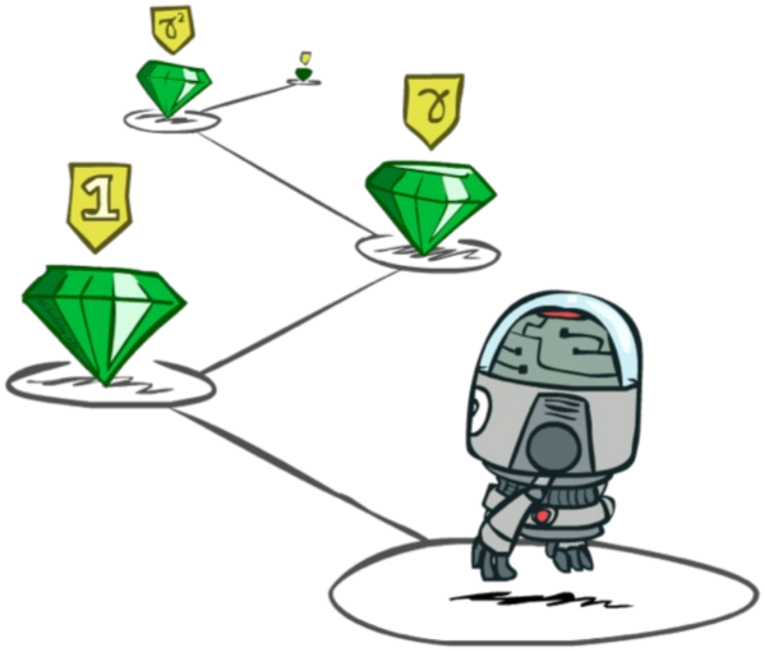
Stationary Preferences

- Theorem: if we assume ***stationary preferences***:
  $[a_1, a_2, \ldots] \succ [b_1, b_2, \ldots] \Leftrightarrow [c, a_1, a_2, \ldots] \succ [c, b_1, b_2, \ldots]$
  then there is only one way to define utilities:

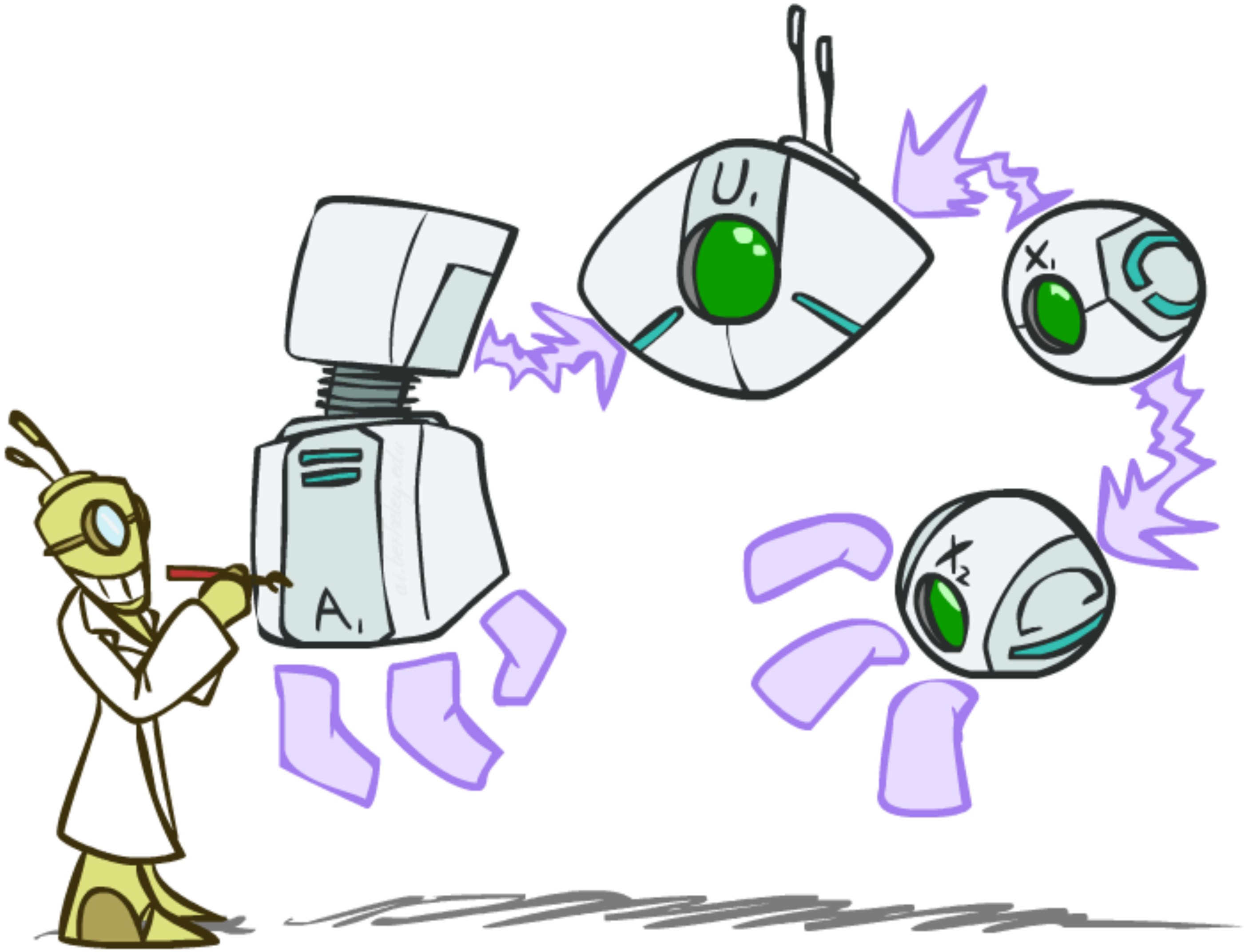
  - ***Additive discounted utility***:

