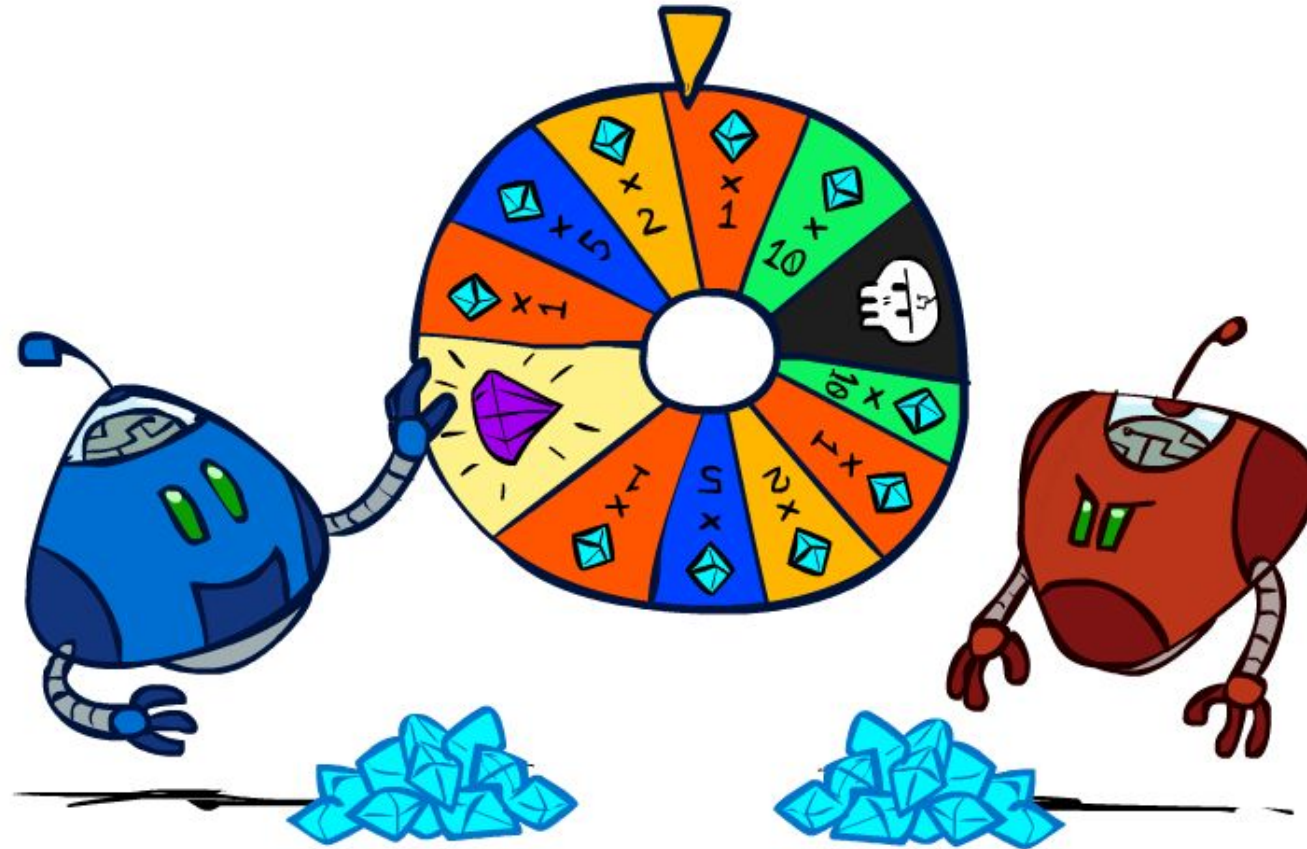


CS 188: Artificial Intelligence

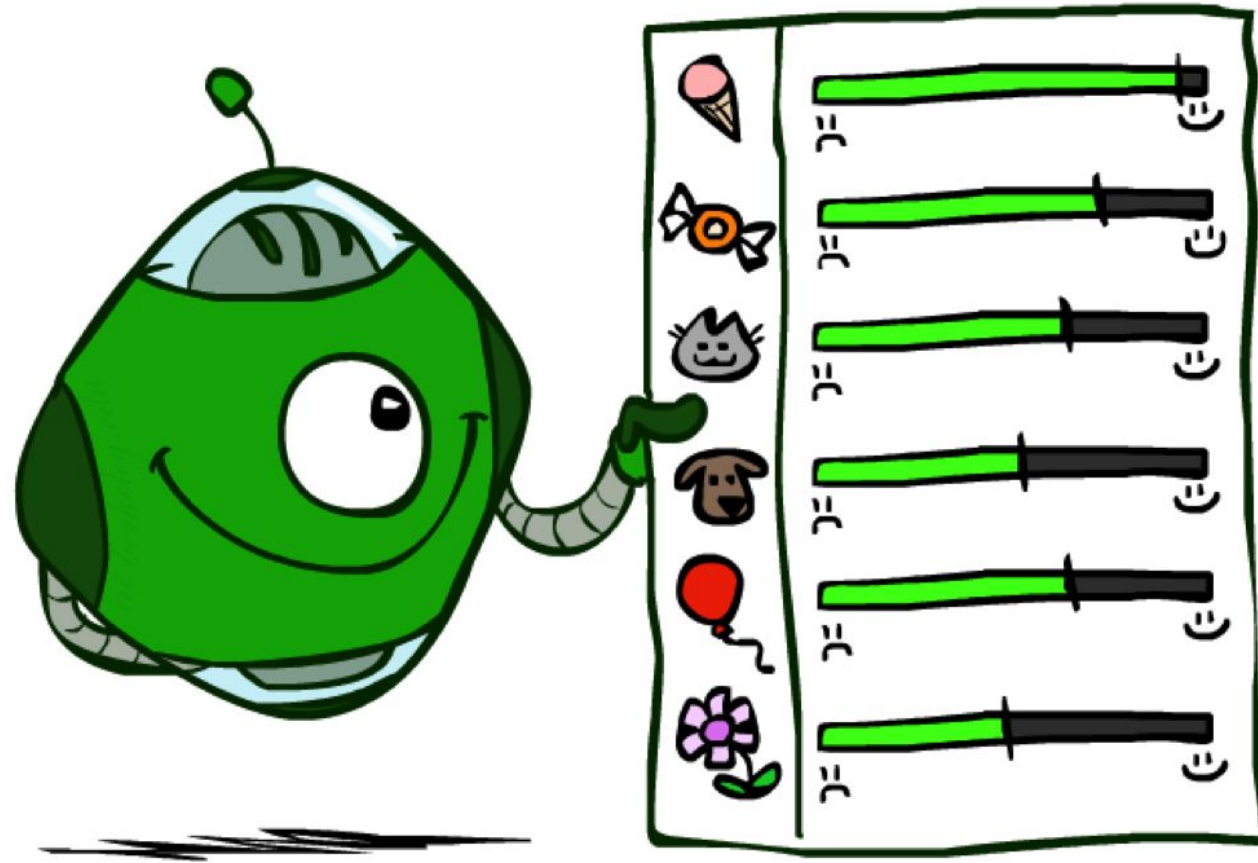
Rational Decisions



Instructor: Dan Klein and Stuart Russell

University of California, Berkeley

Utilities

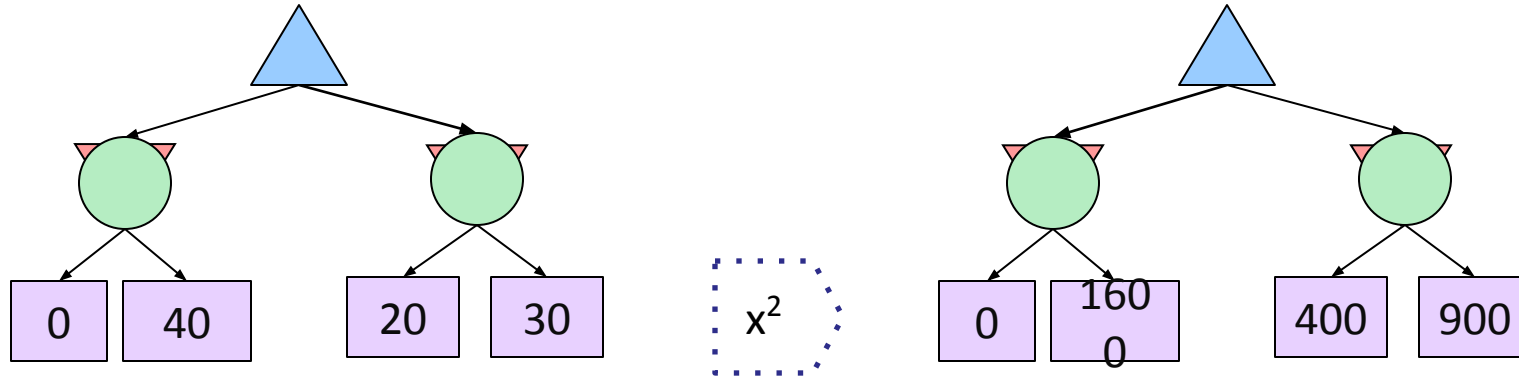


Maximum Expected Utility

- Principle of maximum expected utility:
 - A rational agent should choose 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 even makes sense?
 - What if our behavior (preferences) can't be described by utilities?



