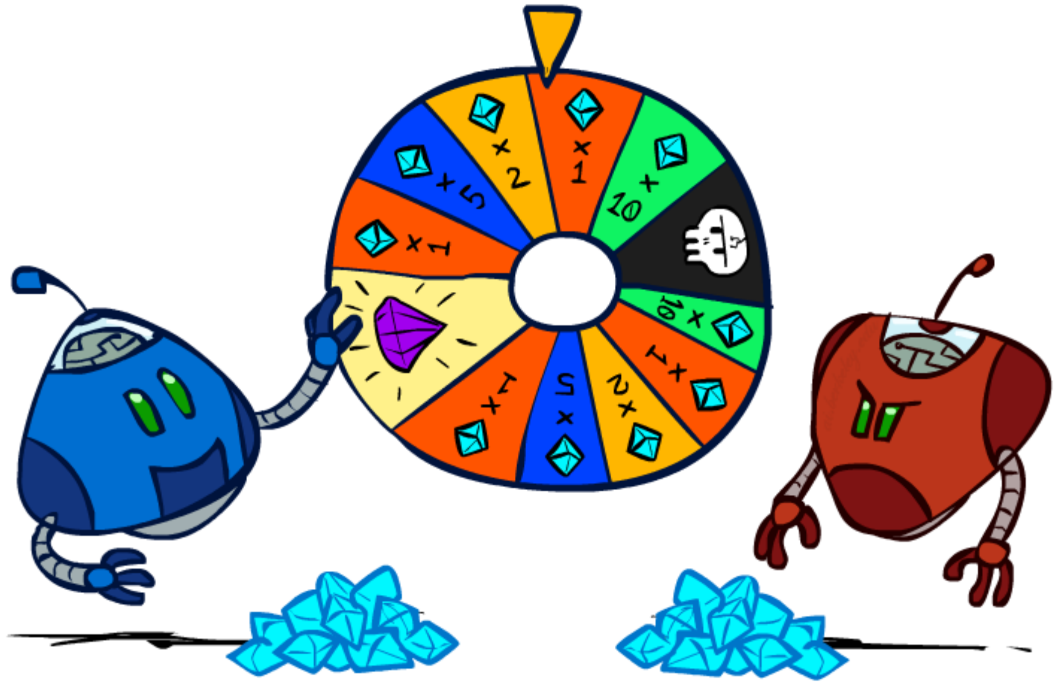
CS 188: Artificial Intelligence

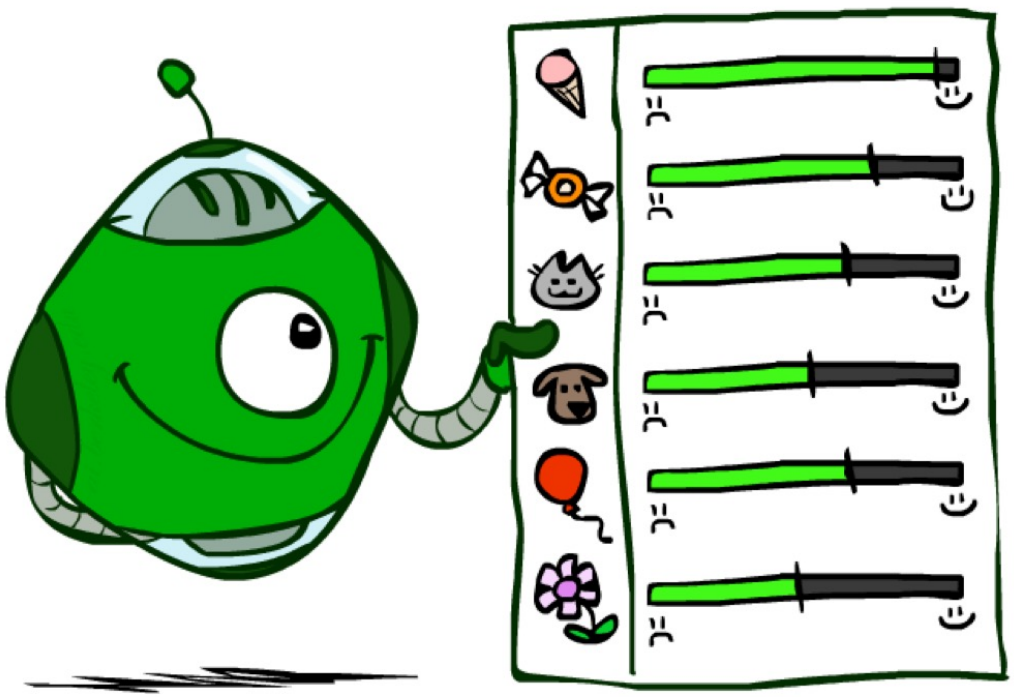
Rational Decisions



Slides from Stuart Russell and Peyrin Kao

University of California, Berkeley

Utilities



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

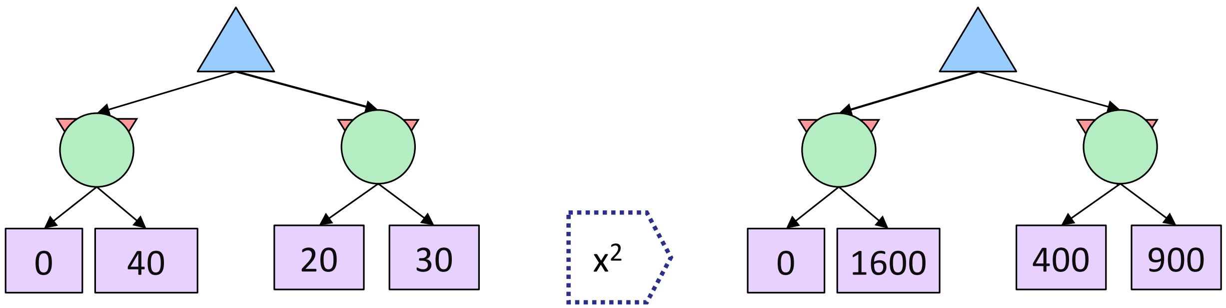
- We hard-wire utilities and let behaviors emerge
  - Why don't we hard-wire behaviors?

Maximum Expected Utility

- Principle of maximum expected utility:
  - A rational agent should chose the action that maximizes its expected utility, given its knowledge

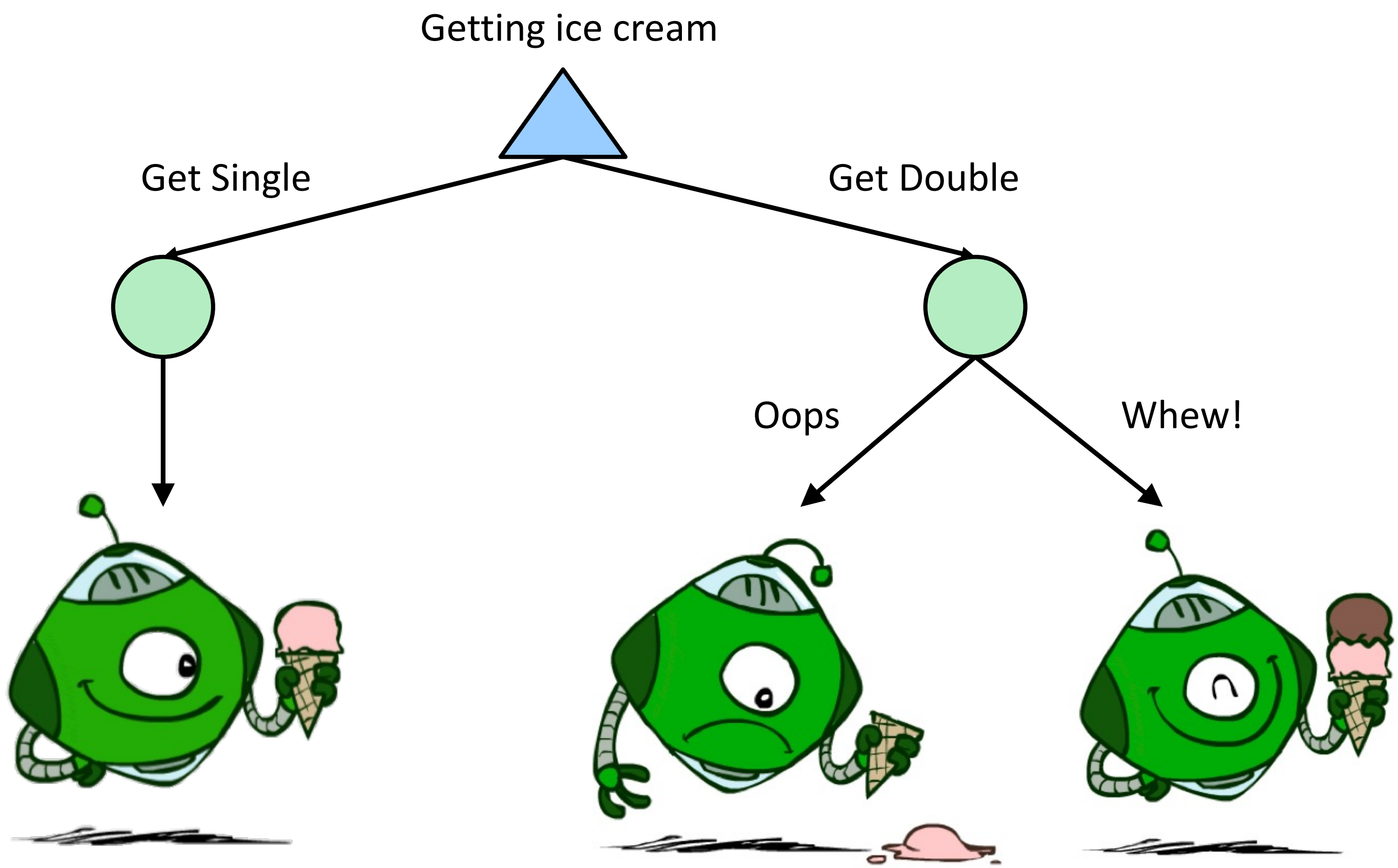
- Questions:
  - Where do utilities come from?
  - How do we know such utilities even exist?
  - How do we know that averaging makes sense?
  - What if our behavior (preferences) can't be described by utilities?