    $$U([r_0, r_1, r_2, \ldots]) = r_0 + \gamma r_1 + \gamma^2 r_2 + \ldots$$

    where $\gamma \in [0,1]$ is the ***discount factor***



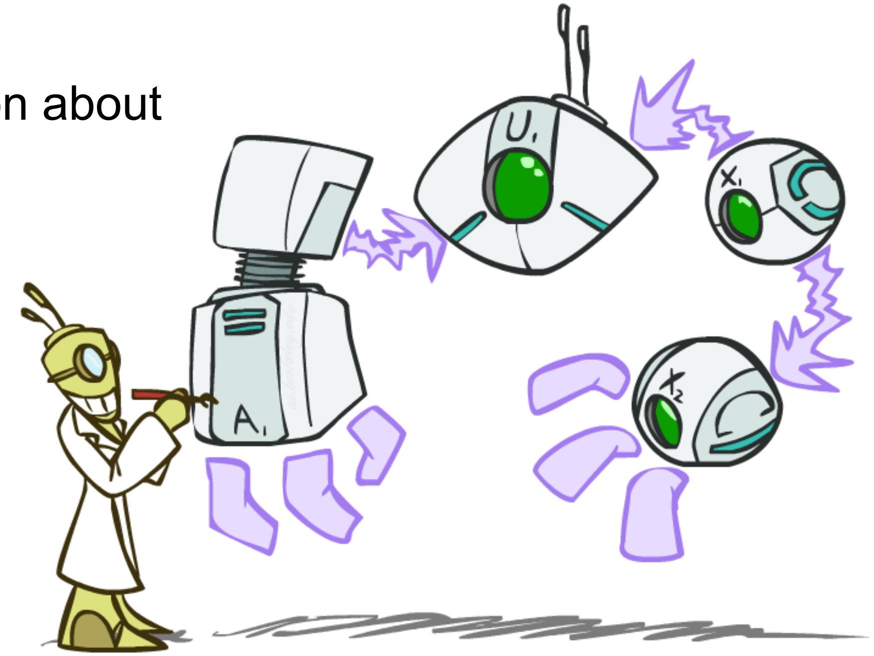
Decision Networks



Decision Networks

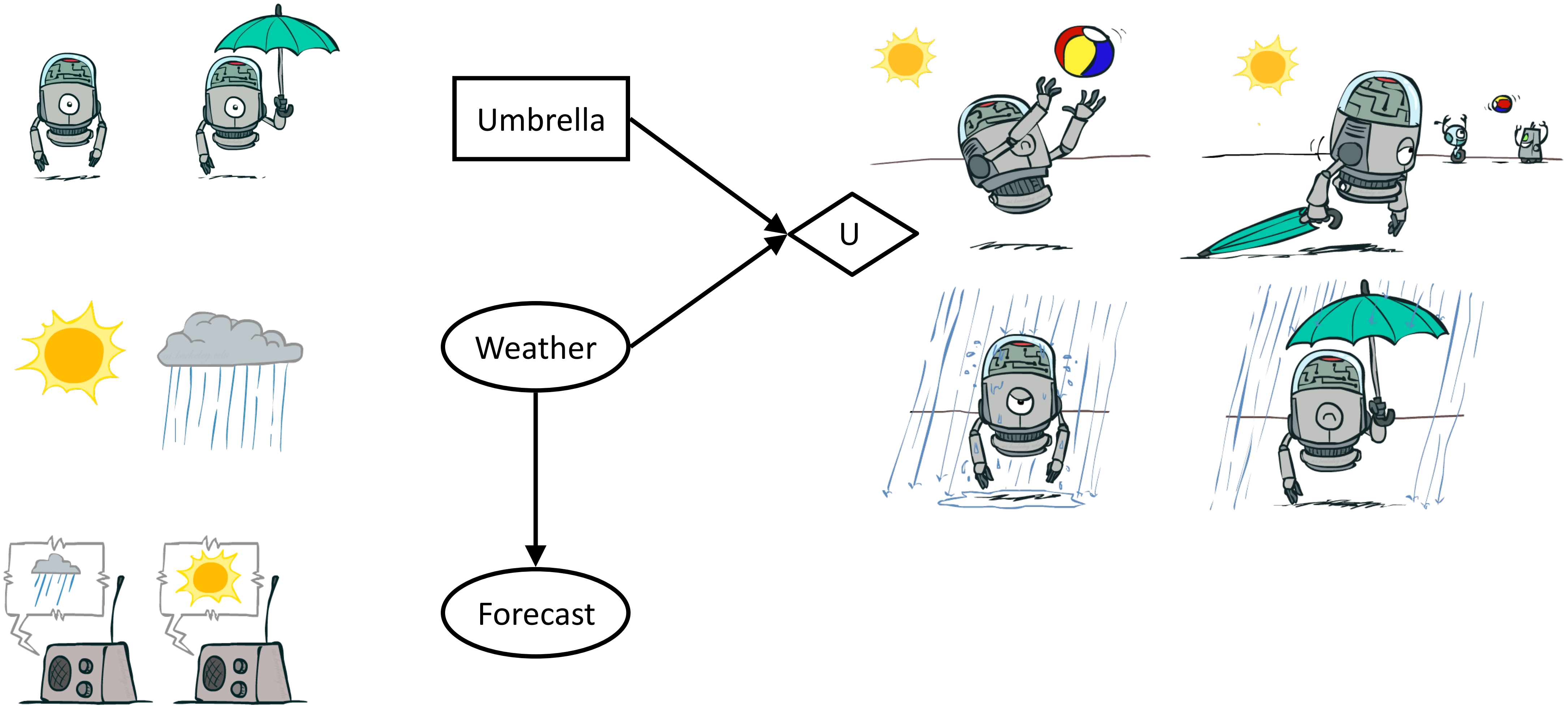
In its most general form, a decision network represents information about

- Its current state
- Its possible actions
- The state that will result from its actions
- The **utility** of that state



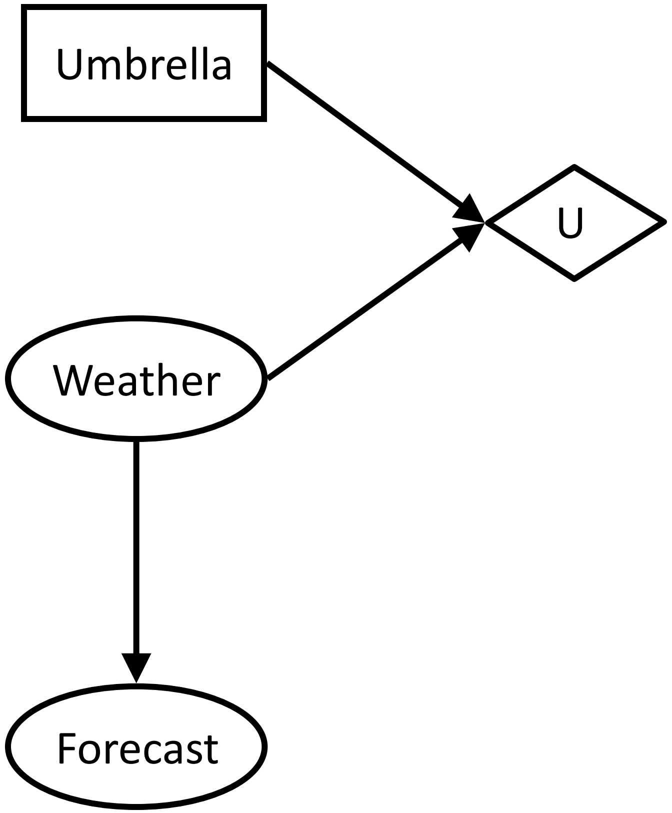
Decision network = Bayes net + Actions + Utilities

Decision Networks



Decision Networks

- A **rational agent** should choose the action that maximizes its expected utility, given its knowledge
- Decision network = Bayes net + Actions + Utilities
  - ***Chance nodes*** (just like BNs)
  - ***Action nodes*** (rectangles, cannot have parents, will have value fixed by algorithm)
  - ***Utility nodes*** (diamond, depends on action and chance nodes)



- Decision algorithm:
  - Fix evidence ***e***
  - For each possible action *a*
    - Fix action node to *a*
    - Compute posterior $P(\mathbf{W}|\mathbf{e},a)$ for parents $\mathbf{W}$ of U
    - Compute expected utility $\sum_{\mathbf{w}} P(\mathbf{w}|\mathbf{e},a)\, U(a,\mathbf{w})$
  - Return action with highest expected utility

Bayes net inference!

Maximum Expected Utility

Umbrella = leave

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

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

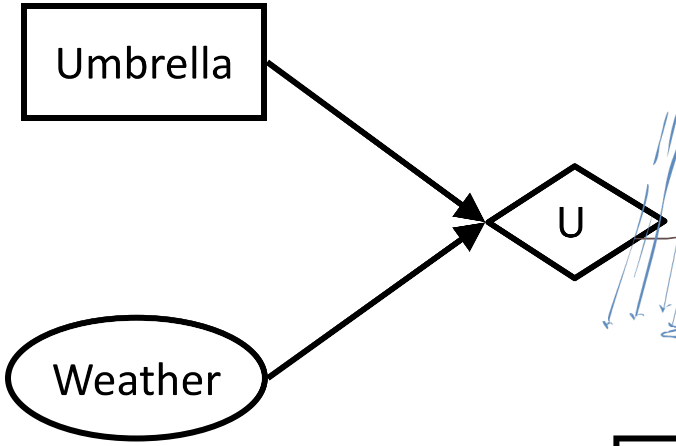
Umbrella = take

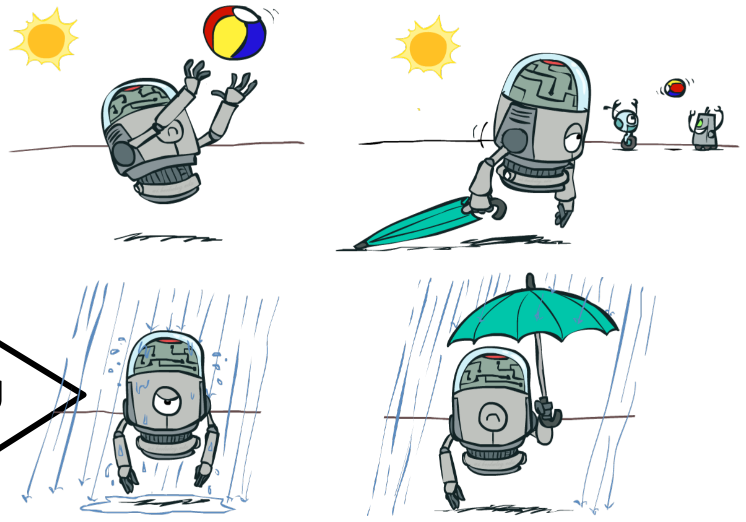
$$\mathrm{EU}(\mathrm{take}) = \sum_w P(w)U(\mathrm{take}, w)$$