The need for numbers



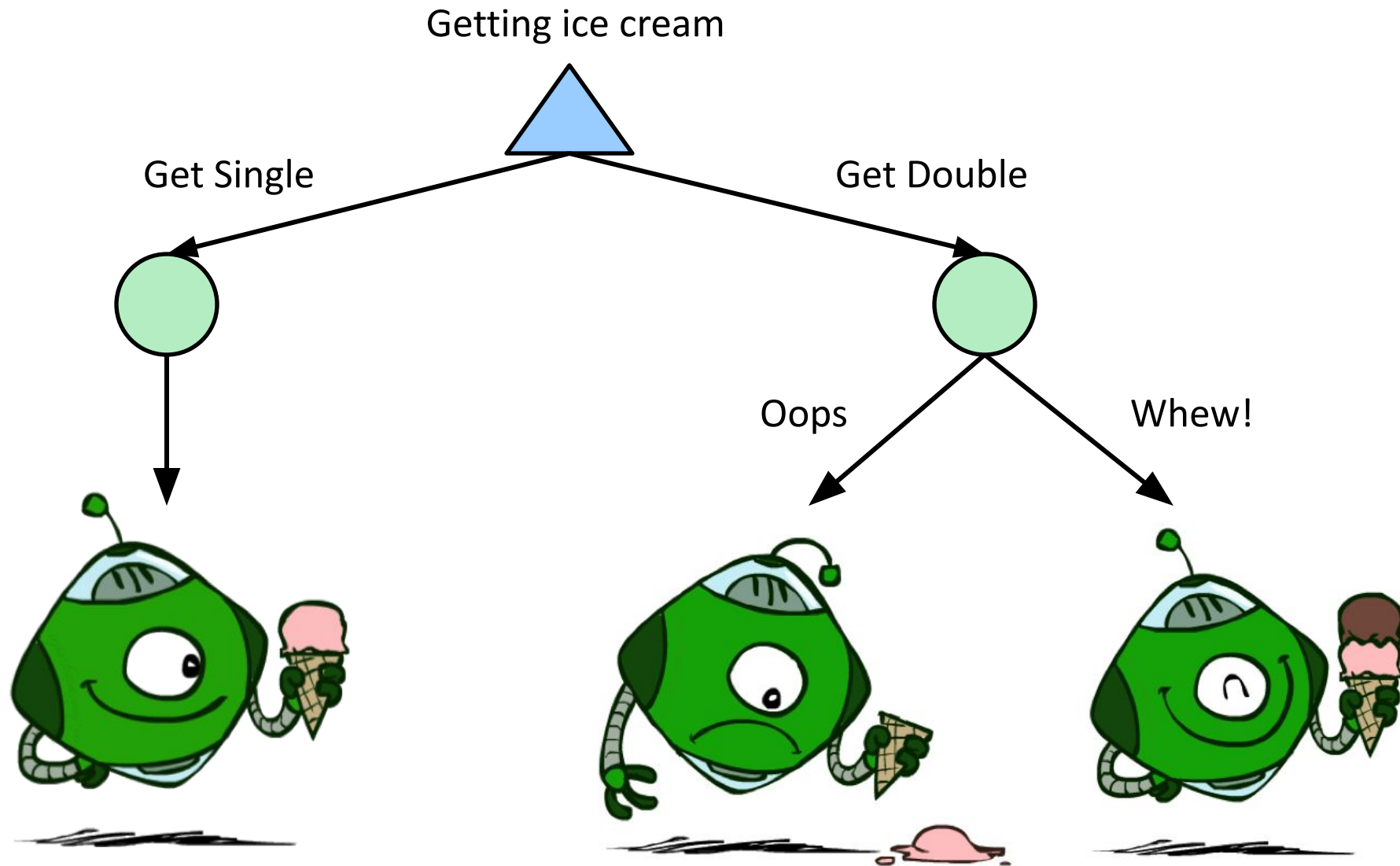
- For worst-case minimax reasoning, terminal value scale doesn't matter
 - We just want better states to have higher evaluations (get the ordering right)
 - The optimal decision is invariant under any **monotonic transformation**
 - The terminal values are really **ordinal utilities** (i.e., ranks)
-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

Utilities

- Utilities are functions from outcomes (possible complete futures) to real numbers that describe an agent's preferences
- Where do utilities come from?
 - In a game, may be simple (+1/-1)
 - Utilities summarize "what the agent wants" (preference rankings over futures)
 - Theorem: any "rational" preferences can be summarized as a utility function
- "Standard model": specify 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:

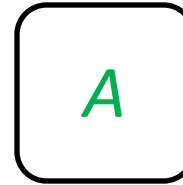
- Prizes: A , B , etc.
- Lotteries: situations with uncertain prizes

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

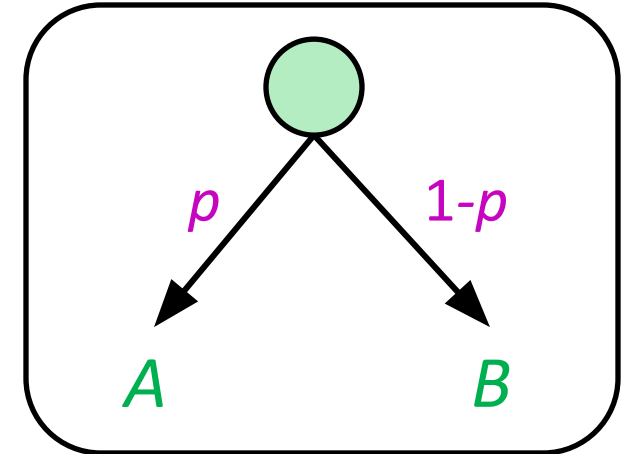
■ Notation:

- Preference: $A \succ B$
- Indifference: $A \sim B$
- Weakly prefers: $A \succeq B$

A Prize



A Lottery

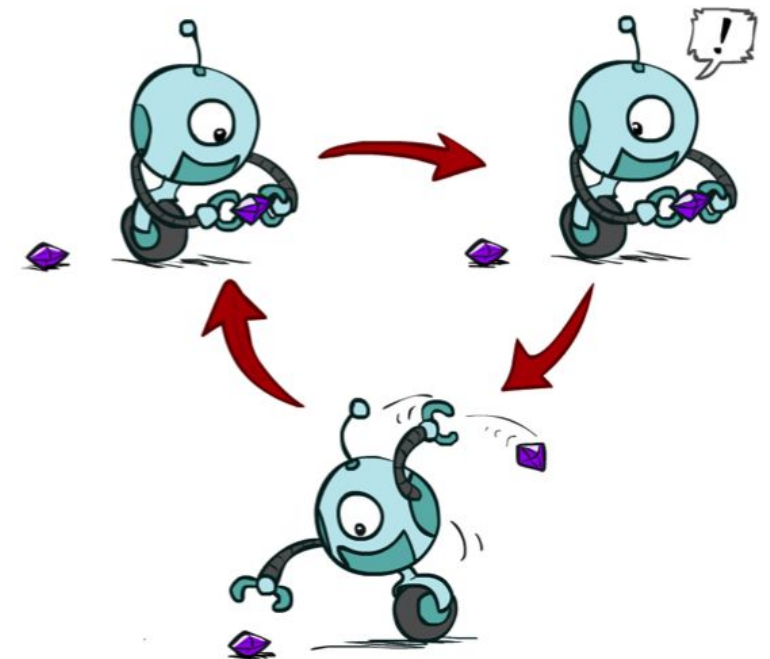


Rational Preferences

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

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

- 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:

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

Independence:

$$(A \sim B) \Leftrightarrow [p, A; 1-p, C] \sim [p, B; 1-p, C]$$



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

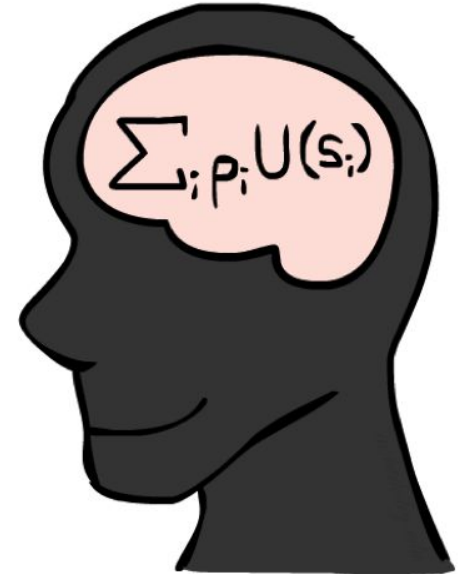
MEU Principle

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

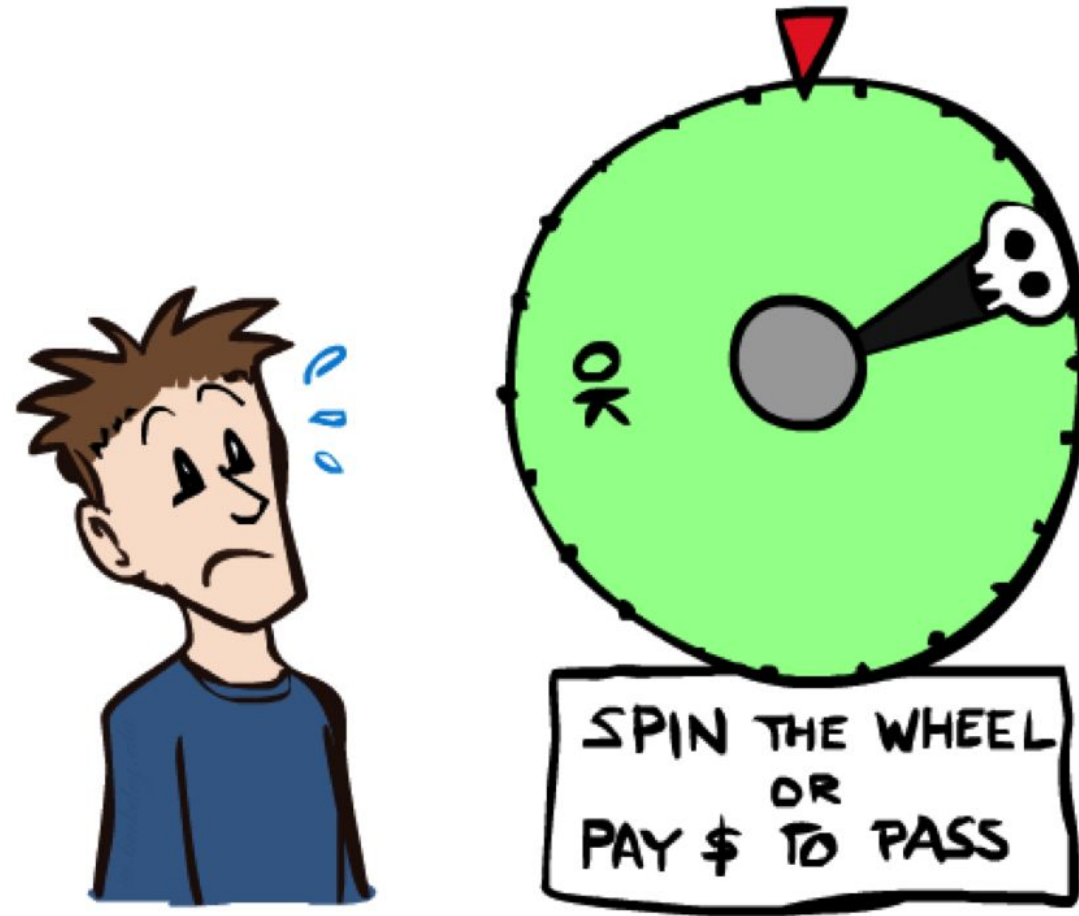
$$U(A) \geq U(B) \Leftrightarrow A \succeq B$$

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

- I.e. values assigned by U preserve preferences of both prizes and lotteries!
 - Optimal policy invariant under **positive affine transformation** $U' = aU + b, a > 0$
- Note: rationality does **not** require representing or manipulating utilities and probabilities
 - E.g., a lookup table for perfect tic-tac-toe

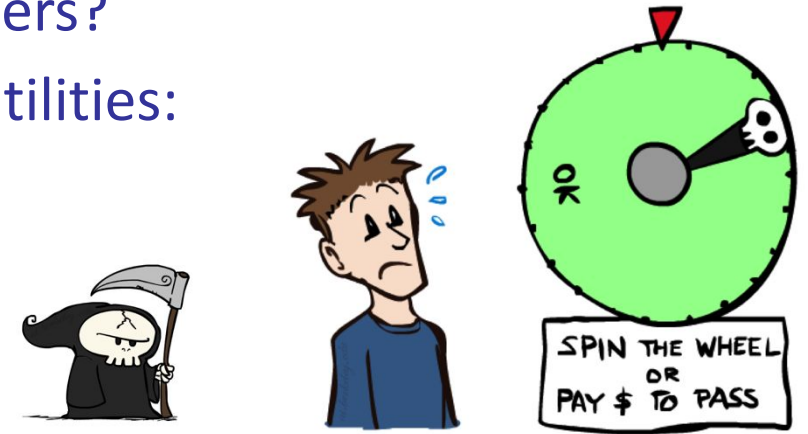


Human Utilities



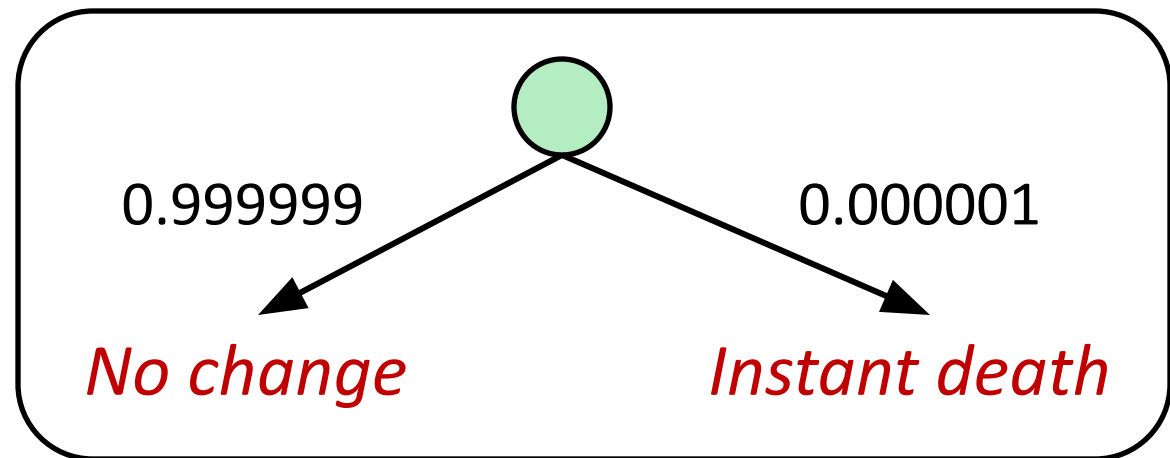
Human Utilities

- Utilities map states/futures 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_T with probability p
 - “worst possible catastrophe” u_L with probability $1-p$
 - Adjust lottery probability p until indifference: $A \sim L_p$
 - Resulting p is a utility in $[0,1]$