Utility magnitudes are meaningful

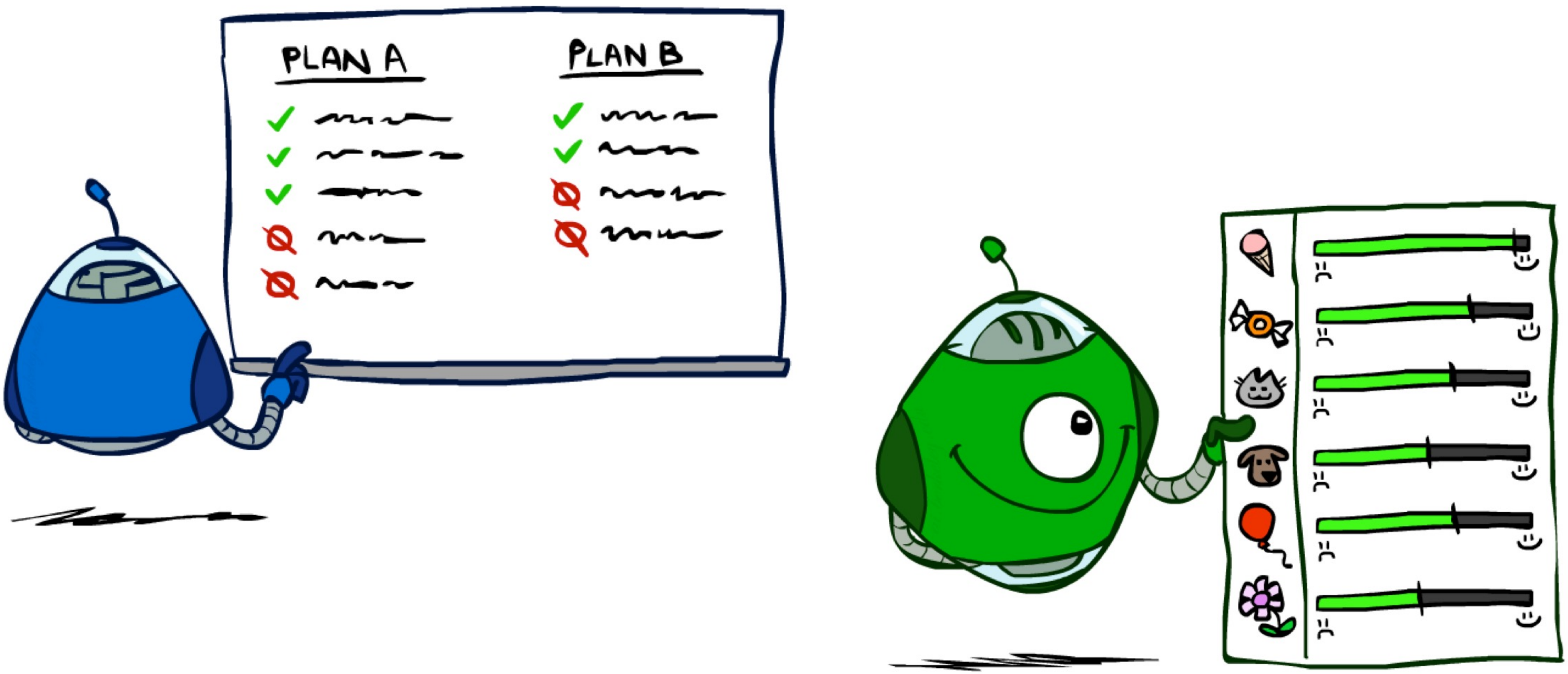


- For worst-case minimax reasoning, terminal value scale doesn't matter
  - We just want better states to have higher evaluations (get the ordering right)
  - The optimal decision is invariant under any ***monotonic transformation***
- For average-case expectimax reasoning, we need ***magnitudes*** to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

Utilities: Uncertain Outcomes



Deriving Utilities from Rational Preferences



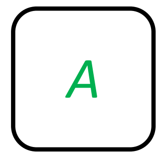
Preferences

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

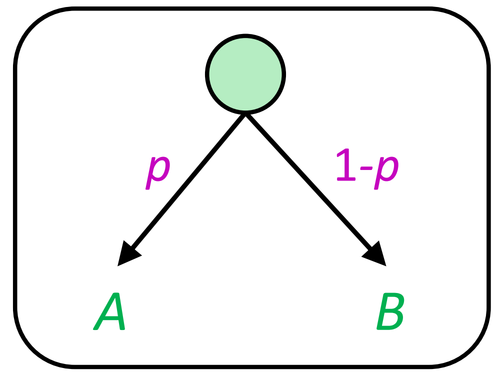
    $L = [p, A;\ (1-p), B]$

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

A Prize



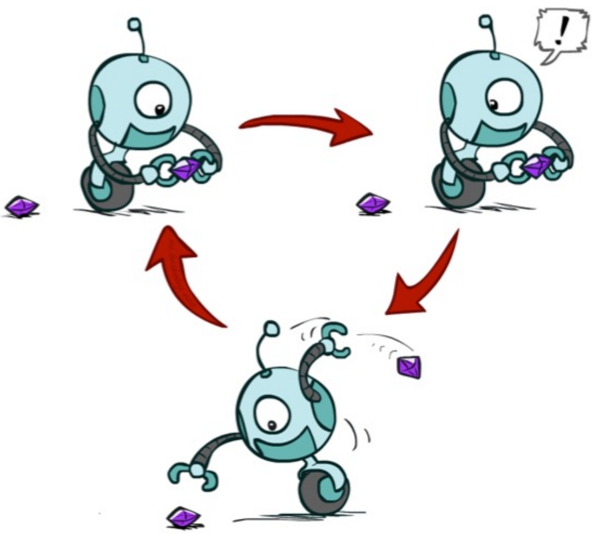
A Lottery



Rational Preferences

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

$$\text{Axiom of Transitivity:} \quad (A \succ B) \wedge (B \succ C) \Rightarrow (A \succ C)$$



- Costs of irrationality:
- An agent with intransitive preferences can be induced to give away all of its money
  - If $B \succ C$, then an agent with C would pay (say) 1 cent to get B
  - If $A \succ B$, then an agent with B would pay (say) 1 cent to get A
  - If $C \succ A$, then an agent with A would pay (say) 1 cent to get C

Rational Preferences

The Axioms of Rationality

Orderability:

$$(A > B) \lor (B > A) \lor (A \sim B)$$

Transitivity:

$$(A > B) \land (B > C) \Rightarrow (A > C)$$

Continuity:

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

Substitutability:

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

Monotonicity:

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

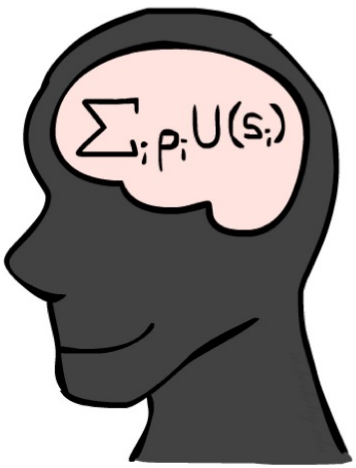
Theorem: Rational preferences imply behavior describable as maximization of expected utility

MEU Principle

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

$$U(A) \geq U(B) \Leftrightarrow A \geq B$$

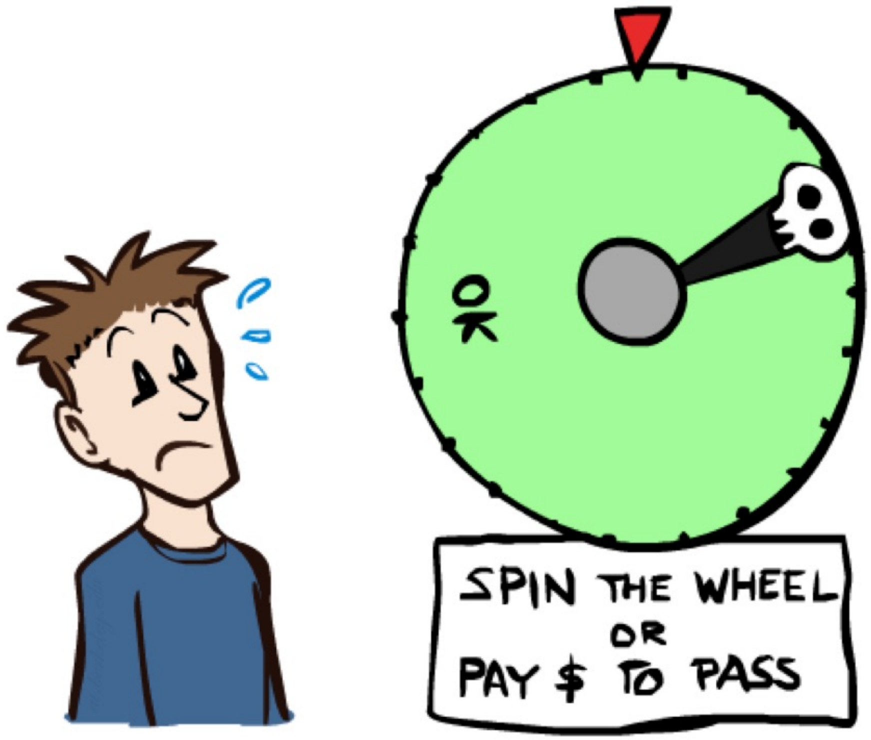
$$U([p_1,S_1; \ldots ; p_n,S_n]) = p_1 U(S_1) + \ldots + p_n U(S_n)$$