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

Optimal decision = leave

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

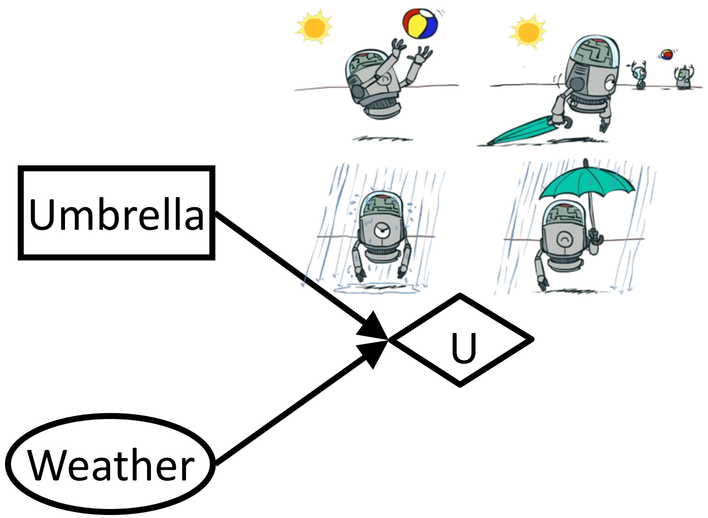


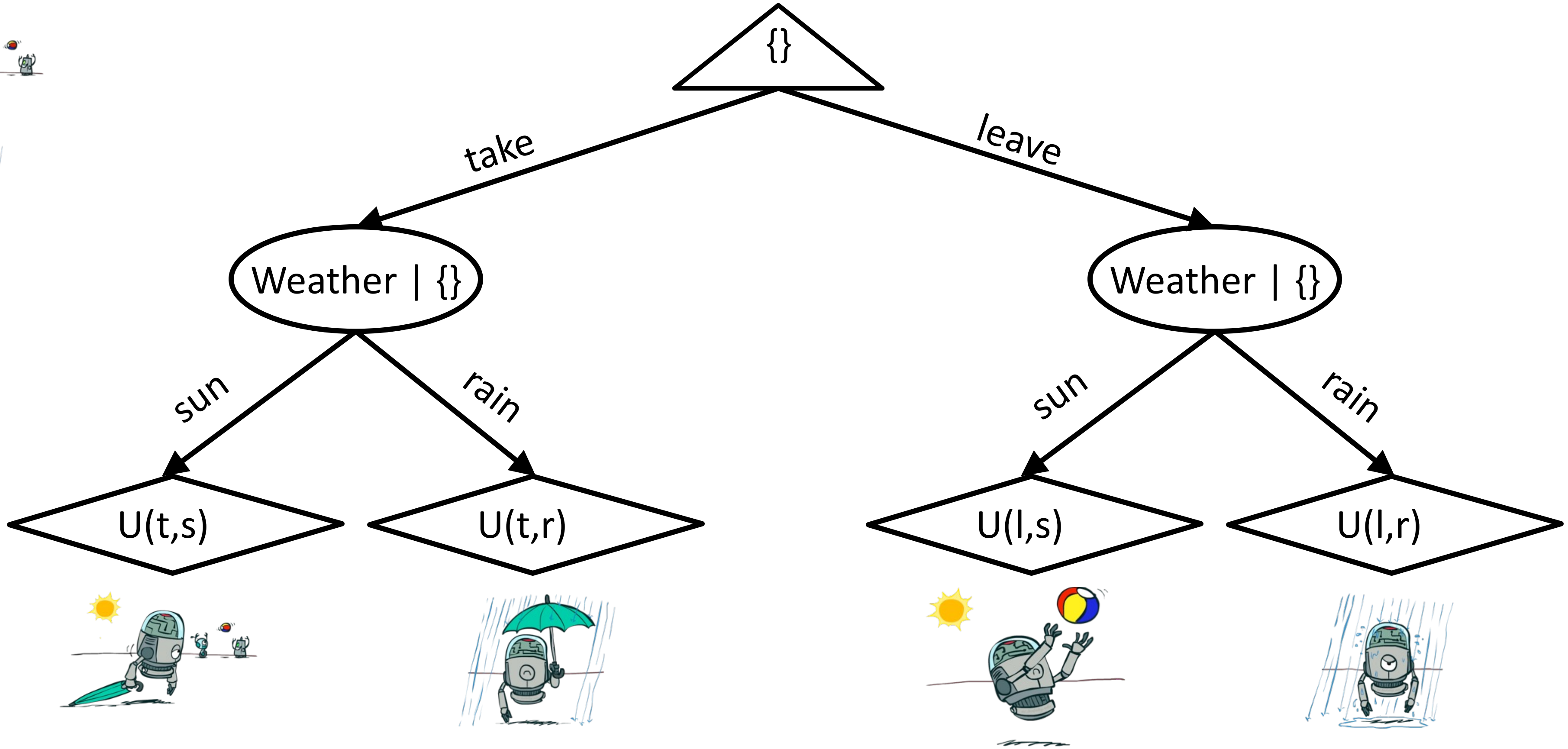


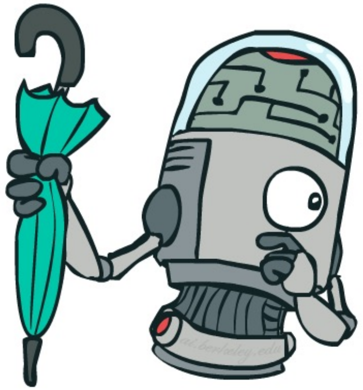
W	P(W)
sun	0.7
rain	0.3

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

Decisions as Outcome Trees



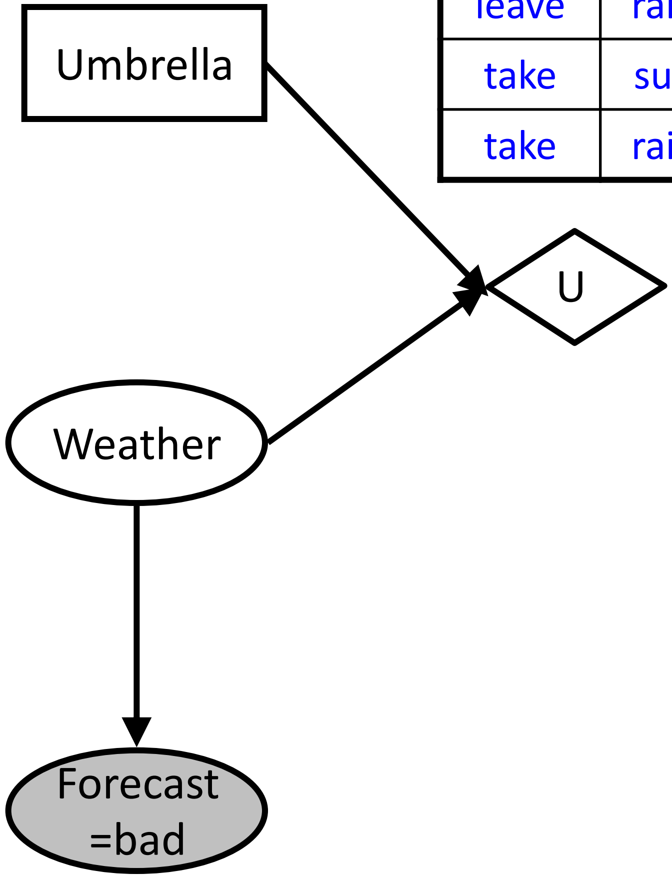




- Almost exactly like expectimax!
- What’s changed?

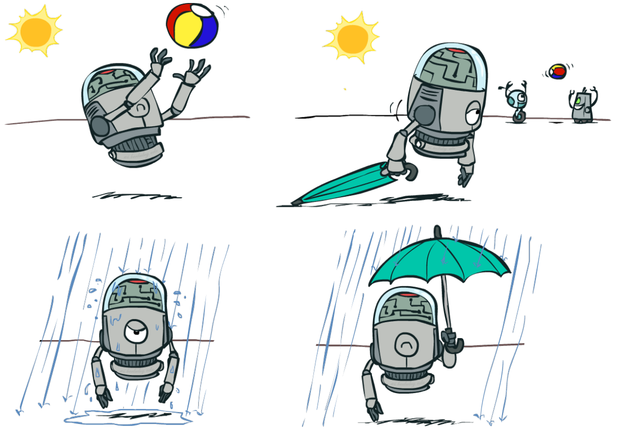
Example: Take an umbrella?