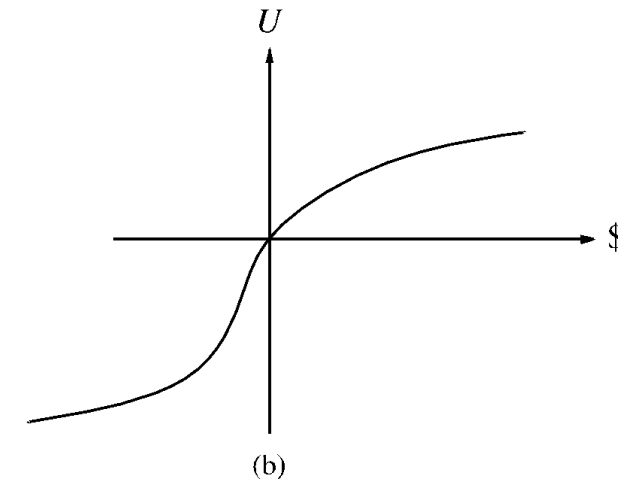
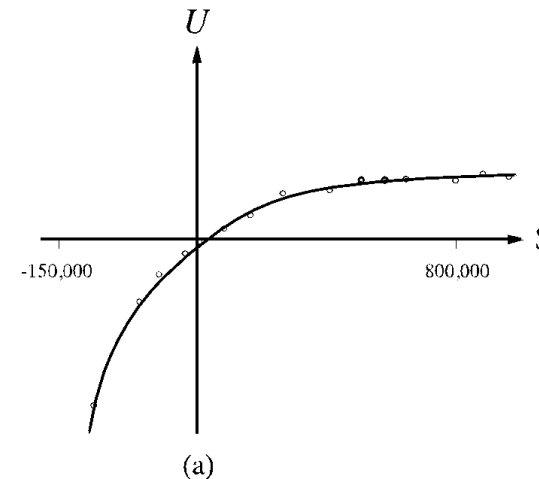
Pay \$50

~



Money

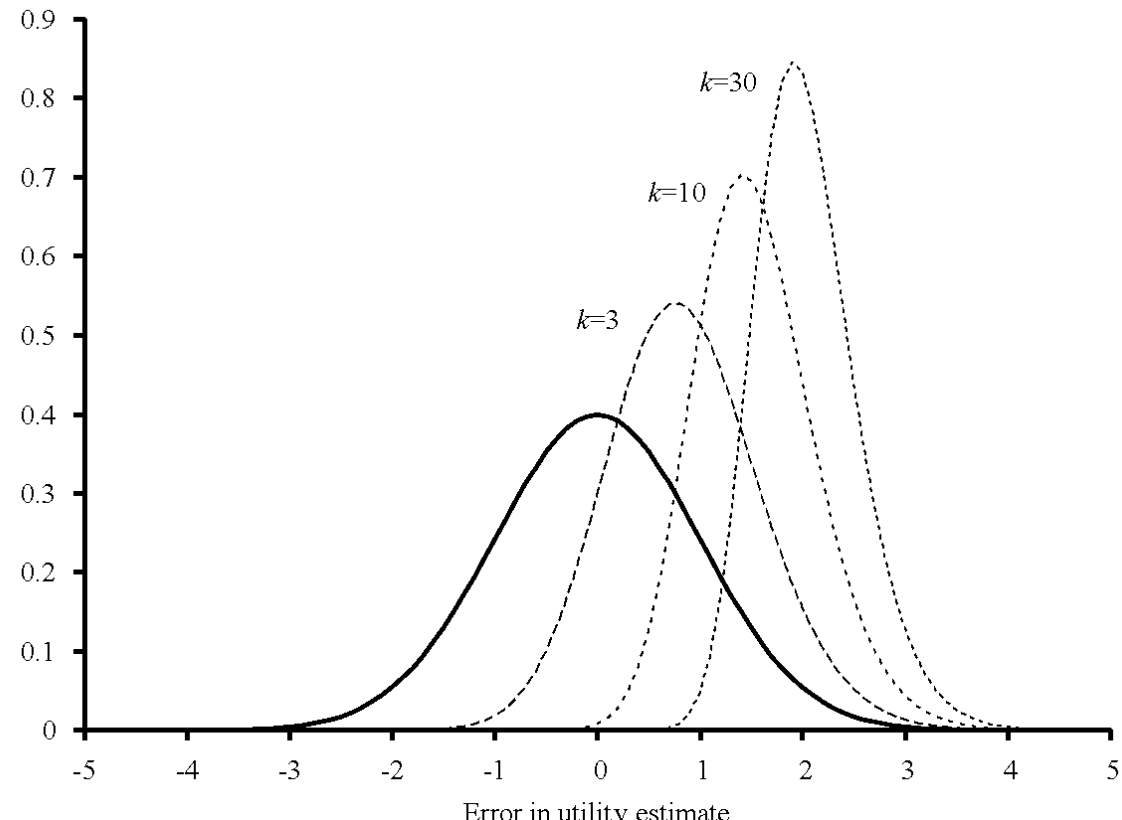
- Money **does not** behave as a utility function, but we can talk about the utility of having money (or being in debt)
- Given a lottery $L = [p, \$X; (1-p), \$Y]$
 - The **expected monetary value** $EMV(L) = pX + (1-p)Y$
 - The utility is $U(L) = pU(\$X) + (1-p)U(\$Y)$
 - Typically, $U(L) < U(EMV(L))$
 - 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 $CE(L)$ is the cash amount such that $CE(L) \sim L$
 - The **insurance premium** is $EMV(L) - CE(L)$
 - If people were risk-neutral, this would be zero!



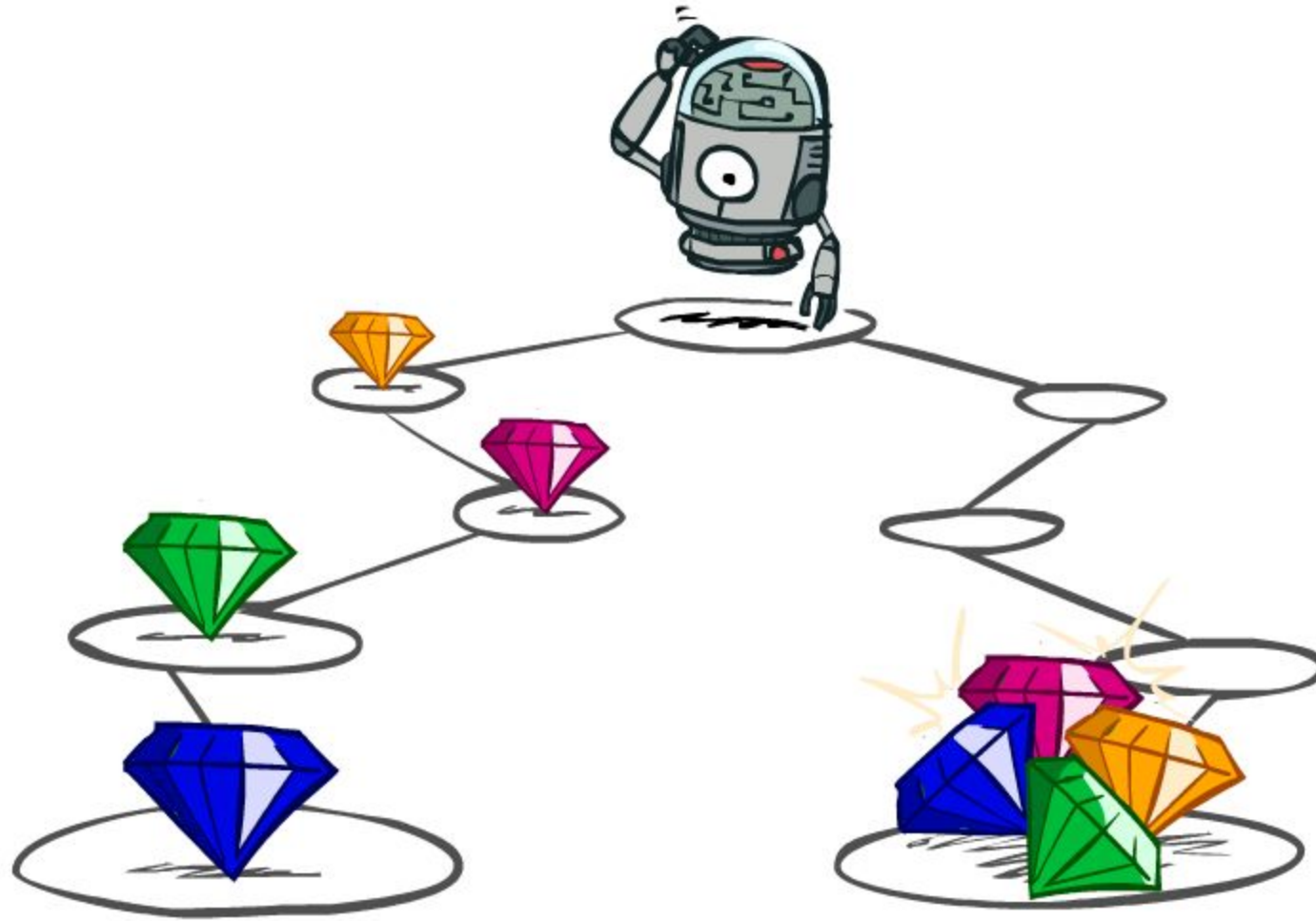
Post-decision Disappointment: the Optimizer's Curse