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

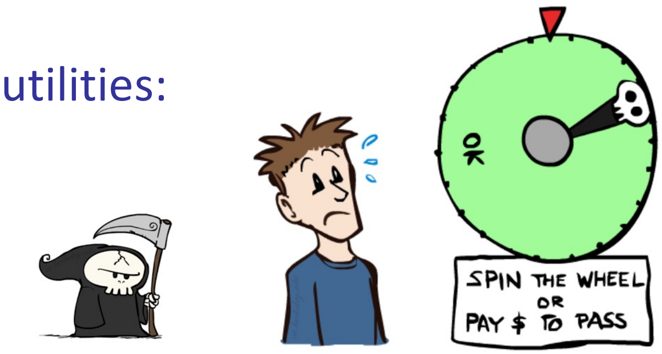
- Maximum expected utility (MEU) principle:
  - Choose the action that maximizes expected utility
  - Note: rationality does ***not*** require representing or manipulating utilities and probabilities
    - E.g., a lookup table for perfect tic-tac-toe

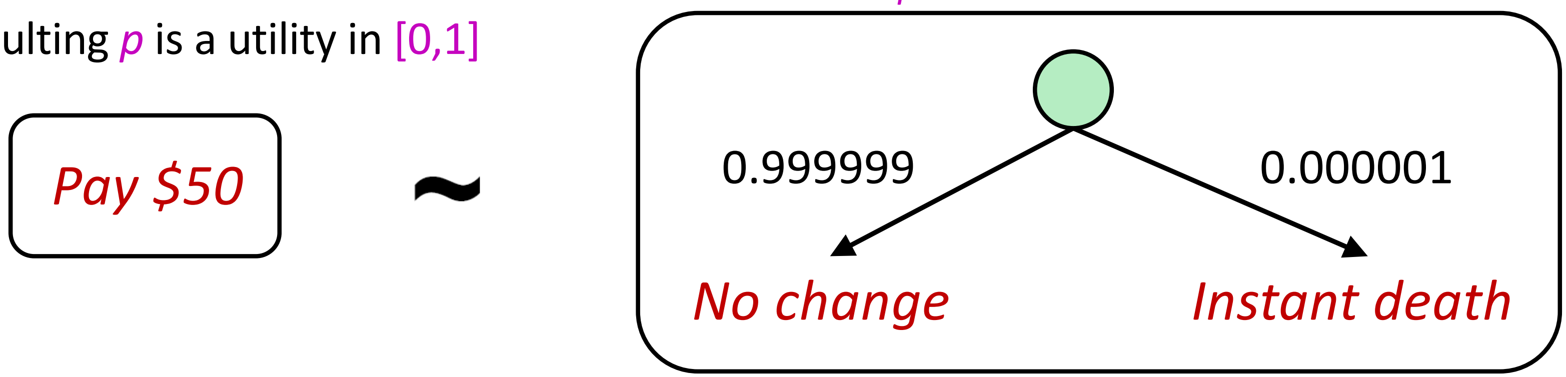
Human Utilities



Human Utilities

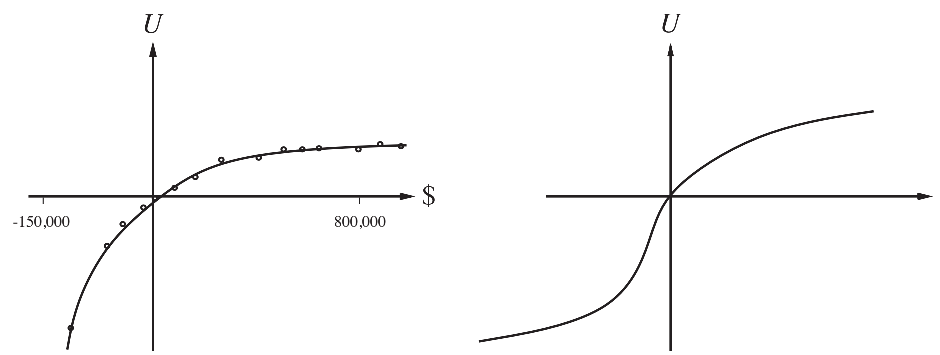
- Utilities map states to real numbers. Which numbers?
- Standard approach to assessment (elicitation) of human utilities:
  - We want to assign a utility to prize *A*
  - Compare a prize *A* to a ***standard lottery*** L_p between
    - "best possible prize" $S_\top$ with probability *p*
    - "worst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





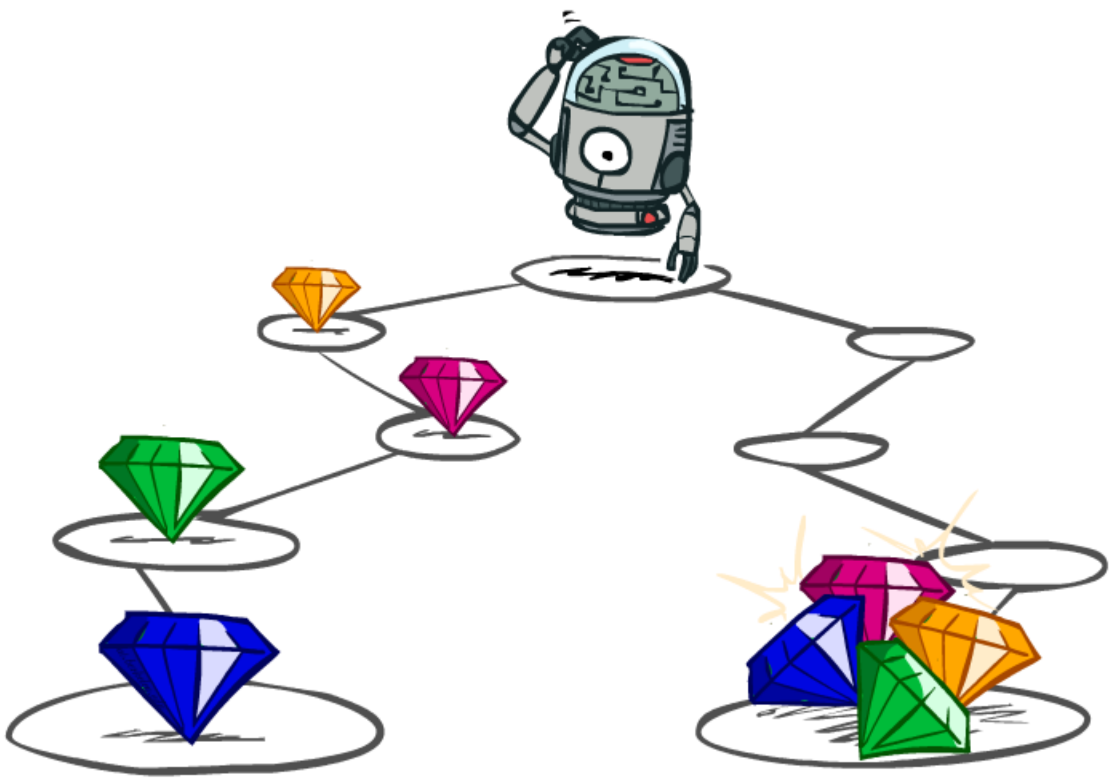
Money

- 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*** $\mathrm{EMV}(L) = pX + (1-p)Y$
  - The utility is $U(L) = pU(\$X) + (1-p)U(\$Y)$
  - Typically, $U(L) < U(\mathrm{EMV}(L))$
  - 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 $\mathrm{CE}(L)$ is the cash amount such that $\mathrm{CE}(L) \sim L$
  - The ***insurance premium*** is $\mathrm{EMV}(L) - \mathrm{CE}(L)$
  - If people were risk-neutral, this would be zero!
    - Pay an insurance premium to get out of a lottery
      - House burns down, cybercriminals take your company's data, you die and leave your family with no income



(a) (b)

Utilities of Sequences

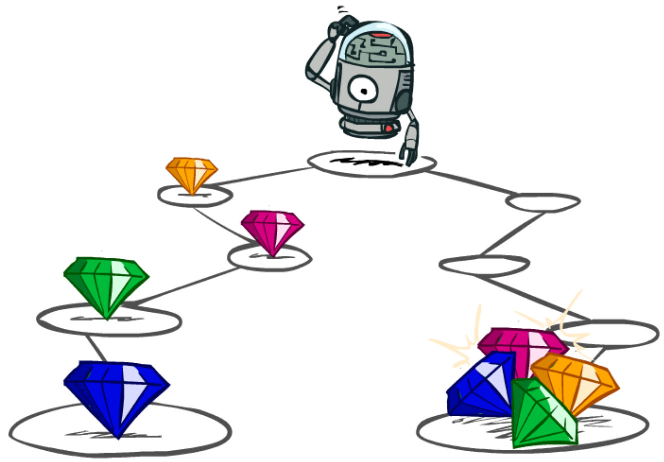


Utilities of Sequences

- 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***:

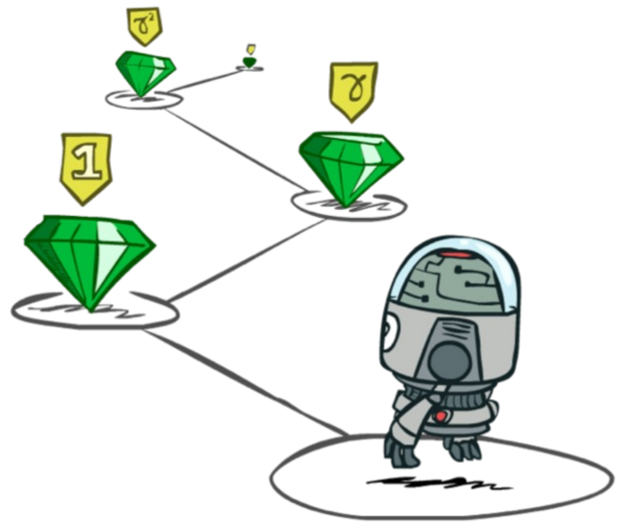
  $[a_1, a_2, \ldots] > [b_1, b_2, \ldots] \Leftrightarrow [c, a_1, a_2, \ldots] > [c, b_1, b_2, \ldots]$

  then there is only one way to define utilities:

  - ***Additive discounted utility***:

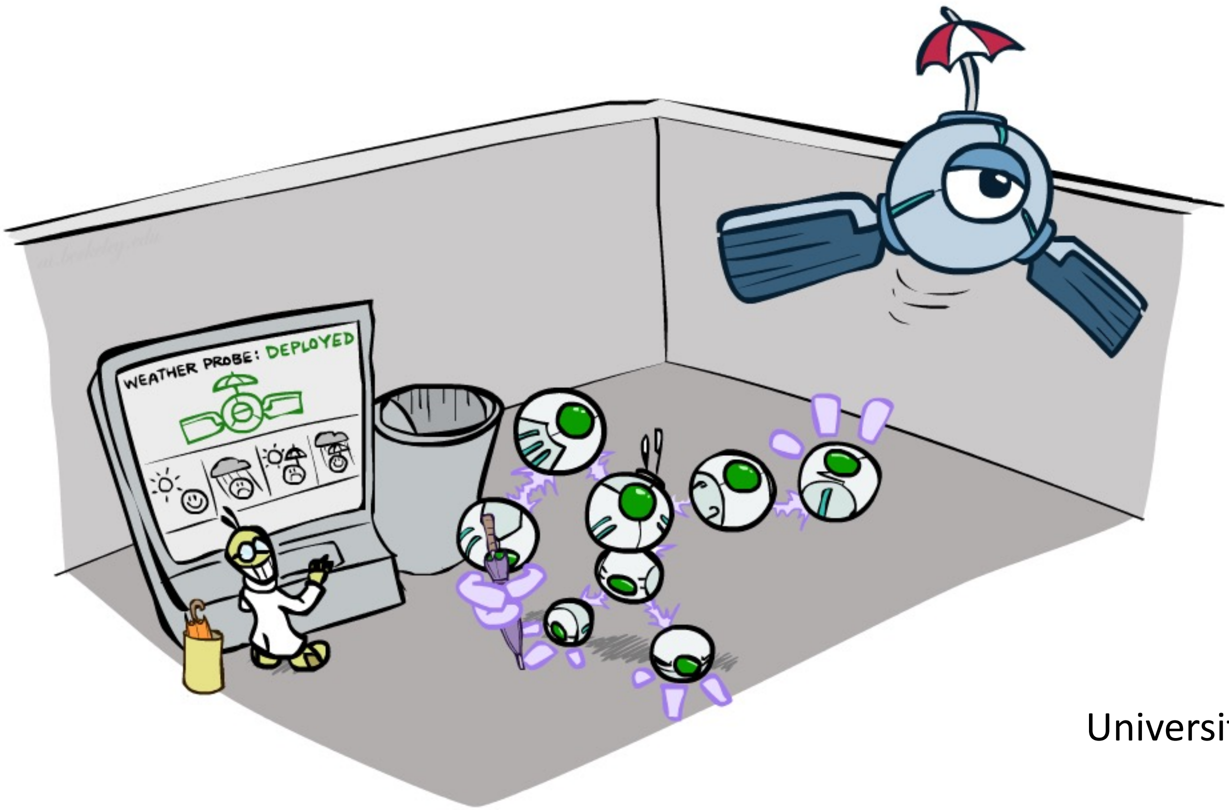
    $U([r_0, r_1, r_2, \ldots]) = r_0 + \gamma r_1 + \gamma^2 r_2 + \ldots$

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



CS 188: Artificial Intelligence

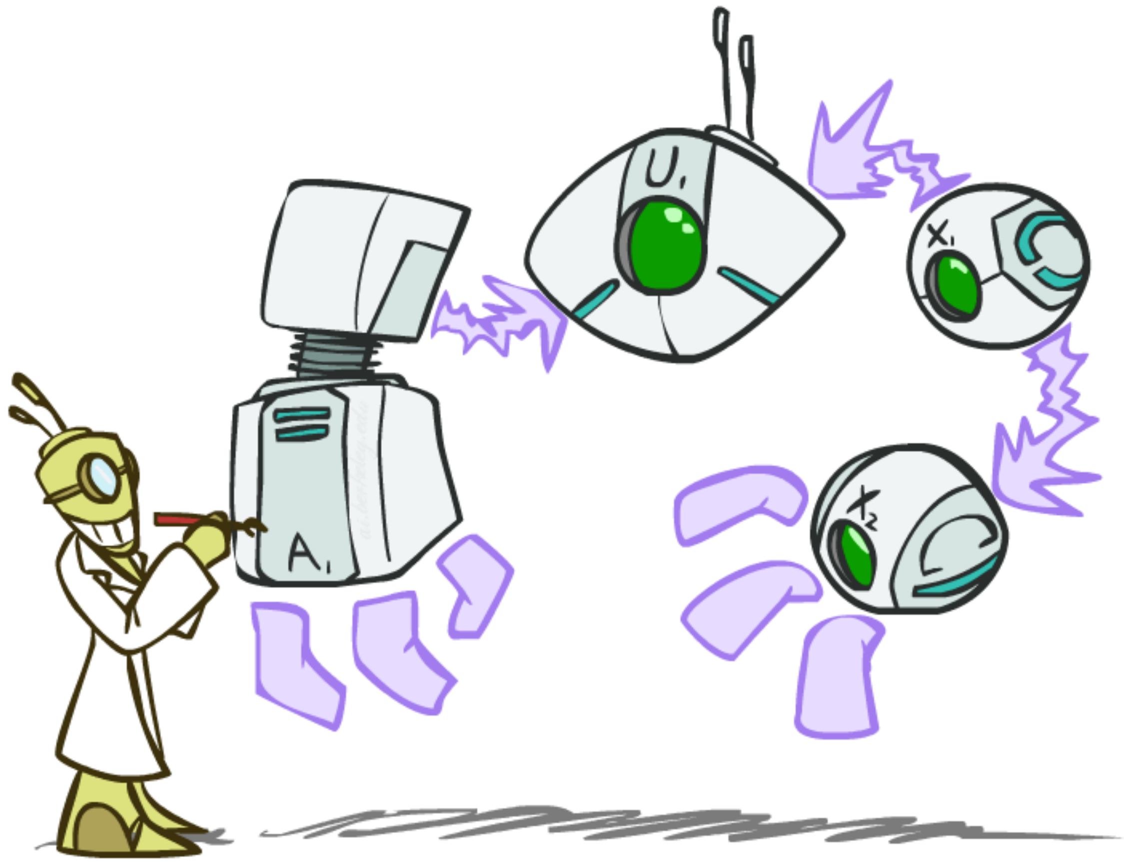
Decision Networks and Value of Information