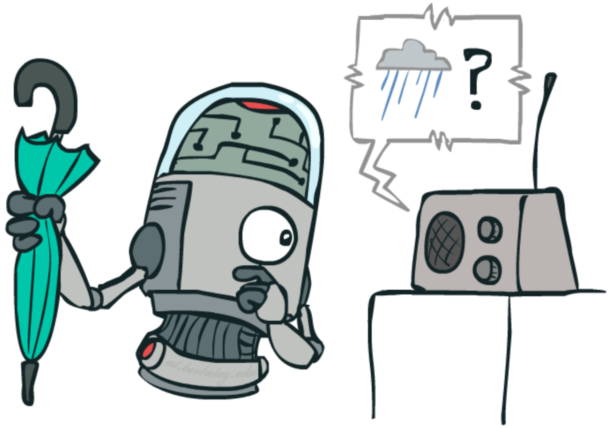
A	W	U(A,W)
leave	sun	100
leave	rain	0
take	sun	20
take	rain	70



W	P(W)
sun	0.7

W	P(F=bad|W)
sun	0.17
rain	0.77



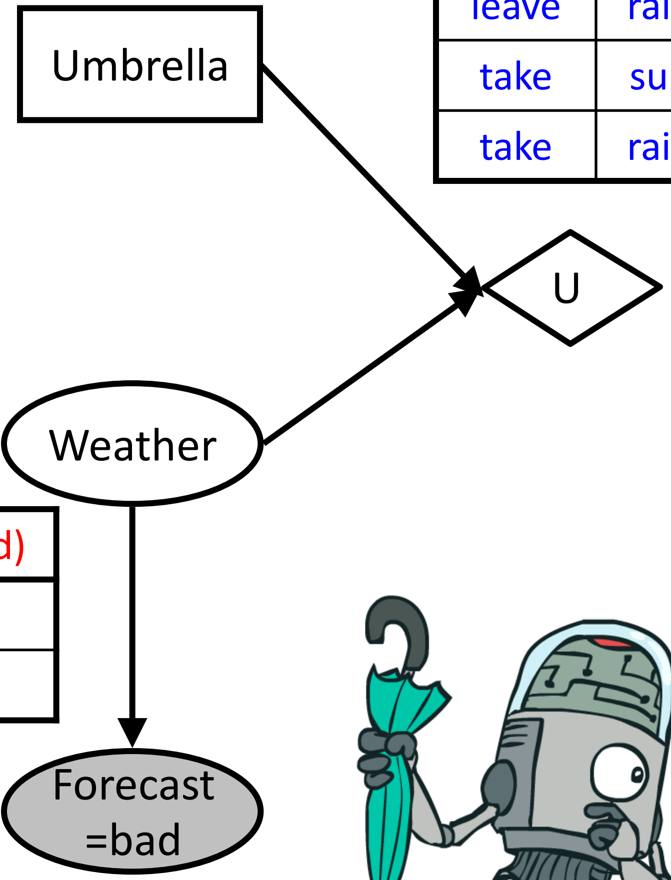


Example: Take an umbrella?

- Decision algorithm:
  - Fix evidence ***e***
  - For each possible action *a*
    - Fix action node to *a*
    - Compute posterior $P(\mathbf{W}|\mathbf{e})$ for parents $\mathbf{W}$ of U
    - Compute expected utility of action a: $\sum_{\mathbf{w}} P(\mathbf{w}|\mathbf{e})\ U(a,\mathbf{w})$
  - Return the action with highest expected utility

Bayes net inference!

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



Umbrella = leave

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

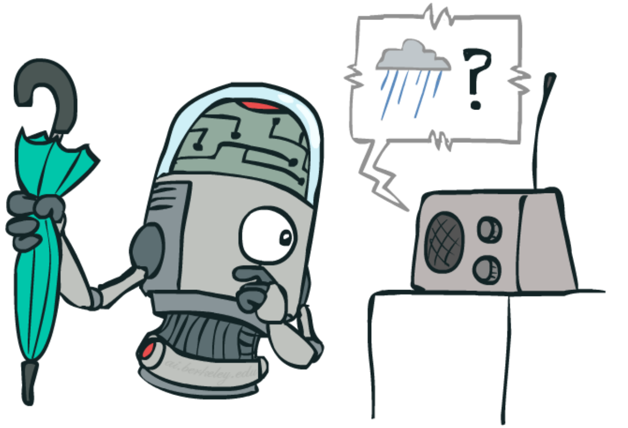
We have: $P(W) \quad P(F|W)$

$$P(W|F) = \frac{P(W,F)}{\sum_w P(w,F)} = \frac{P(F|W)P(W)}{\sum_w P(F|w)P(w)}$$

W	P(W)
sun	0.7

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

W	P(F=bad|W)
sun	0.17
rain	0.77

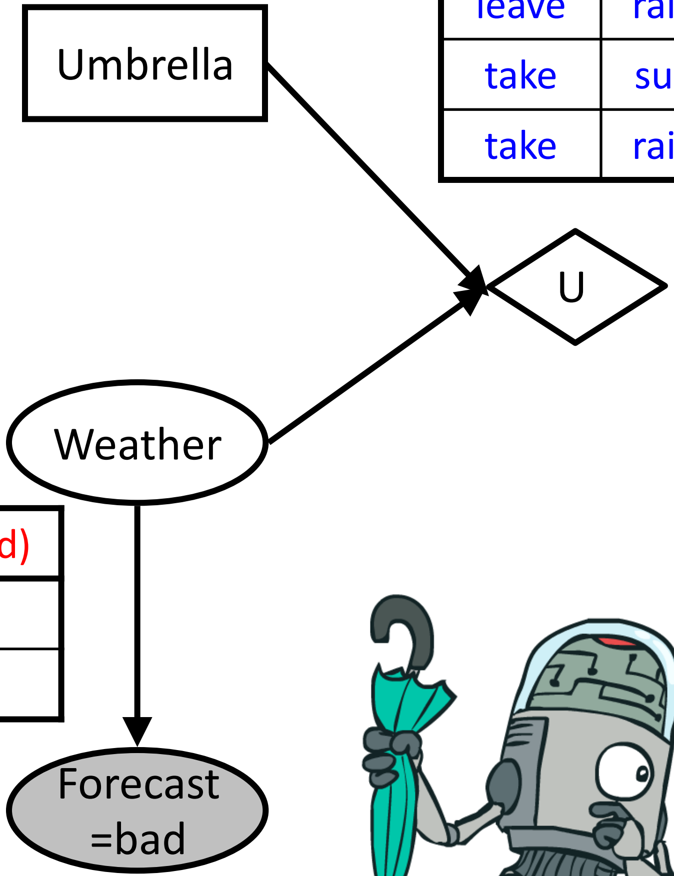


Example: Take an umbrella?

- Decision algorithm:
  - Fix evidence ***e***
  - For each possible action *a*
    - Fix action node to *a*
    - Compute posterior $P(\boldsymbol{W}|\boldsymbol{e},a)$ for parents $\boldsymbol{W}$ of U
    - Compute expected utility of action a: $\sum_{\boldsymbol{w}} P(\boldsymbol{w}|\boldsymbol{e},a)\ U(a,\boldsymbol{w})$
  - Return action with highest expected utility

Bayes net inference!