- 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, \dots, 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 $\sim N(0,1)$;
Max of k estimates:

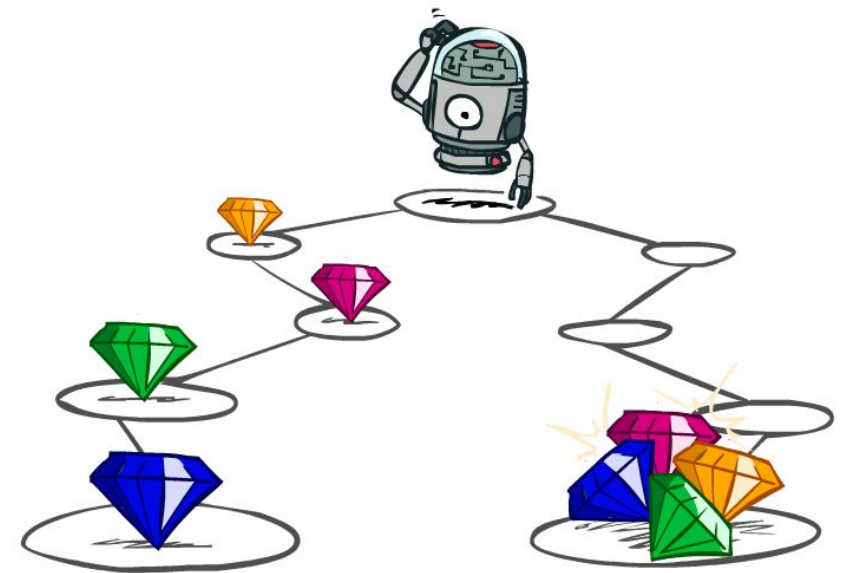


Utilities of Sequences



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**:

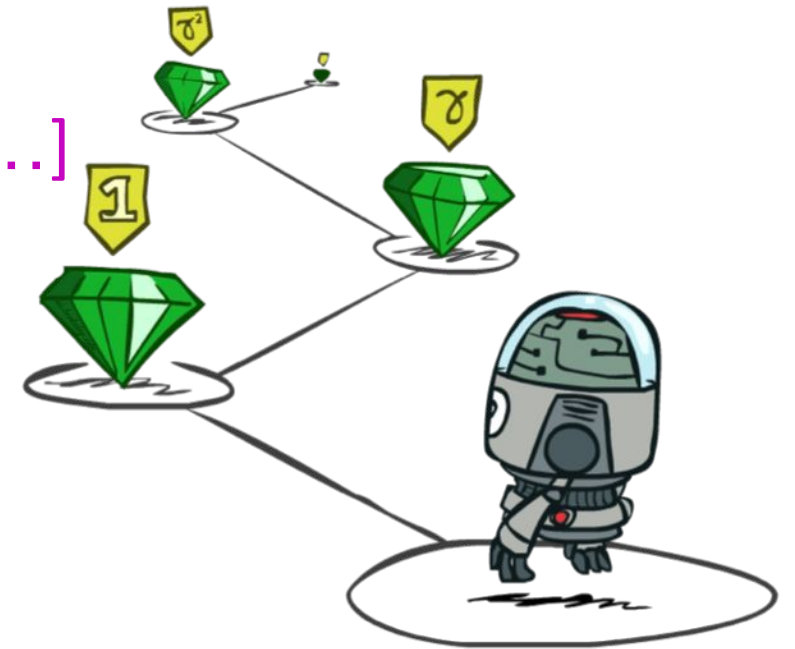
$$[a_1, a_2, \dots] > [b_1, b_2, \dots] \Leftrightarrow [c, a_1, a_2, \dots] > [c, b_1, b_2, \dots]$$

then there is only one way to define utilities:

- **Additive discounted utility**:

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

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



Invariance for sequences

- Invariance for utilities (reminder):

Optimal policy invariant under positive affine transformation $U' = aU + b, a > 0$

- Invariance for rewards:

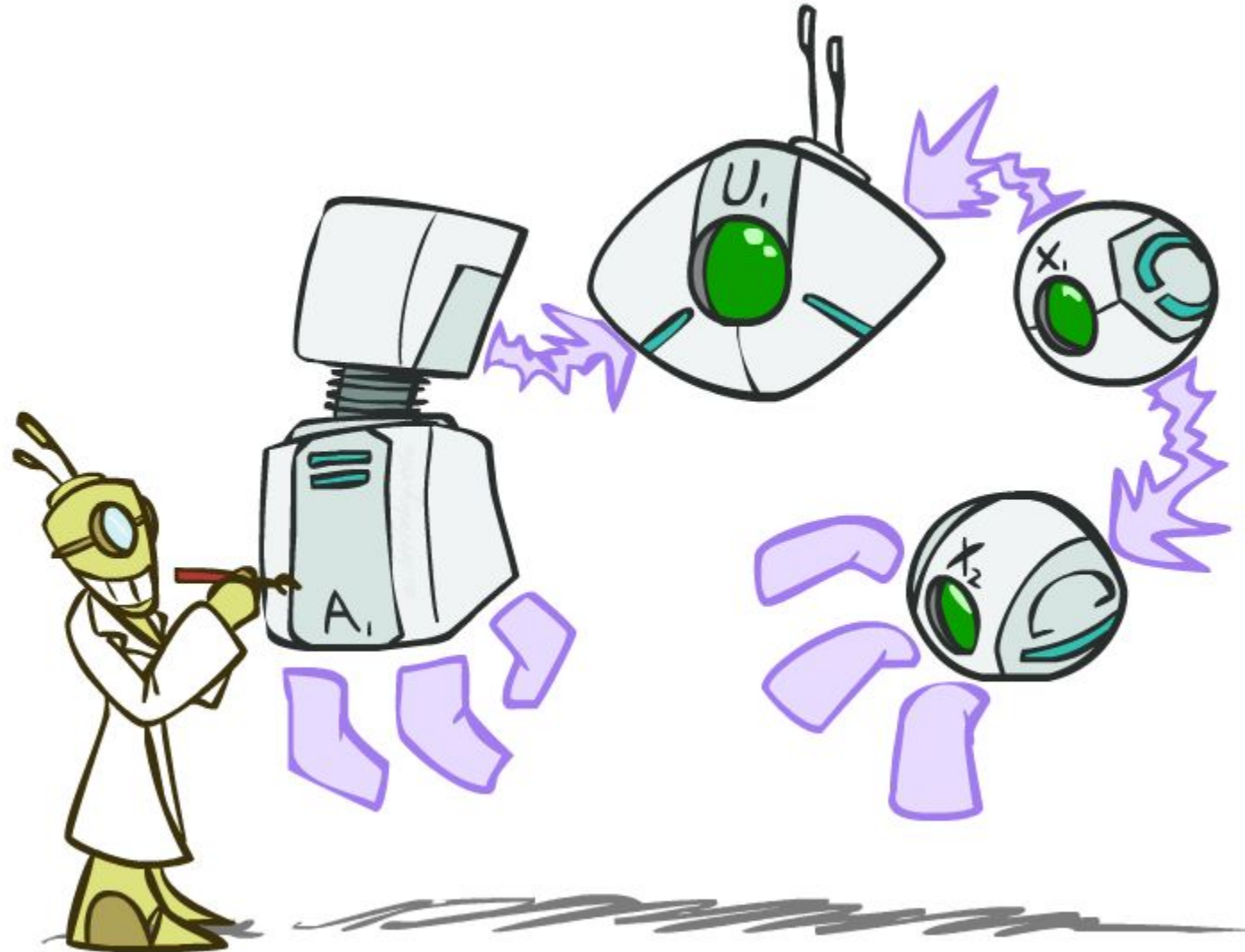
Optimal policy is also invariant under potential transformation:

$$R'(s, a, s') = R(s, a, s') + \gamma \Phi(s') - \Phi(s)$$

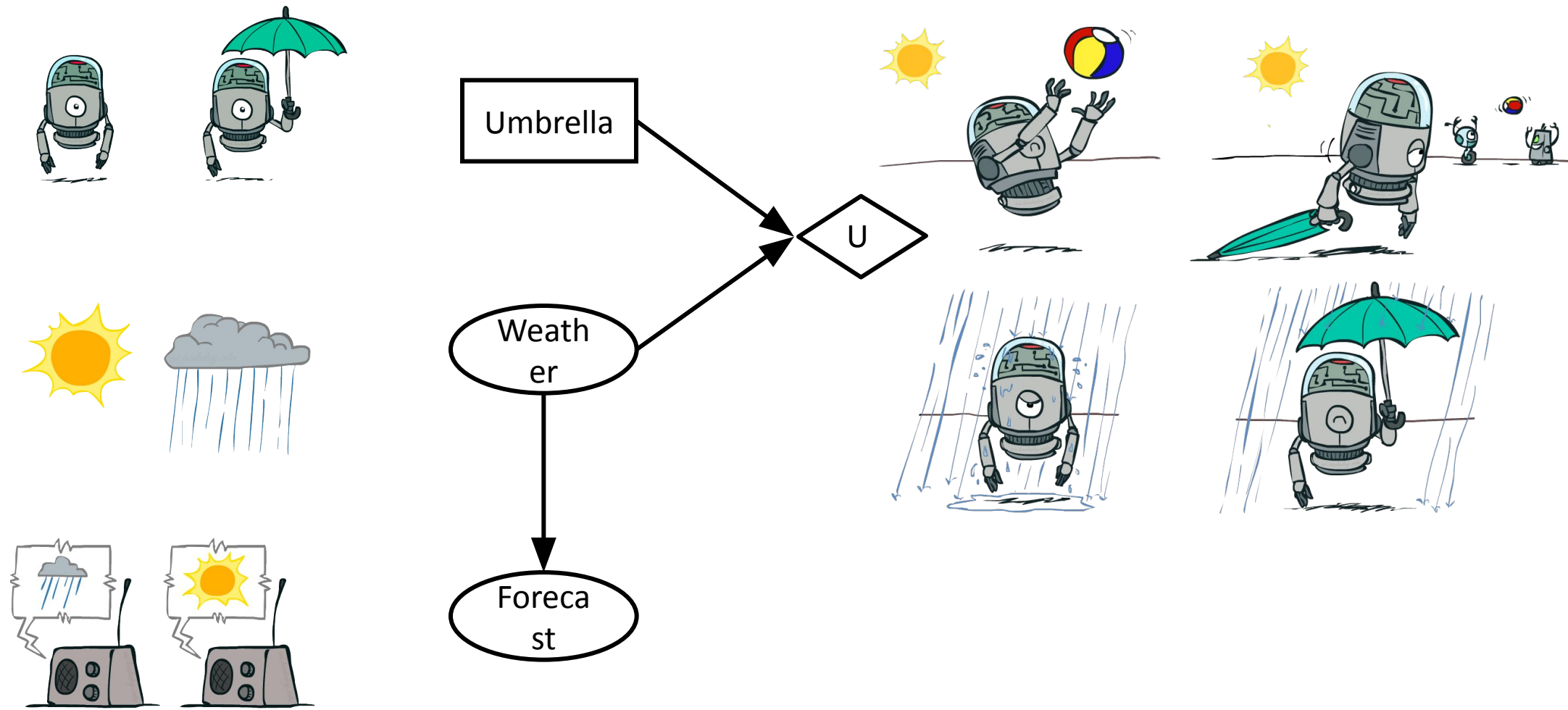
where Φ is **any function of state**

- These **shaping rewards** can massively speed up RL
- Soccer example: $R(s, a, s') = +3$ for a win, $+1$ for a draw, 0 for a loss
 $\Phi(s) = (100 * \text{goal difference}) + (10/\text{distance to goal}) + 0.1(\text{possession})$

Decision Networks



Decision Networks



Decision Networks

- Decision network = Bayes net + Actions + Utilities



- Action nodes** (rectangles, cannot have parents, will have value fixed by algorithm)

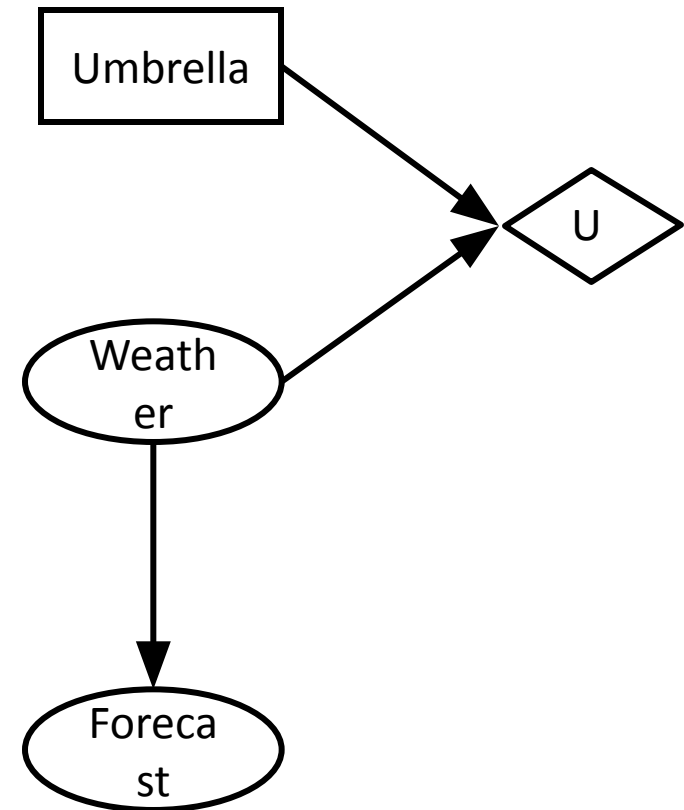


- Utility nodes** (diamond, depends on action and chance nodes)

- Decision algorithm:

- Fix evidence e
- For each possible action a
 - Fix action node to a
 - Compute posterior $P(W|e,a)$ for parents W of U
 - Compute expected utility $\sum_w P(w|e,a) U(a,w)$
- Return action with highest expected utility

Bayes net inference!



Example: Take an umbrella?

- Decision algorithm:
 - Fix evidence e
 - For each possible action a
 - Fix action node to a
 - Compute posterior $P(W|e,a)$ for parents W of U
 - Compute expected utility of action a : $\sum_w P(w|e,a) U(a,w)$
 - 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