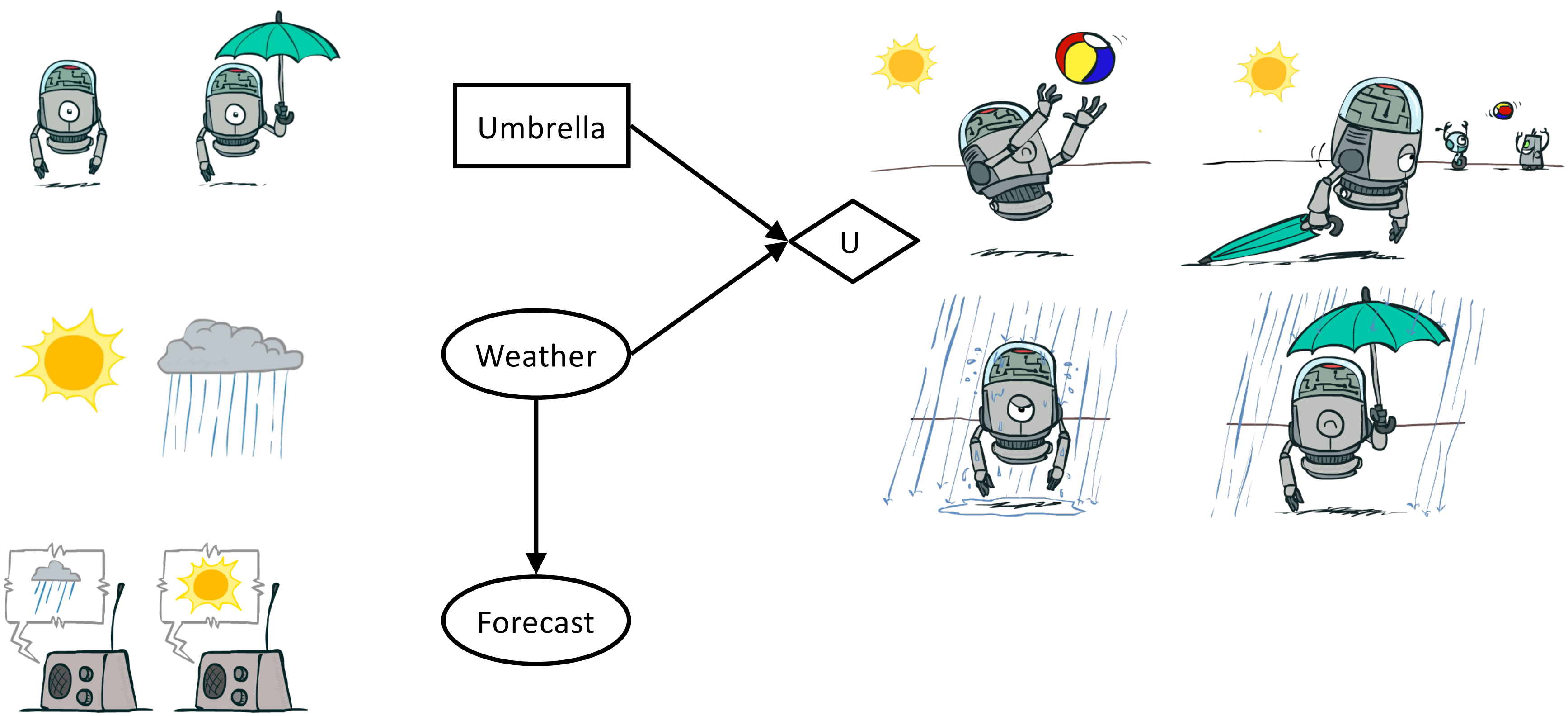
Spring 2024

University of California, Berkeley

Decision Networks



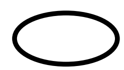
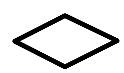
Decision Networks

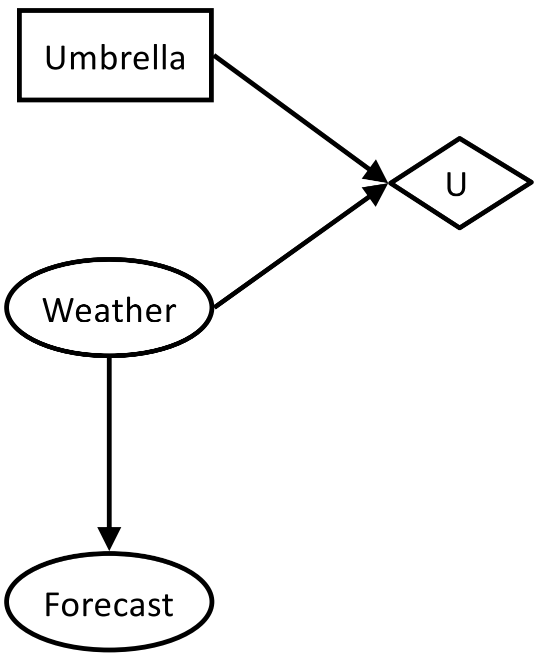


Decision Networks

- **MEU: choose the action which maximizes the expected utility given the evidence**

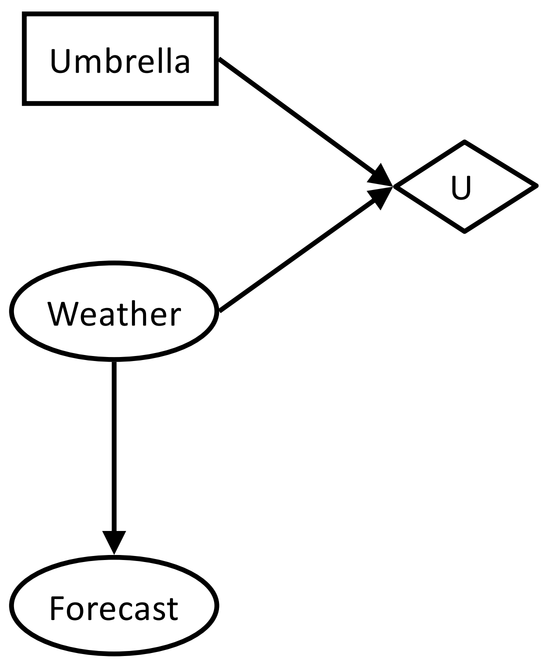
- Can directly operationalize this with decision networks
  - Bayes nets with nodes for utility and actions
  - Lets us calculate the expected utility for each action

- New node types:
  - Chance nodes (just like BNs) 
  - Actions (rectangles, cannot have parents, act as observed evidence) 
  - Utility node (diamond, depends on action and chance nodes) 



Decision Networks

- Action selection
  - Instantiate all evidence
  - Set action node(s) each possible way
  - Calculate posterior for all parents of utility node, given the evidence
  - Calculate expected utility for each action
  - Choose maximizing action



Decision Networks

Umbrella = leave

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

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

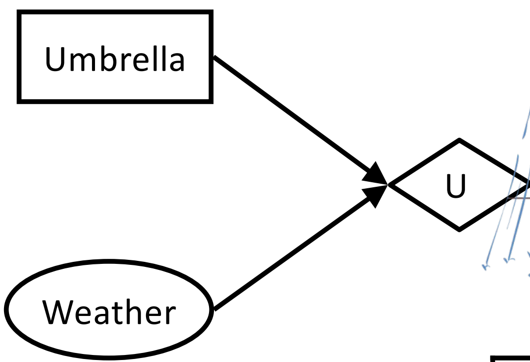
Umbrella = take

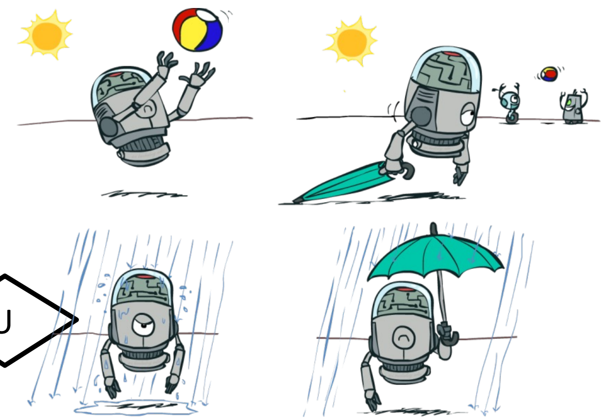
$$\mathrm{EU}(\mathrm{take}) = \sum_w P(w)U(\mathrm{take}, w)$$

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

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

Decision Networks: Notation

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

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

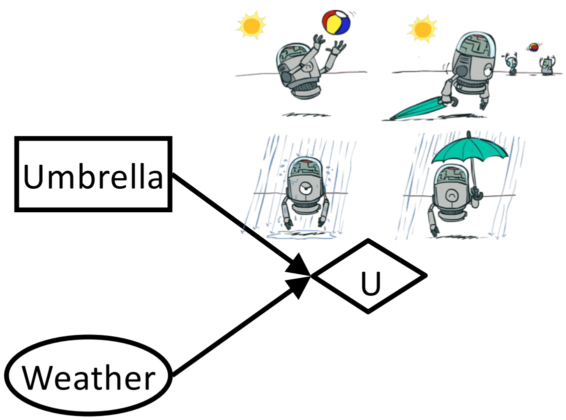
$$= 0.7 \cdot 20 + 0.3 \cdot 70 = 35$$

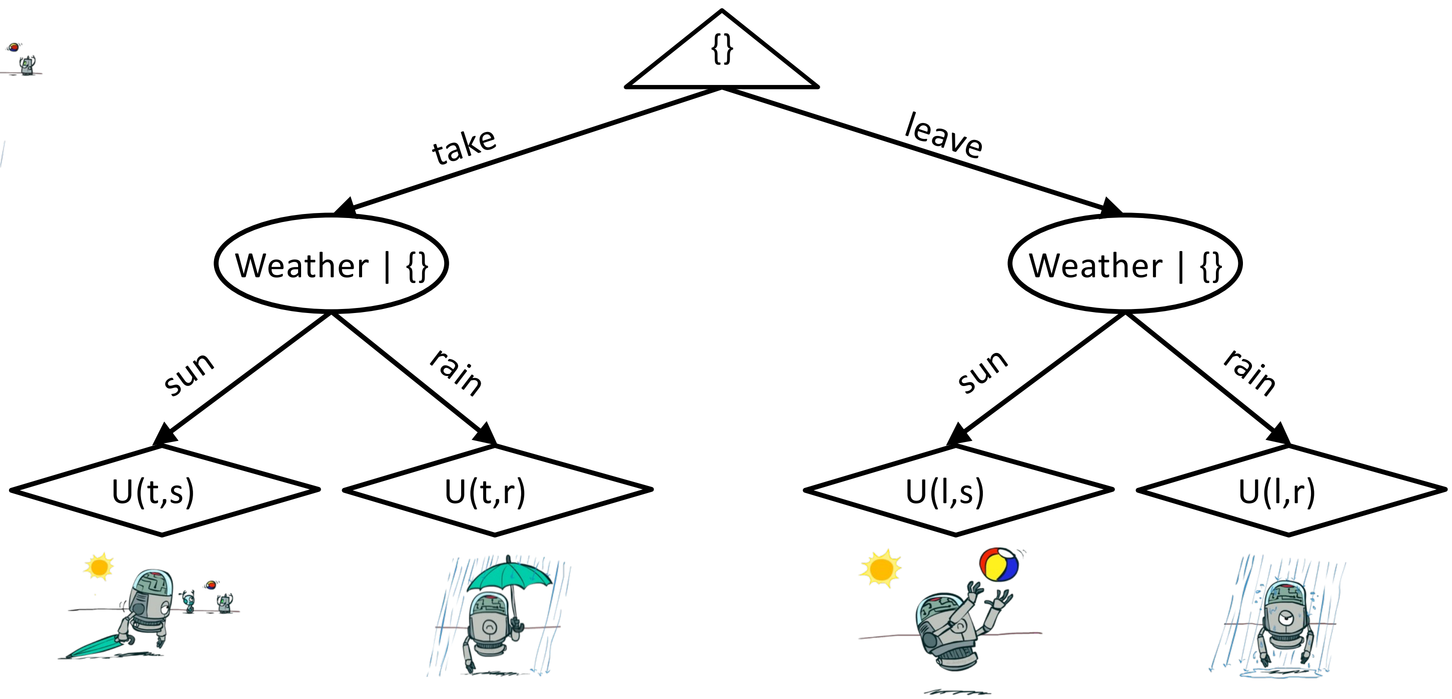
Optimal decision = leave

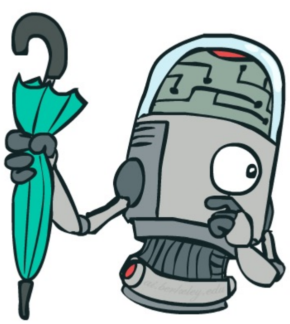
$$\mathrm{MEU}(\emptyset) = \max_a \mathrm{EU}(a) = 70$$