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



W	P(W)
sun	0.7

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

W	P(F=bad|W)
sun	0.17
rain	0.77

Umbrella = leave

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

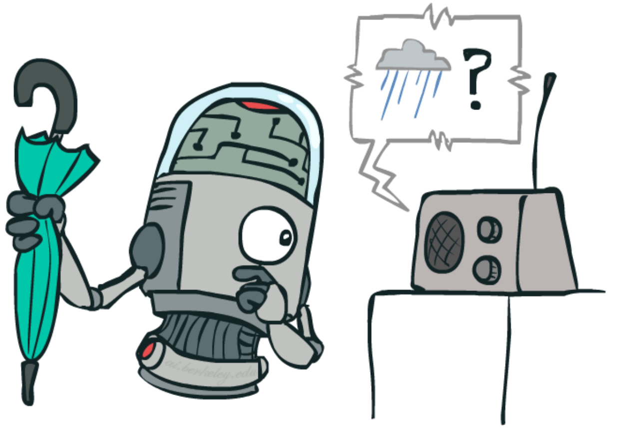
$$= 0.34 \times 100 + 0.66 \times 0 = 34$$

Umbrella = take

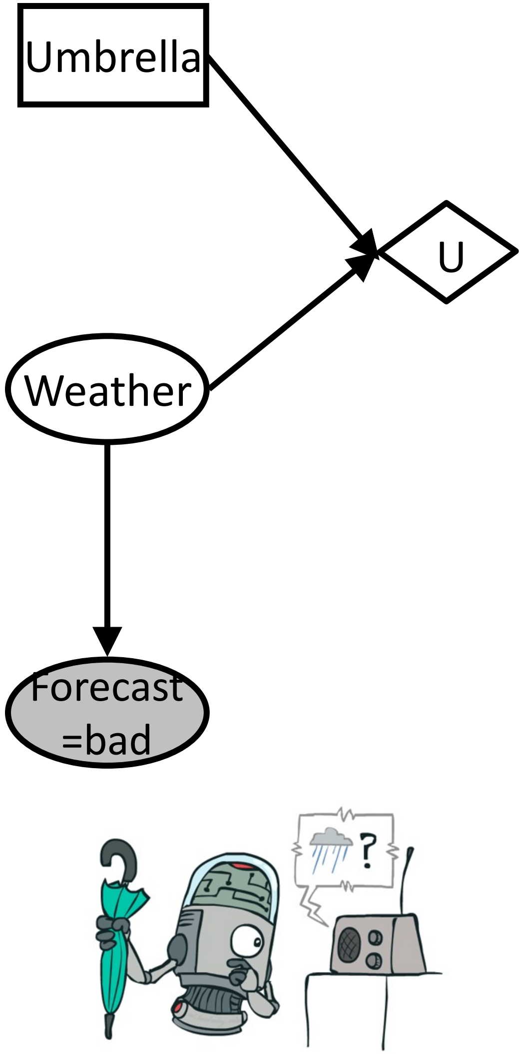
$$EU(\text{take}|F=\text{bad}) = \sum_w P(w|F=\text{bad})\ U(\text{take},w)$$

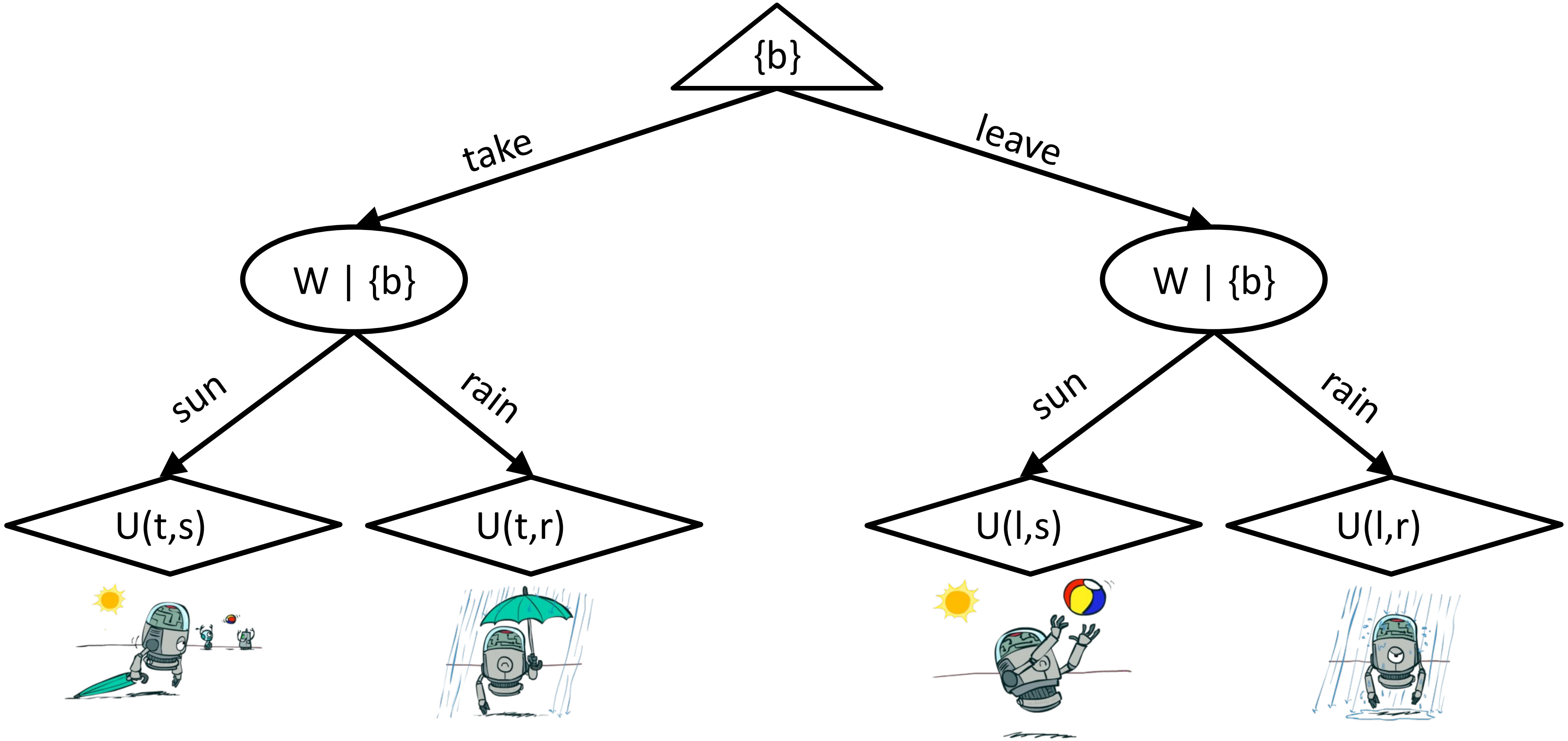
$$= 0.34 \times 20 + 0.66 \times 70 = 53$$

Optimal decision = take!