Umbrella = leave

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

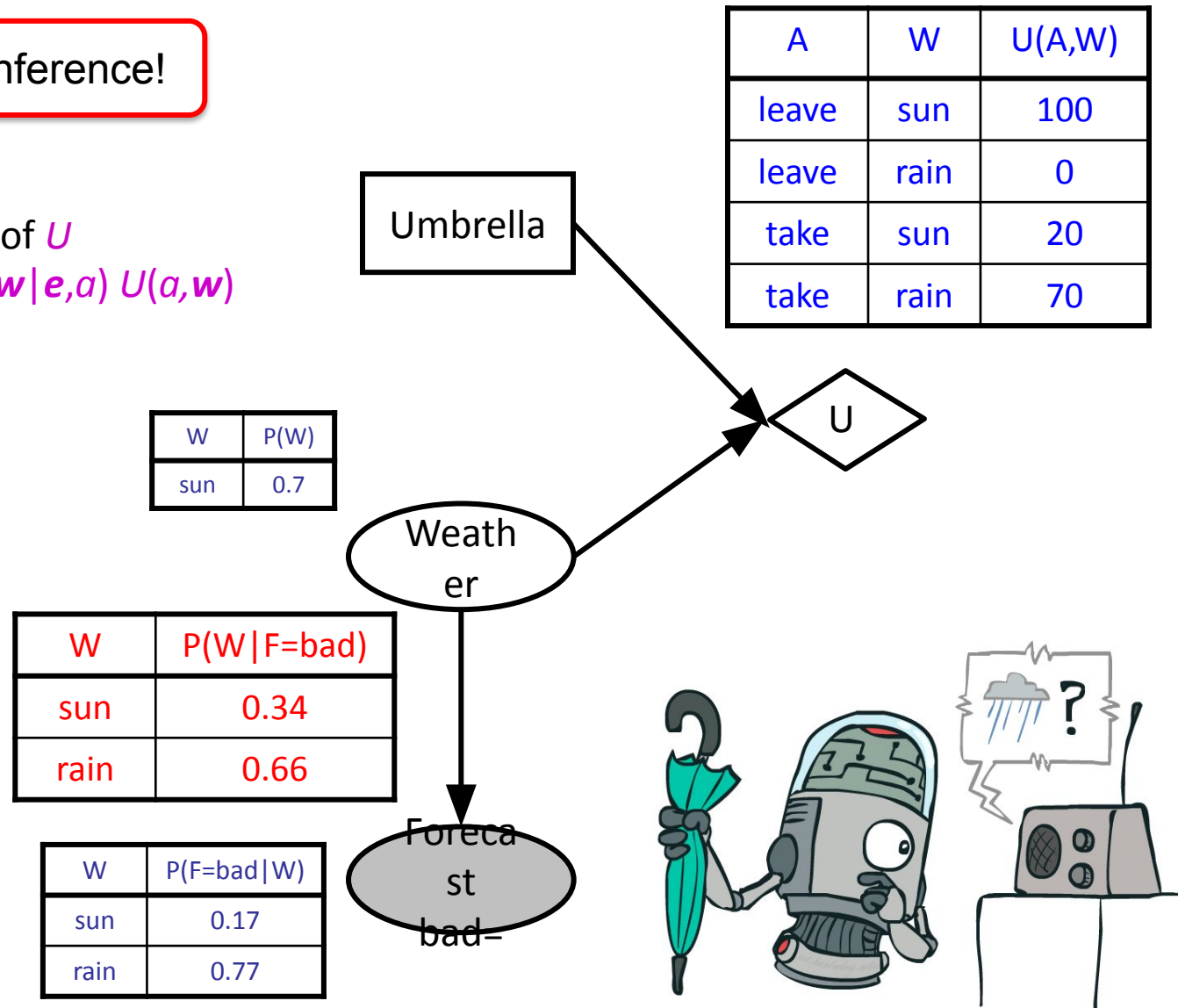
$$= 0.34 \times 100 + 0.66 \times 0 = 34$$

Umbrella = take

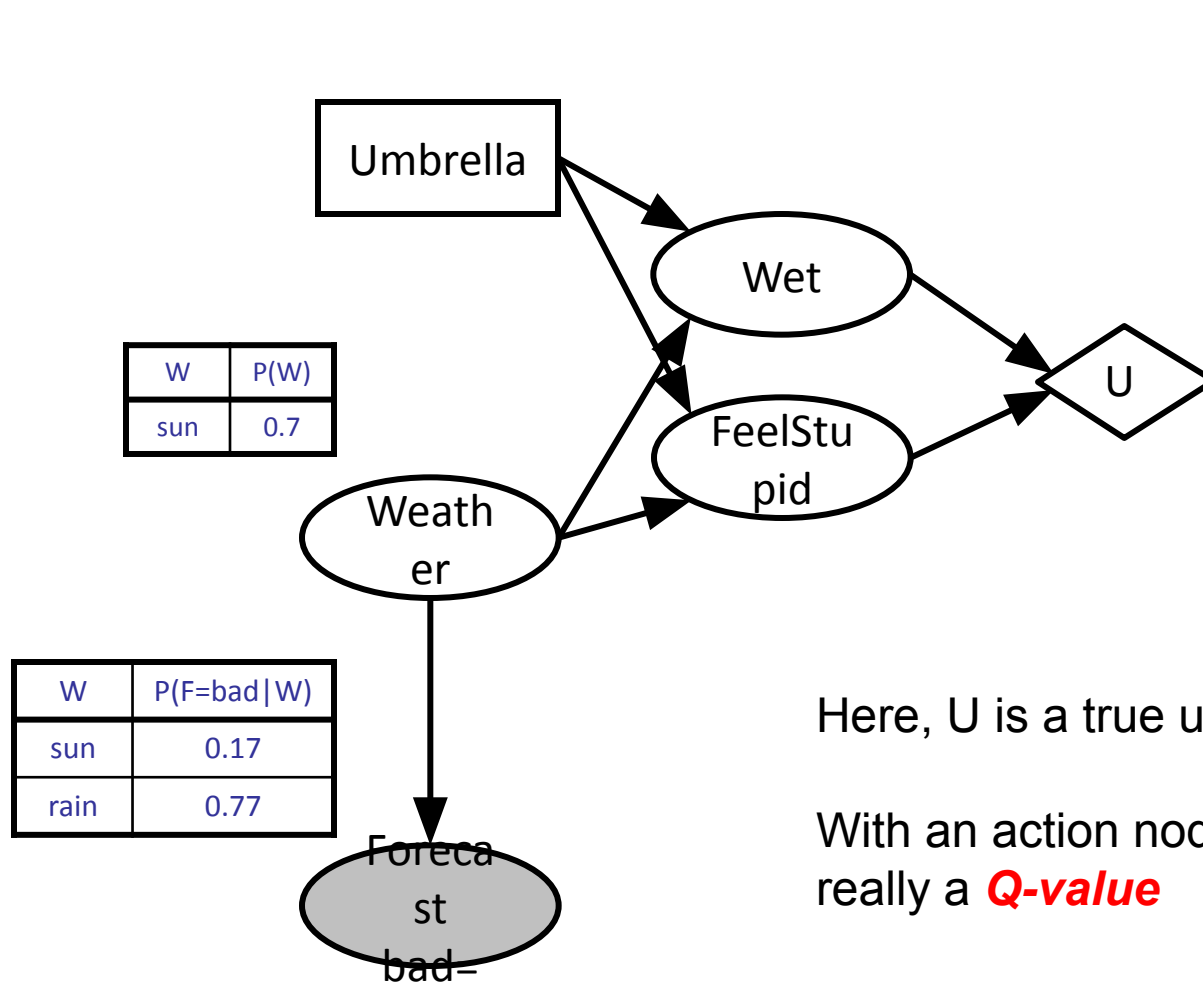
$$EU(\text{take}|F=\text{bad}) = \sum_w P(w|F=\text{bad}) U(\text{take},w)$$

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

Optimal decision = take!



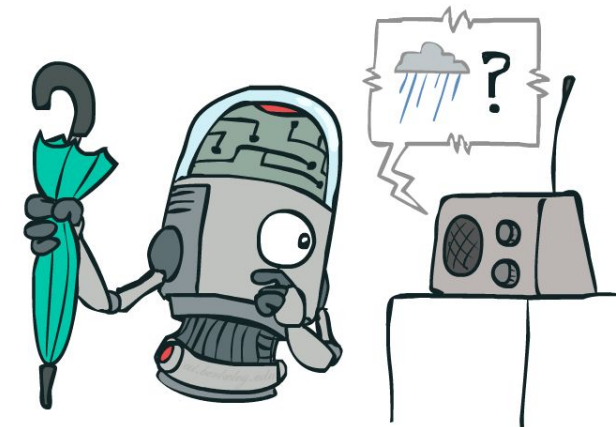
Decision network with utilities on outcome states



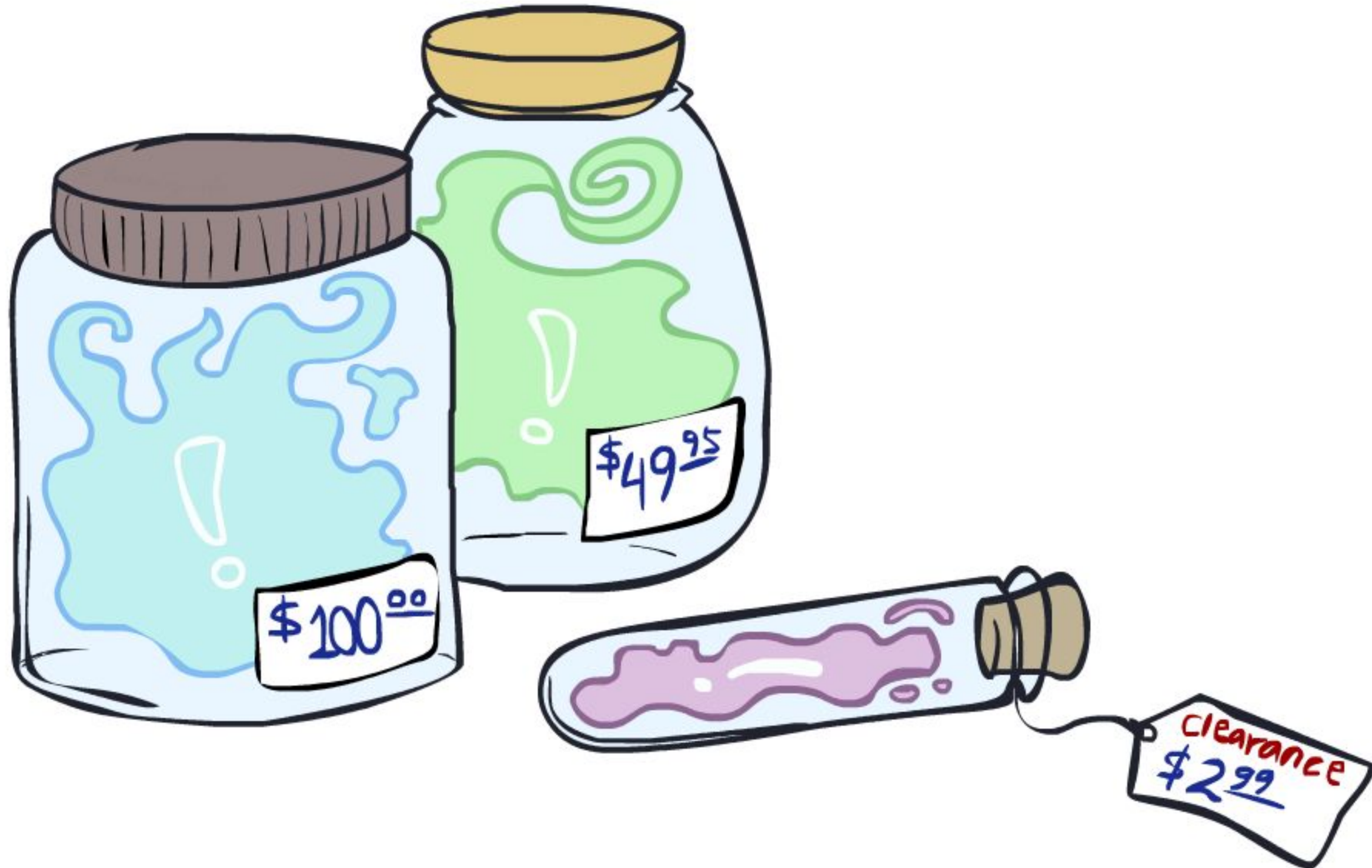
Wet	FeelStupid	U
dry	false	100
wet	true	0
dry	true	20
damp	false	70

Here, U is a true utility.

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



Value of Information



Value of information

- Suppose you haven't yet seen the forecast

- $EU(\text{leave} \mid) = 0.7 \times 100 + 0.3 \times 0 = 70$
 - $EU(\text{take} \mid) = 0.7 \times 20 + 0.3 \times 70 = 35$

- What if you look at the forecast?**

- If Forecast=good

- $EU(\text{leave} \mid F=\text{good}) = 0.89 \times 100 + 0.11 \times 0 = 89$
 - $EU(\text{take} \mid F=\text{good}) = 0.89 \times 20 + 0.11 \times 70 = 25$

- If Forecast=bad

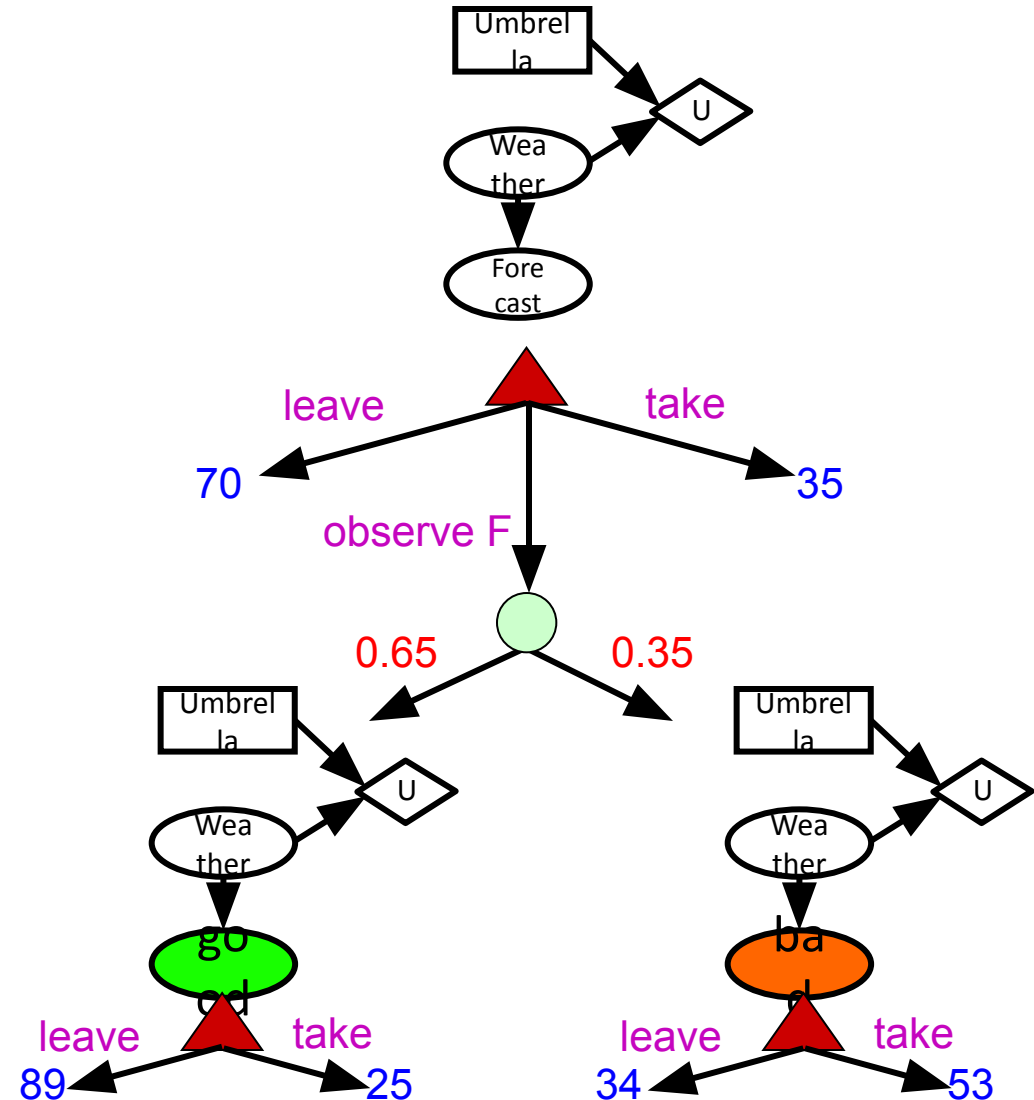
- $EU(\text{leave} \mid F=\text{bad}) = 0.34 \times 100 + 0.66 \times 0 = 34$
 - $EU(\text{take} \mid F=\text{bad}) = 0.34 \times 20 + 0.66 \times 70 = 53$

- $P(\text{Forecast}) = \langle 0.65, 0.35 \rangle$

- Expected utility given forecast

- $= 0.65 \times 89 + 0.35 \times 53 = 76.4$

- Value of information** = $76.4 - 70 = 6.4$



Video of Demo Ghostbusters with VPI



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
- Key point: decision network contains everything needed to compute it!
- $VPI(E_j | e) = \left[\sum_{e_j} P(e_j | e) \max_a EU(a | e_j, e) \right] - \max_a EU(a | e)$

VPI Properties

VPI is non-negative! $VPI(E_i | e) \geq 0$



VPI is not (usually) additive: $VPI(E_i, E_j | e) \neq VPI(E_i | e) + VPI(E_j | e)$



VPI is order-independent: $VPI(E_i, E_j | e) = VPI(E_j, E_i | e)$