- EU(leave) = Expected Utility of taking action leave
  - In the parentheses, we write an action
  - Calculating EU requires taking an expectation over chance node outcomes
- MEU(ø) = Maximum Expected Utility, given no information
  - In the parentheses, we write the evidence (which nodes we know)
  - Calculating MEU requires taking a maximum over several expectations (one EU per action)

Decisions as Outcome Trees







- Almost exactly like expectimax
- What's changed?

Example: Decision Networks

Umbrella = leave

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

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

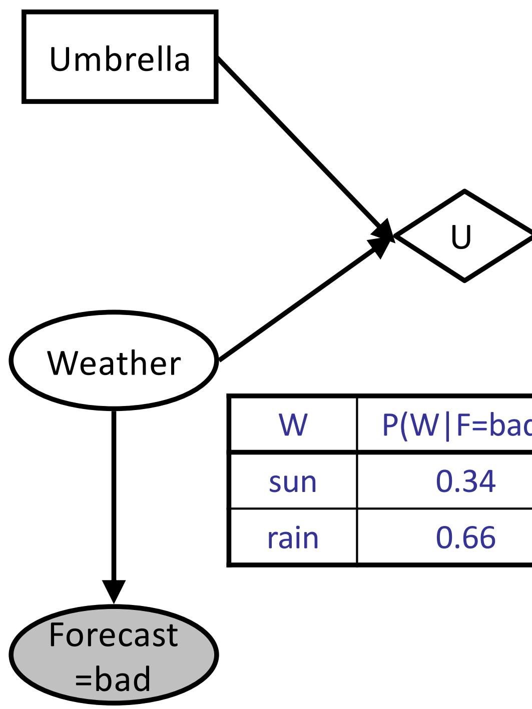
Umbrella = take

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

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

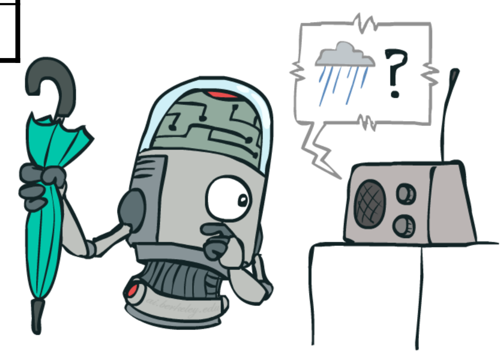
Optimal decision = take

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



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



Decision Networks: Notation

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

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

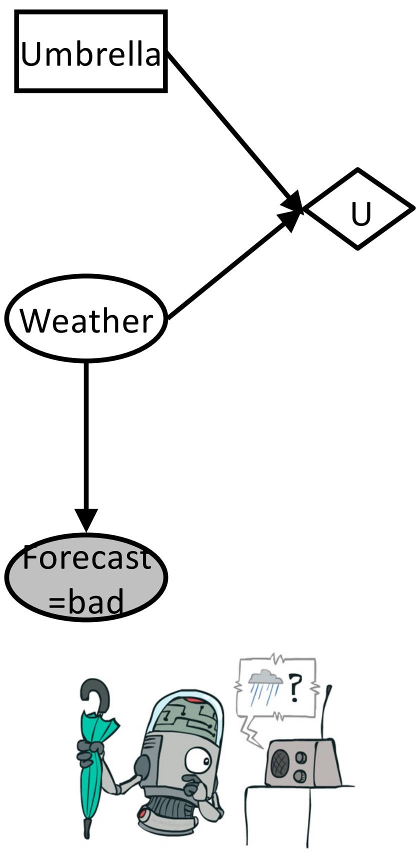
$$= 0.34 \cdot 20 + 0.66 \cdot 70 = 53$$

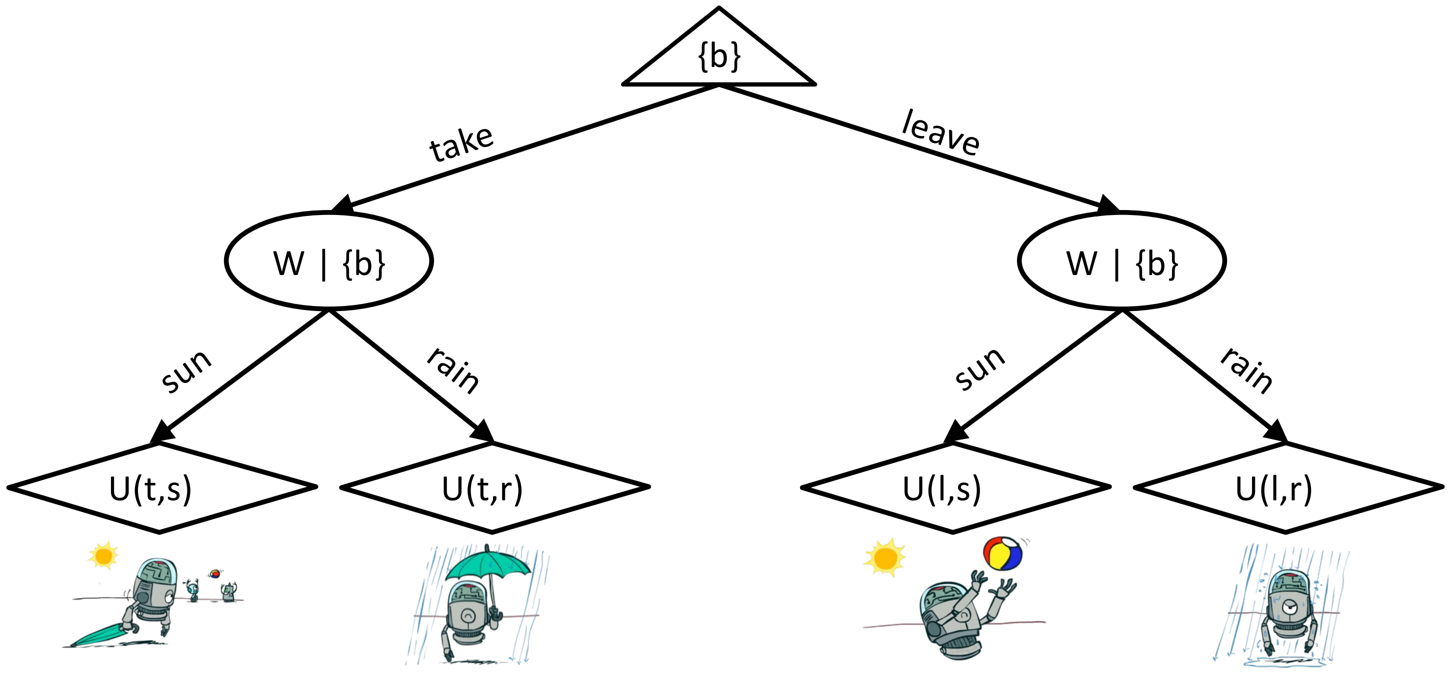
Optimal decision = take

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

- EU(leave|bad) = Expected Utility of choosing leave, given you know the forecast is bad
  - Left side of conditioning bar: Action being taken
  - Right side of conditioning bar: The random variable(s) we know the value of (evidence)
- MEU(F=bad) = Maximum Expected Utility, given you know the forecast is bad
  - In the parentheses, we write the evidence (which nodes we know)