Decisions as Outcome Trees

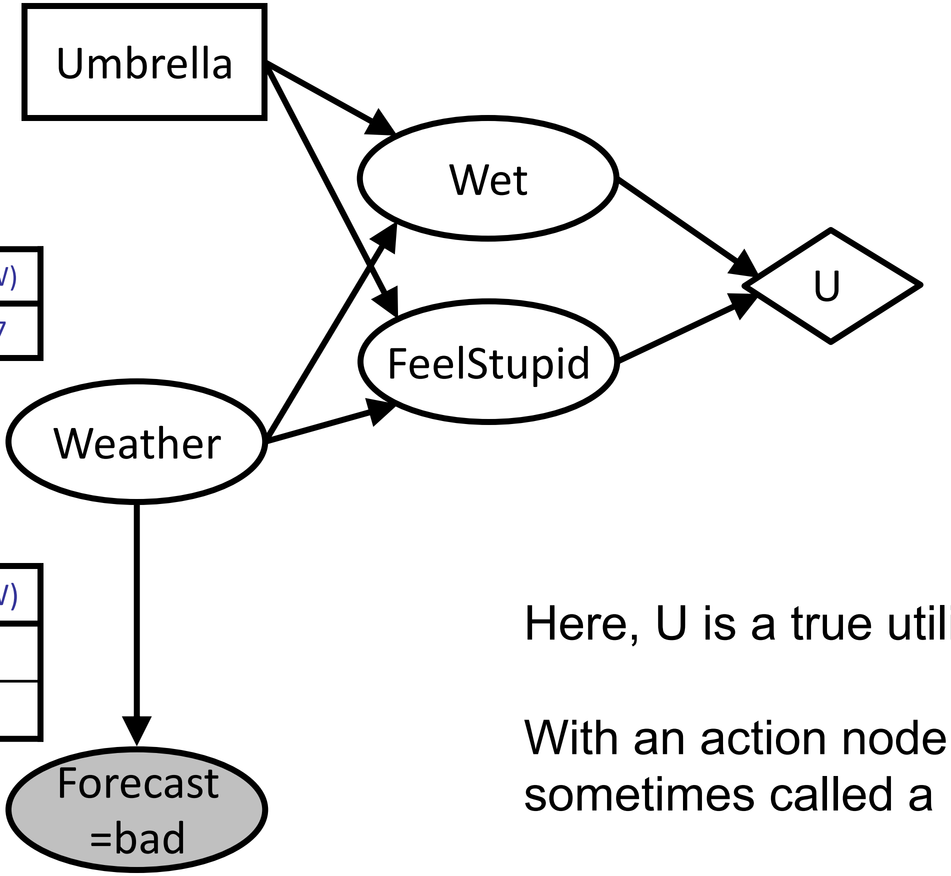




Decision network with utilities on outcome states

W	P(W)
sun	0.7

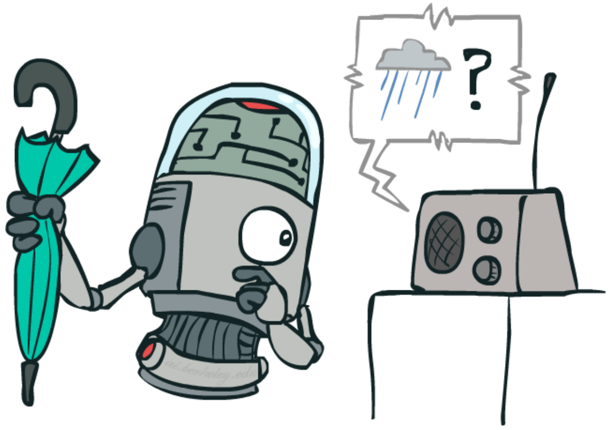
W	P(F=bad|W)
sun	0.17
rain	0.77



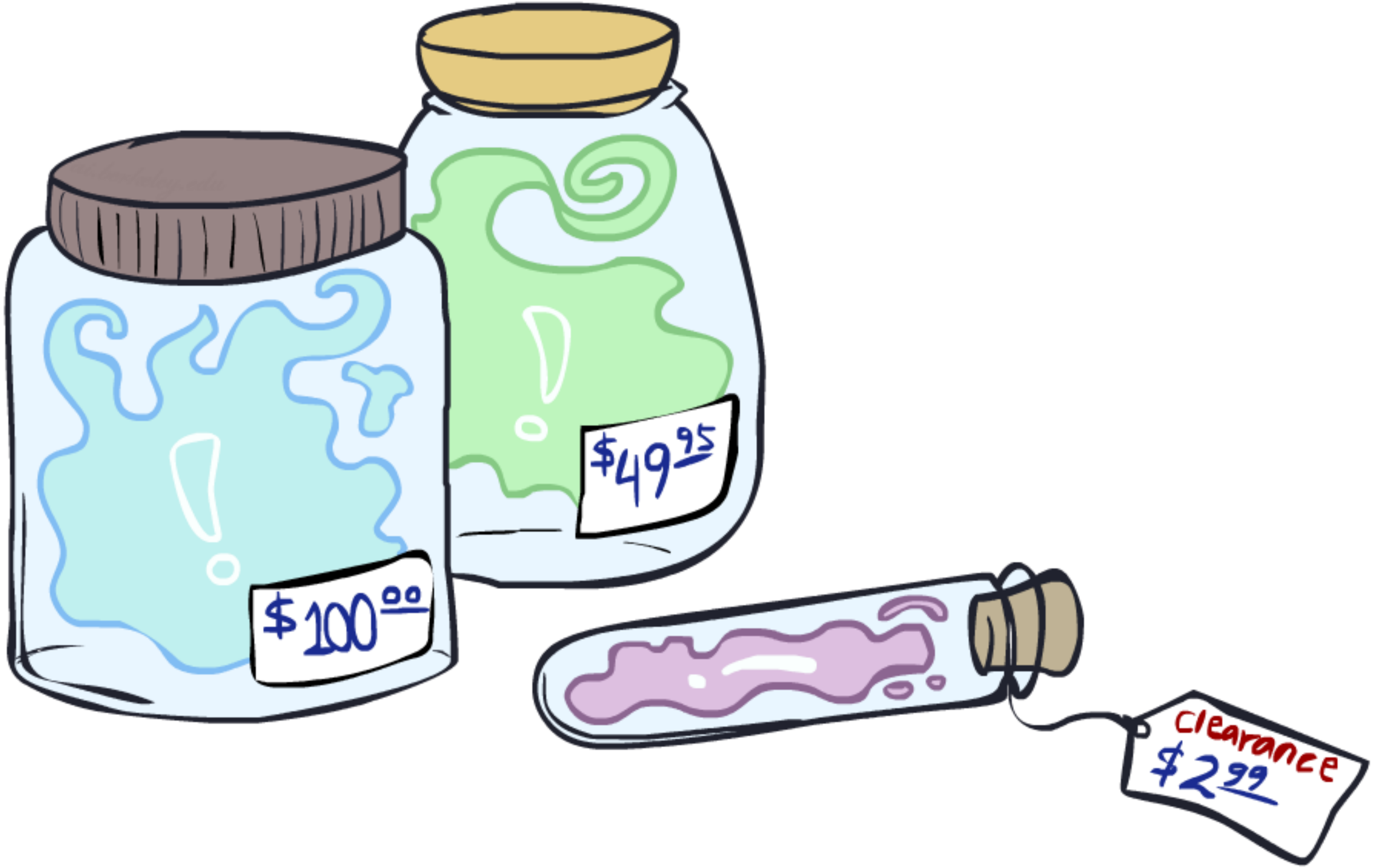
Wet	FeelStupid	U
dry	false	100
wet	true	0
dry	true	20
damp	false	70

Here, U is a true utility.

With an action node as parent, it is sometimes called a ***Q-value***



Value of Information



A question to motivate VPI

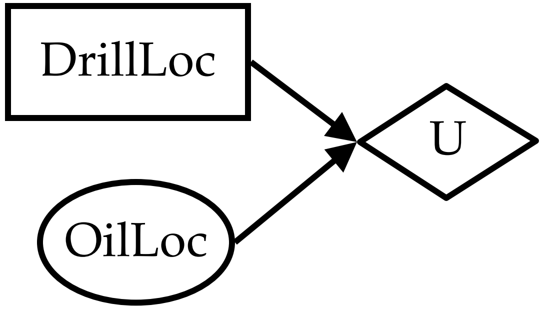
How valuable is a weather forecast?

How useful is it to get the evidence?

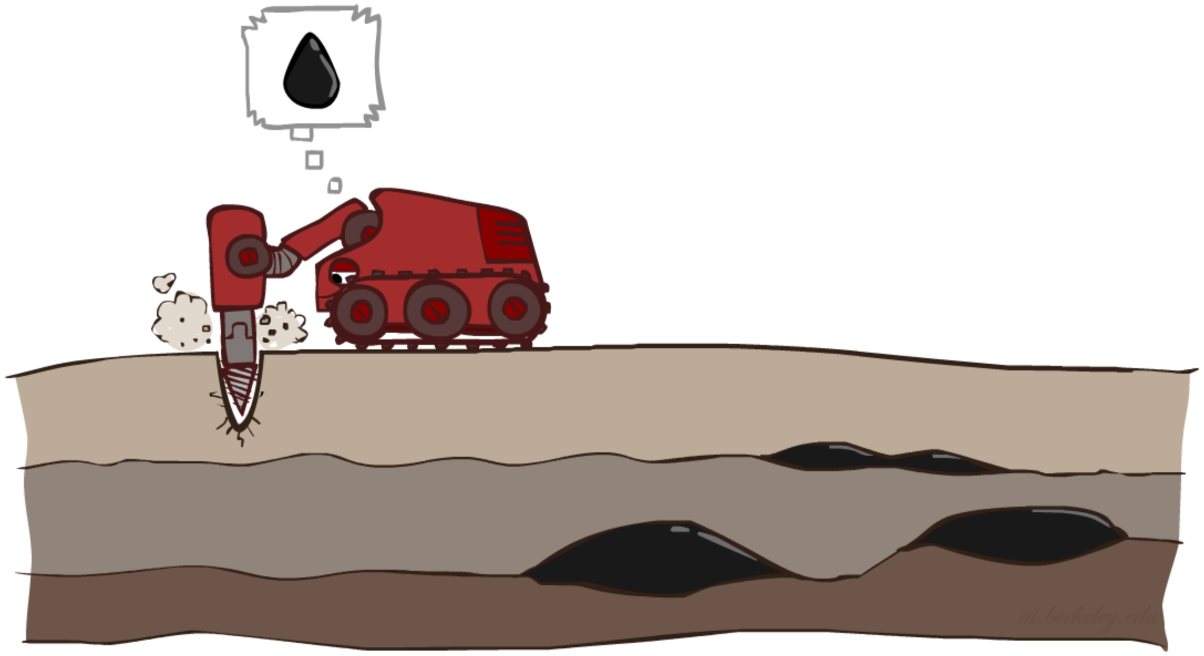
Value of Perfect Information

- Idea: compute value of acquiring evidence
  - Can be done directly from decision network
- Example: buying oil drilling rights
  - Two blocks A and B, exactly one has oil, worth k
  - You can drill in one location
  - Prior probabilities 0.5 each, & mutually exclusive
  - Drilling in either A or B has EU = k/2, MEU = k/2
- Question: what's the value of information of O?
  - Value of knowing which of A or B has oil
  - Value is expected gain in MEU from new info
  - If we know OilLoc, MEU is k (either way)
  - Gain in MEU from knowing OilLoc?
  - VPI(OilLoc) = k – k/2 = k/2
  - Fair price of information: 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