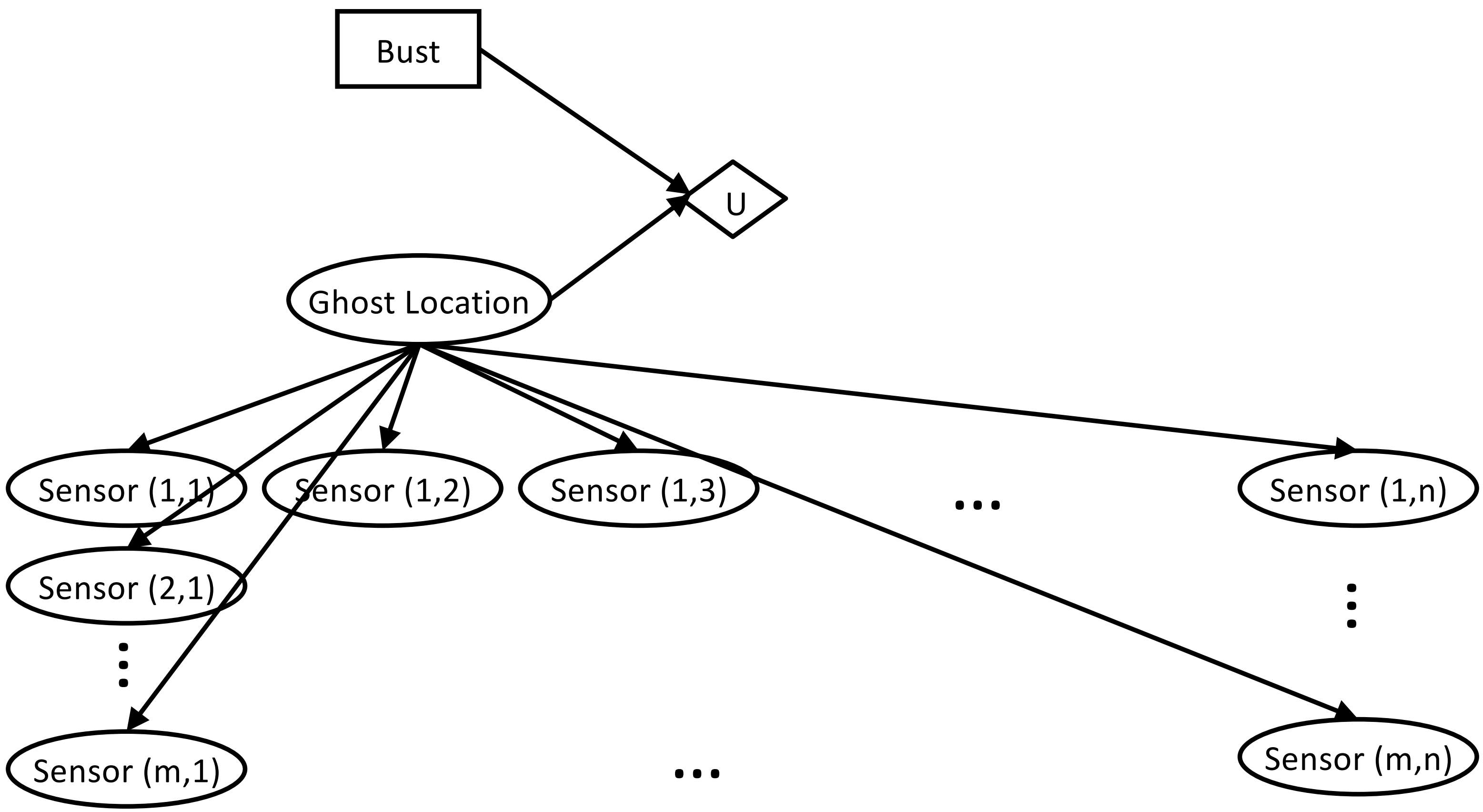
Decisions as Outcome Trees





Ghostbusters Decision Network

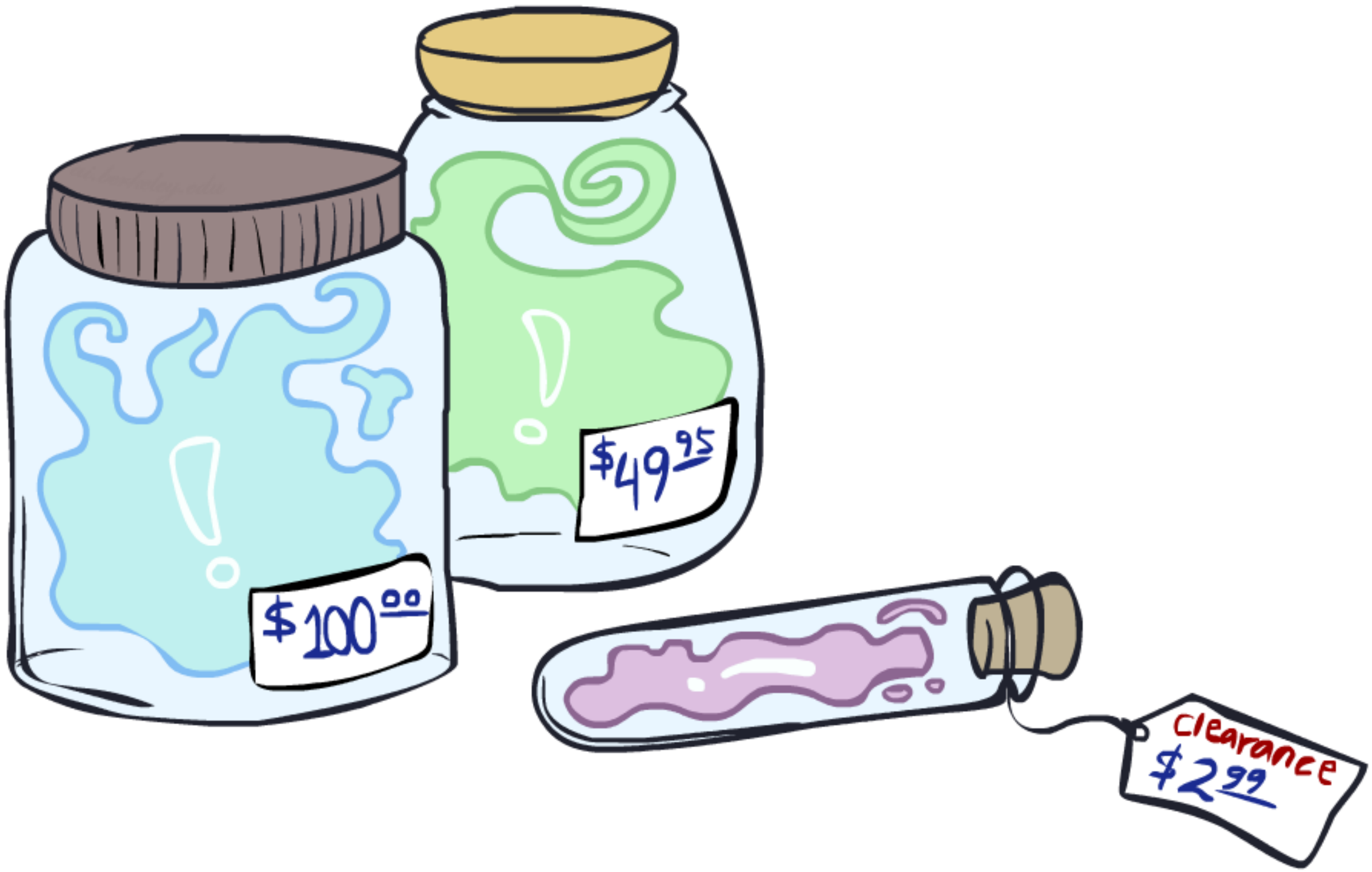
Demo: Ghostbusters with probability



Video of Demo Ghostbusters with Probability

- Game:
  - Costs 1 to make a measurement
  - Measurement gives noisy estimate of distance to ghost
  - When we blast, game is over
  - If we blast the ghost, we get utility of 250

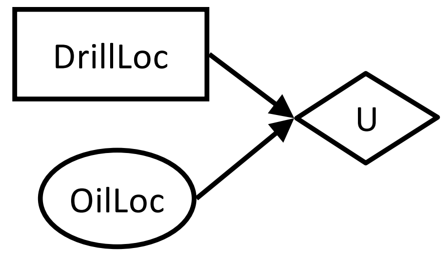
Value of Information



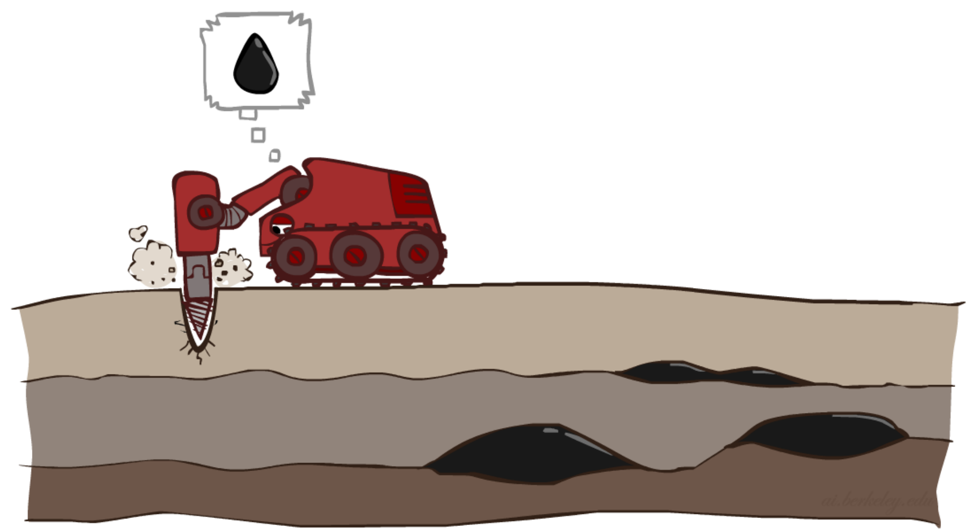
Value of Information

- 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
  - Survey may say "oil in a" or "oil in b"
  - 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

O	P
a	1/2
b	1/2



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



Value of Information Example: Weather

MEU with no evidence

$$\text{MEU}(\emptyset) = \max_a \text{EU}(a) = 70$$

MEU if forecast is bad

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

argmax_a is "take umbrella"

MEU if forecast is good

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

argmax_a is "leave umbrella"