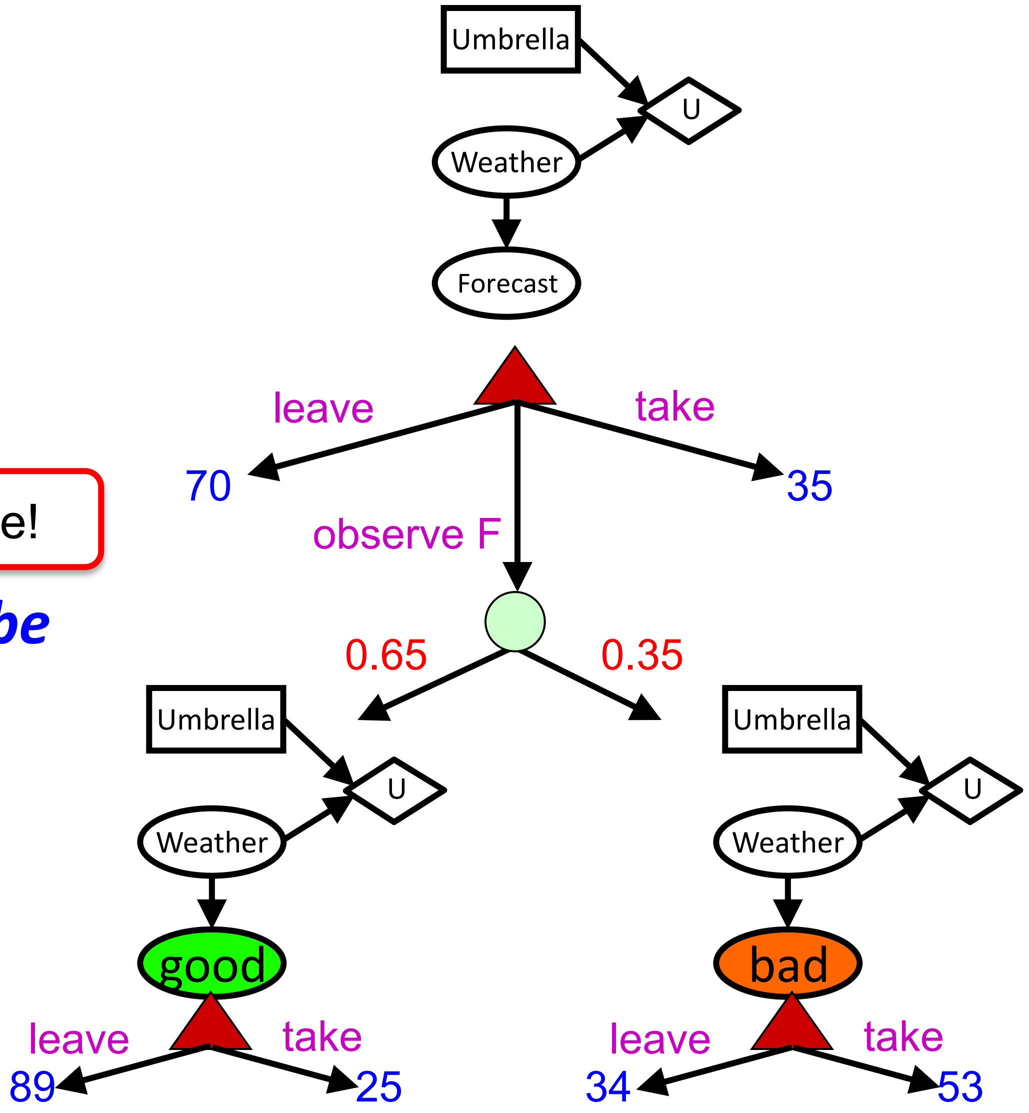


Value of information

- Before you see the forecast (no evidence)
  - $MEU(\emptyset) = \max_a EU(a) = 70$
- ***What if you look at the forecast?***
- If Forecast=bad
  - $MEU(F=bad) = \max_a EU(a \mid F=bad) = 53$
- If Forecast=good
  - $MEU(F=good) = \max_a EU(a \mid$
- ***But, we don't know what the ecast will be ahead of time!***
- So we need a distribution of P(F)
- Expected utility given forecast
  - $= 0.35 \times 53 + 0.65 \times 89 = 76.4$
- ***Value of information*** $= 76.4-70 = 6.4$

Bayes net inference!

F	P(F)
good	0.65
bad	0.35



Value of Information

- Assume we have evidence E=e. Value if we act now:

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

- Assume we see that E' = e'. Value if we act then:

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

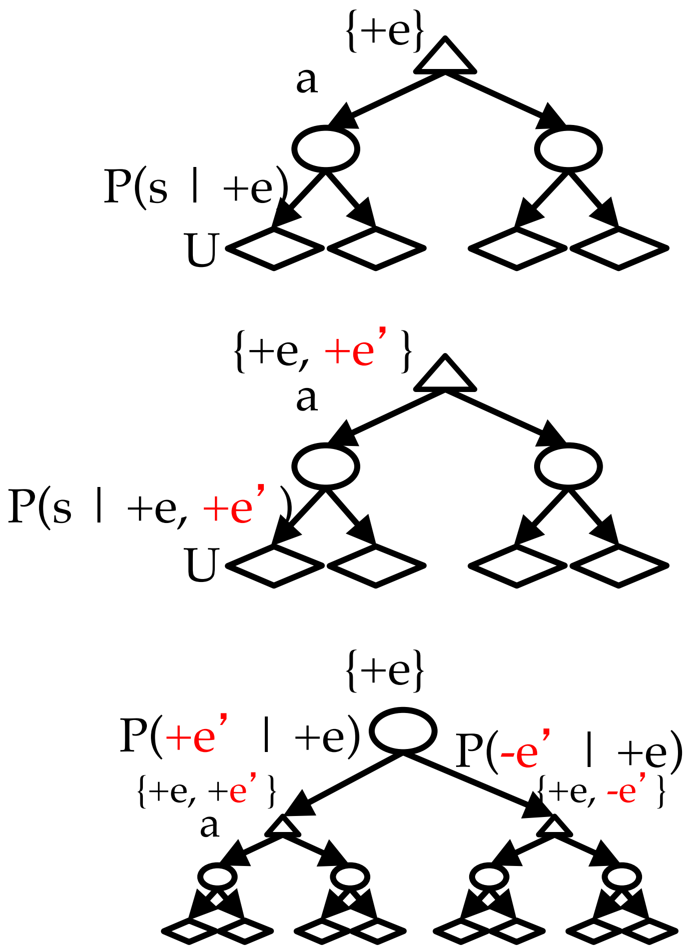
- BUT E' is a random variable whose value is unknown, so we don't know what e' will be

- Expected value if E' is revealed and then we act:

$$\text{MEU}(e,E') = \sum_{e'} P(e'|e)\text{MEU}(e,e')$$

- Value of information: how much MEU goes up by revealing E' first then acting, over acting now:

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



Value of Information

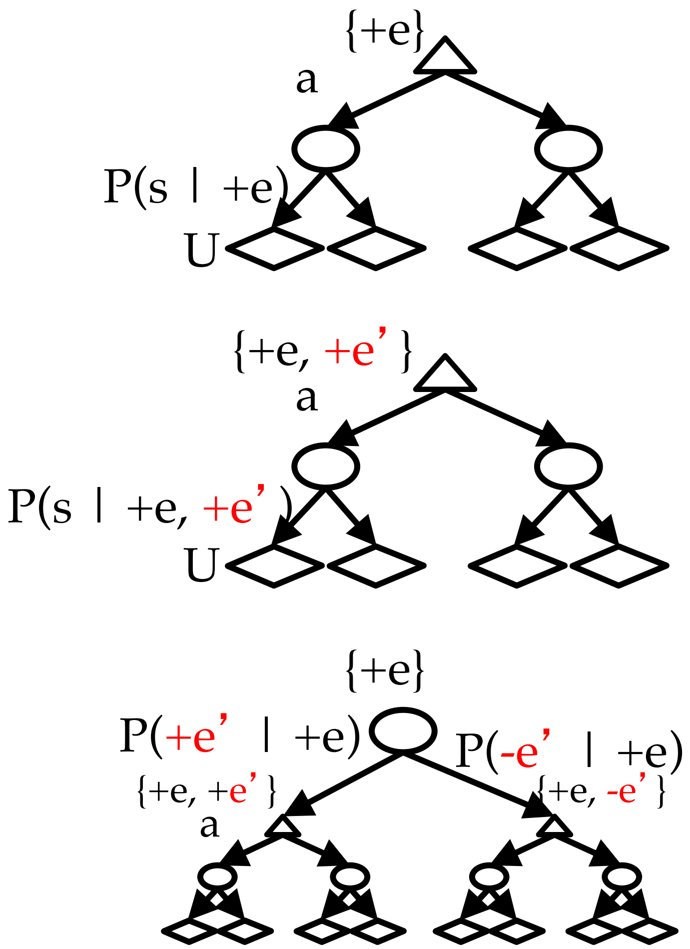
$$\text{MEU}(e, E') = \sum_{e'} P(e'|e)\text{MEU}(e, e')$$

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

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

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

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



VPI Properties

VPI is non-negative! $\mathrm{VPI}(E_i \mid e) \geq 0$

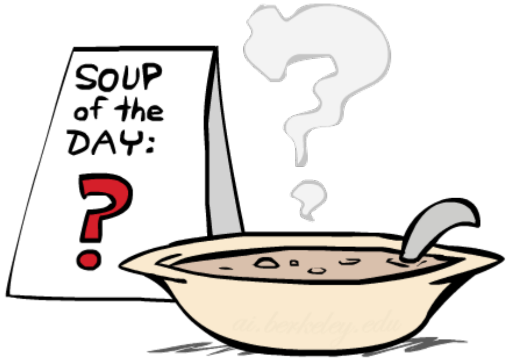
VPI is not (usually) additive: $\mathrm{VPI}(E_i, E_j \mid e) \neq \mathrm{VPI}(E_i \mid e) + \mathrm{VPI}(E_j \mid e)$

VPI is order-independent: $\mathrm{VPI}(E_i, E_j \mid e) = \mathrm{VPI}(E_j, E_i \mid e)$

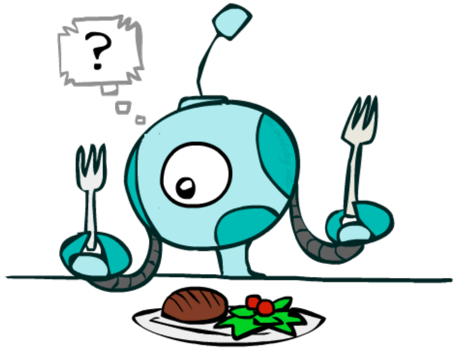
Quick VPI Questions

- The soup of the day is either clam chowder or split pea, but you wouldn't order either one. What's the value of knowing which it is?

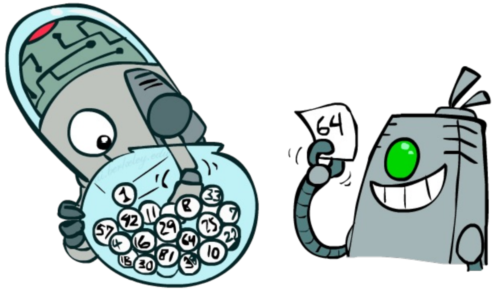
  (zero/positive)



- There are two kinds of plastic forks at a picnic. One kind is slightly sturdier. What's the value of knowing which?



- You're playing the lottery. The prize will be $0 or $100. You can play any number between 1 and 100 (chance of winning is 1%). What is the value of knowing the winning number?

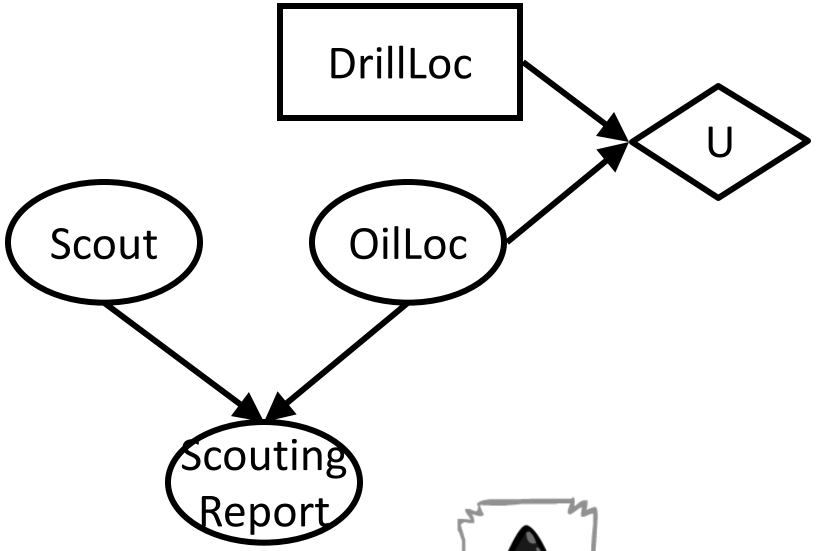


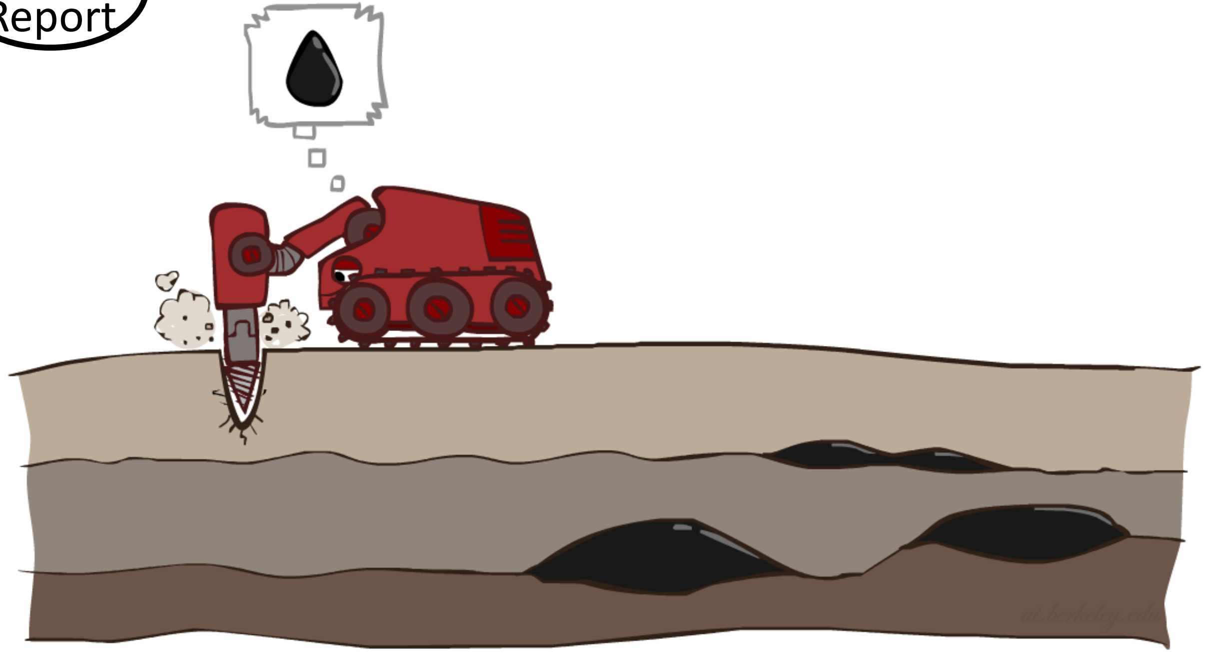
Value of Imperfect Information?

- No such thing
- Information corresponds to the observation of a node in the decision network
- If data is “noisy” that just means we don’t observe the original variable, but another variable which is a noisy version of the original one

VPI Question

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





- Generally:

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

  Then VPI(Z | CurrentEvidence) = 0