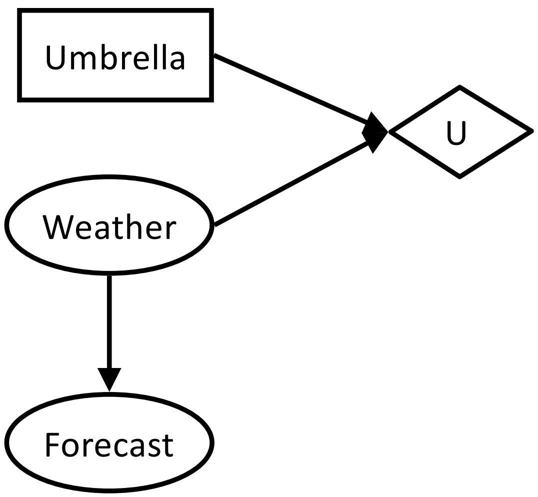
Forecast distribution

F	P(F)
good	0.59
bad	0.41

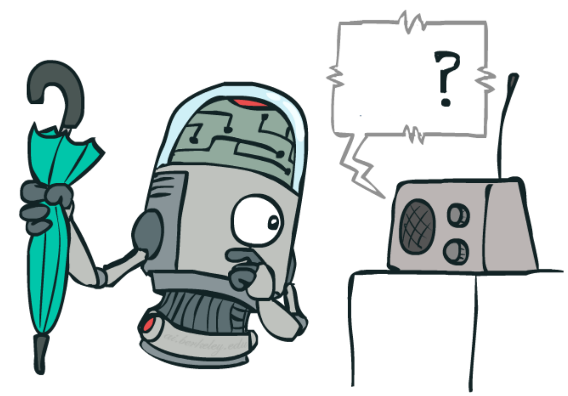
$$\text{MEU}(F) = 0.59 \cdot (95) + 0.41 \cdot (53) = 77.8$$

$$77.8 - 70 = 7.8$$

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



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



“Value of Perfect 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:

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

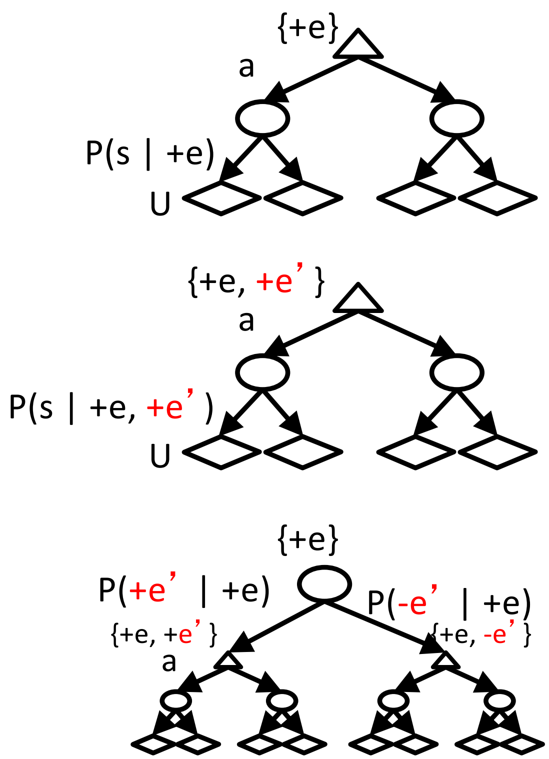
- BUT E’ is a random variable whose value is unknown, so we don’t know what e’ will be

- 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:

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



VPI: Notation

- MEU(e) = Maximum Expected Utility, given evidence E=e
  - In the parentheses, we write the evidence (which nodes we know)
  - Calculating MEU requires taking a maximum over several expectations (one EU per action)
- VPI(E'|e) = Expected gain in utility for knowing the value of E', given that I know the value of e so far
  - Left side of conditioning bar: The random variable(s) we want to know the value of revealing
  - Right side of conditioning bar: The random variable(s) we already know the value of
  - Calculating VPI requires taking an expectation over several MEUs (one MEU per possible outcome of E', because we don't know the value of E')

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

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

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

VPI: Computation Workflow

$$\text{MEU}(e) = \max_a \text{EU}(a|e)$$

$$\text{MEU}(e, e') = \max_a \text{EU}(a|e, e')$$ (calculate this for all values e' that E' could take)

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

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

Video of Demo Ghostbusters with VPI

- Game:
  - Costs 1 to make a measurement
  - Measurement gives noisy estimate of distance to ghost
  - When we blast, game is over
  - If we blast the ghost, we get utility of 250

VPI Properties

- Nonnegative

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

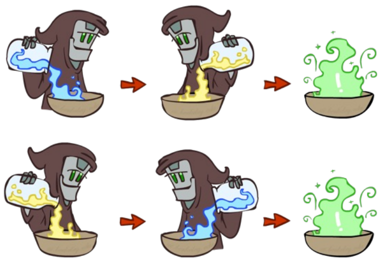
(Positive if different observed values of e' lead to different optimal decisions)

- Subadditive

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

(think of observing the same E_j twice)

- Order-independent



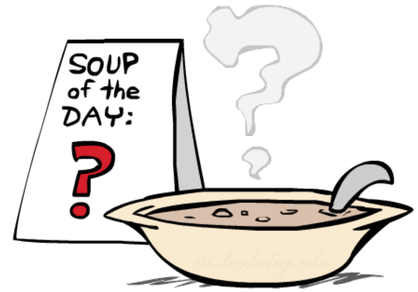
$$\text{VPI}(E_j, E_k|e) = \text{VPI}(E_j|e) + \text{VPI}(E_k|e, E_j)$$
$$= \text{VPI}(E_k|e) + \text{VPI}(E_j|e, E_k)$$

Value of information contd.

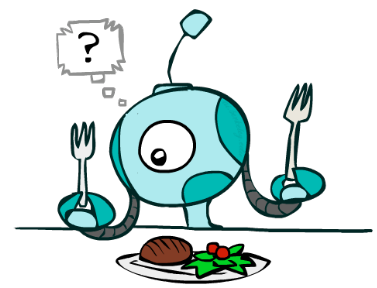
- General idea: value of information = ***expected improvement in decision quality*** from observing value of a variable
  - E.g., oil company deciding on seismic exploration and test drilling
  - E.g., doctor deciding whether to order a blood test
  - E.g., person deciding on whether to look before crossing the road
- ***Decision network contains everything needed to compute it!***
- $\mathrm{VPI}(E_i \mid e) = \left[\, \sum_{e_i} P(e_i \mid e) \max_a \mathrm{EU}(a \mid e_i, e) \,\right] - \max_a \mathrm{EU}(a \mid e)$

Quick VPI Questions

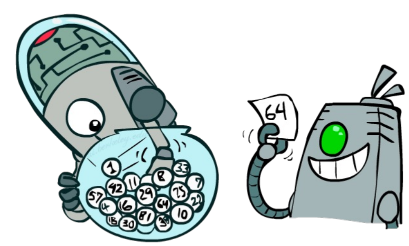
- The soup of the day is either clam chowder or split pea, but you wouldn't order either one. What's the value of knowing which it is?



- 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?

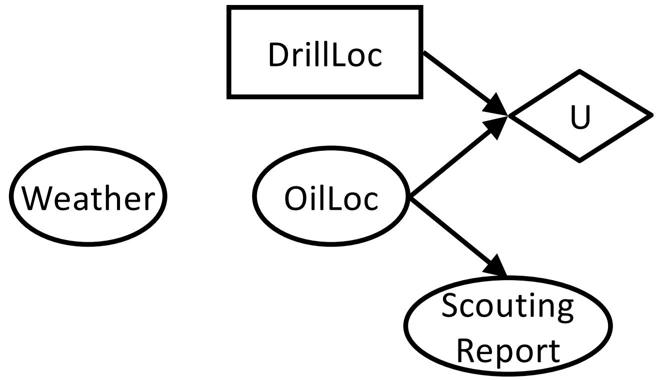
- No such thing (as we formulate it)
- Information corresponds to the observation of a node in the decision network
- If data is “noisy” that just means we don’t observe the original variable, but another variable which is a noisy version of the original one

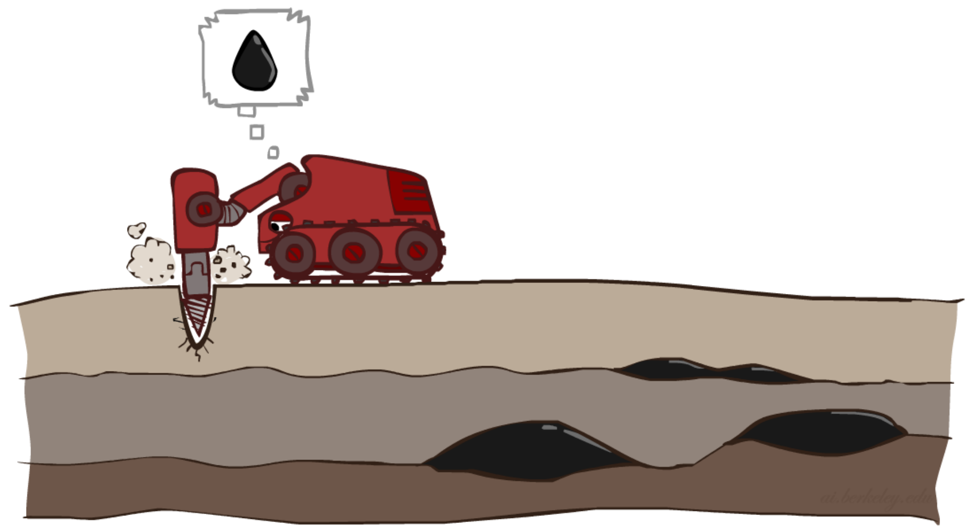
VPI Question

- VPI(ScoutingReport) ?
- VPI(OilLoc) ?
- VPI(Weather) ?
- VPI(OilLoc | ScoutingReport) vs VPI(ScoutingReport | OilLoc) ?
- Generally:

  VPI(Z | CurrentEvidence) = 0

  if $\quad$ Parents(U) $\perp\!\!\!\perp$ Z $\;|\;$ CurrentEvidence





Bonus slide (if time)

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: