

Utility

Many slides courtesy of
Dan Klein, Stuart Russell,
or Andrew Moore

CS 5300 / CS 6300
Artificial Intelligence
Spring 2010

Hal Daumé III
hal@cs.utah.edu

www.cs.utah.edu/~hal/courses/2010S_AI

Announcements

- P1 solution up on Thursday
- HW grades will be out today
- P2 up
 - By end of today, you can complete it
 - Feel free to use anything from our P1 or your P1

Hal's Lottery

- You pay $\$M$ to enter my lottery
- I put $\$1$ in the pot
- Now, I start flipping fair coins
 - If the coin = heads, I double the pot
 - If the coin = tails, the game ends and you get the pot
- How much would you pay ($\$M$) to enter my lottery?
- Note, $\$1 = 30$ minutes on P2

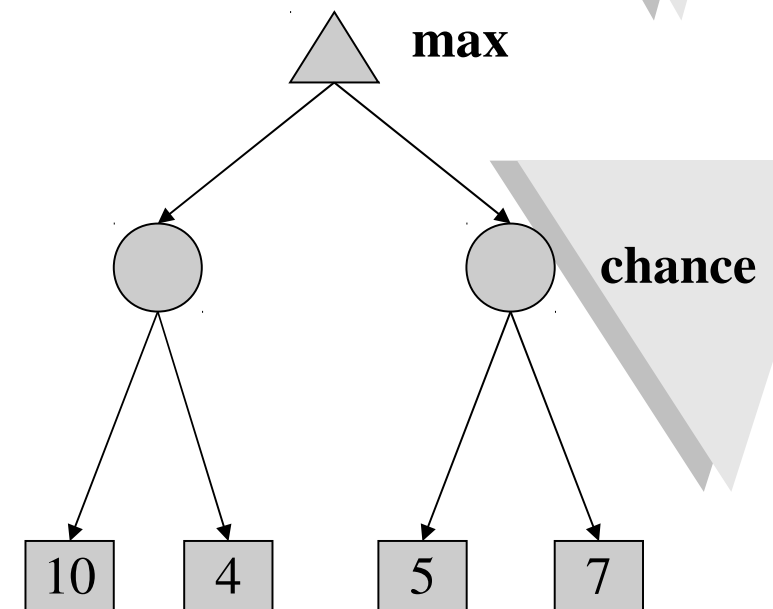
Where we are and where we're going

- Where we've been:
 - Single agent, known world, known rewards
 - Multi-agent, known world, known rewards
- Where we're going:
 - Stochastic, known world, known rewards
(Markov Decision Processes)
 - Stochastic, ~known world, unknown rewards
(Reinforcement Learning)

Expectimax Search Trees

- What if we don't know what the result of an action will be? E.g.,
 - In solitaire, next card is unknown
 - In minesweeper, mine locations
 - In pacman, the ghosts act randomly
- Can do **expectimax search**
 - Chance nodes, like min nodes, except the outcome is uncertain
 - Calculate **expected utilities**
 - Max nodes as in minimax search
 - Chance nodes take average (expectation) of value of children
- Later, we'll learn how to formalize the underlying problem as a

Markov Decision Process



Maximum Expected Utility

- Why should we average utilities? Why not minimax?
- Principle of maximum expected utility: an agent should choose the action which **maximizes its expected utility, given its knowledge**
- General principle for decision making
- Often taken as the definition of rationality
- We'll see this idea over and over in this course!
- Let's decompress this definition...

Reminder: Probabilities

- A **random variable** is an event whose outcome is unknown
- A **probability distribution** is an assignment of weights to outcomes
- Example: traffic on freeway?
 - Random variable: T = whether there's traffic
 - Outcomes: T in {none, light, heavy}
 - Distribution: $P(T=\text{none}) = 0.25$, $P(T=\text{light}) = 0.55$, $P(T=\text{heavy}) = 0.20$
- Some laws of probability (more later):
 - Probabilities are always non-negative
 - Probabilities over all possible outcomes sum to one
- As we get more evidence, probabilities may change:
 - $P(T=\text{heavy}) = 0.20$, $P(T=\text{heavy} \mid \text{Hour}=8\text{am}) = 0.60$
 - We'll talk about methods for reasoning about probabilities later

What are Probabilities?

- **Objectivist / frequentist answer:**
 - Averages over repeated *experiments*
 - E.g. empirically estimating $P(\text{rain})$ from historical observation
 - Assertion about how future experiments will go (in the limit)
 - New evidence changes the *reference class*
 - Makes one think of *inherently random* events, like rolling dice

- **Subjectivist / Bayesian answer:**
 - Degrees of belief about unobserved variables
 - E.g. an agent's belief that it's raining, given the temperature
 - E.g. pacman's belief that the ghost will turn left, given the state
 - Often *learn* probabilities from past experiences (more later)
 - New evidence *updates beliefs* (more later)

Dutch Books

Horse	Odds	Price
1	Even	\$100
2	3 to 1	\$50
3	4 to 1	\$40
4	9 to 1	\$20

- If your internal beliefs don't obey the rules of probability:
 - I can construct a Dutch book
 - \implies I can take infinite amounts of money from you!

Uncertainty Everywhere

- Not just for games of chance!
 - I'm sniffing: am I sick?
 - Email contains "FREE!": is it spam?
 - Tooth hurts: have cavity?
 - 60 min enough to get to the airport?
 - Robot rotated wheel three times, how far did it advance?
 - Safe to cross street? (Look both ways!)

- Why can a random variable have uncertainty?
 - Inherently random process (dice, etc)
 - Insufficient or weak evidence
 - Ignorance of underlying processes
 - Unmodeled variables
 - The world's just noisy!

- Compare to *fuzzy logic*, which has *degrees of truth*, or rather than just *degrees of belief*

Reminder: Expectations

- Often a quantity of interest depends on a random variable
- The expected value of a function is its average output, weighted by a given distribution over inputs
- Example: How late if I leave 60 min before my flight?
 - Lateness is a function of traffic:
 $L(\text{none}) = -10$, $L(\text{light}) = -5$, $L(\text{heavy}) = 15$
 - What is my expected lateness?
 - Need to specify some belief over T to weight the outcomes
 - Say $P(T) = \{\text{none: } 2/5, \text{light: } 2/5, \text{heavy: } 1/5\}$
 - The expected lateness:

$$E_{P(T)}[L(T)] = \frac{2}{5} \times (-10) + \frac{2}{5} \times (-5) + \frac{1}{5} \times (15)$$

$$P(\text{none})L(\text{none}) + P(\text{light})L(\text{light}) + P(\text{heavy})L(\text{heavy})$$

Reminder: Expectations

- Real valued functions of random variables:

$$f : X \rightarrow R$$

- Expectation of a function of a random variable

$$E_{P(X)}[f(X)] = \sum_x f(x)P(x)$$

- Example: Expected value of a fair die roll

X	P	f
1	1/6	1
2	1/6	2
3	1/6	3
4	1/6	4
5	1/6	5
6	1/6	6

$$1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6}$$

$$= 3.5$$

Two Envelopes Problem

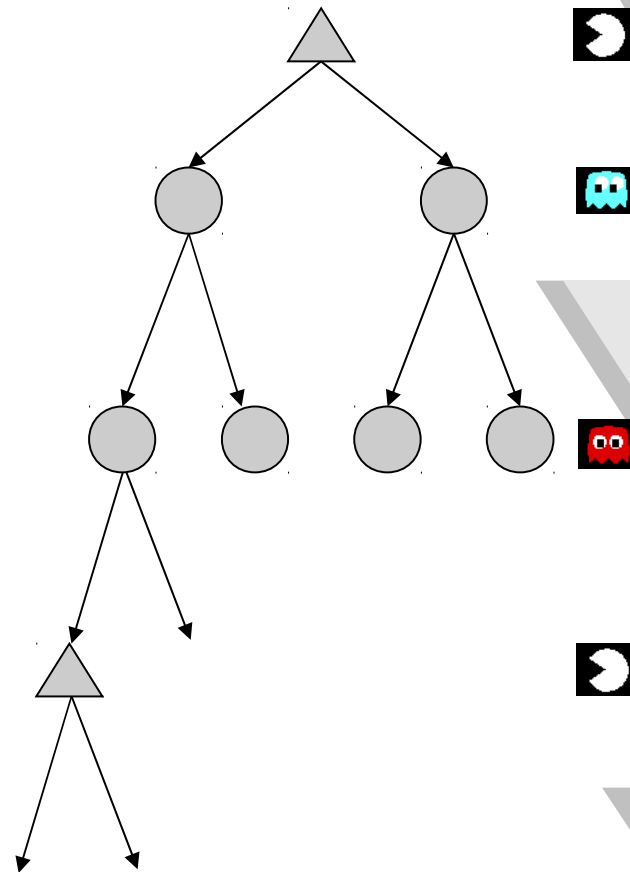
- One envelope contains \$100, the other \$200
- Pick an envelope, then I'll let you switch if you want
- Pick an envelope A
- $p(A \text{ is } \$100) = p(A \text{ is } \$200) = 0.5$
- if A is \$100, then other contains \$200
if A is \$200, then other contains \$100
- So other contains $2 \cdot A$ with $p=0.5$ and $A/2$ with $p=0.5$
- $E[\text{money in other}]$
- So you should swap...
- and swap...
- and swap...

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 set of preferences between outcomes can be summarized as a utility function (provided the preferences meet certain conditions)
- In general, we hard-wire utilities and let actions emerge (why don't we let agents decide their own utilities?)
- More on utilities soon...

Expectimax Search

- In expectimax search, we have a probabilistic model of how the opponent (or environment) will behave in any state
 - Model could be a simple uniform distribution (roll a die)
 - Model could be sophisticated and require a great deal of computation
 - We have a node for every outcome out of our control: opponent or environment
 - The model might say that adversarial actions are likely!
- For now, assume for any state we magically have a distribution to assign probabilities to opponent actions / environment outcomes



Having a probabilistic belief about an agent's action does not mean that agent is flipping any coins!

Expectimax Pseudocode

```
def value(s)
```

```
  if s is a max node return maxValue(s)
```

```
  if s is an exp node return expValue(s)
```

```
  if s is a terminal node return evaluation(s)
```

```
def maxValue(s)
```

```
  values = [value(s') for s' in successors(s)]
```

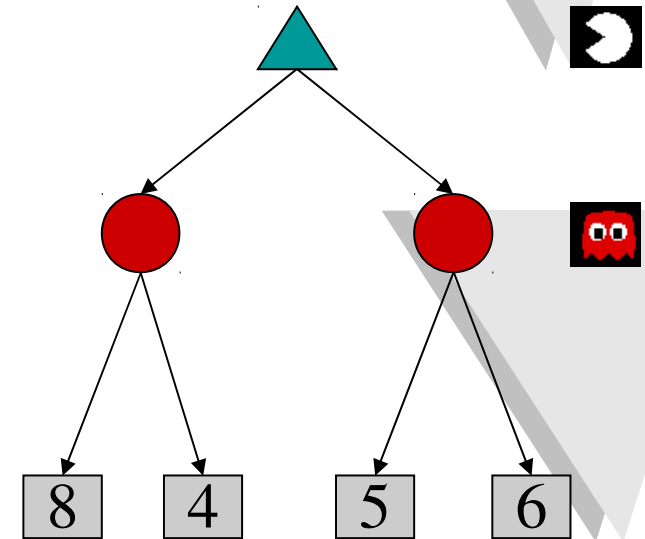
```
  return max(values)
```

```
def expValue(s)
```

```
  values = [value(s') for s' in successors(s)]
```

```
  weights = [probability(s, s') for s' in successors(s)]
```

```
  return expectation(values, weights)
```



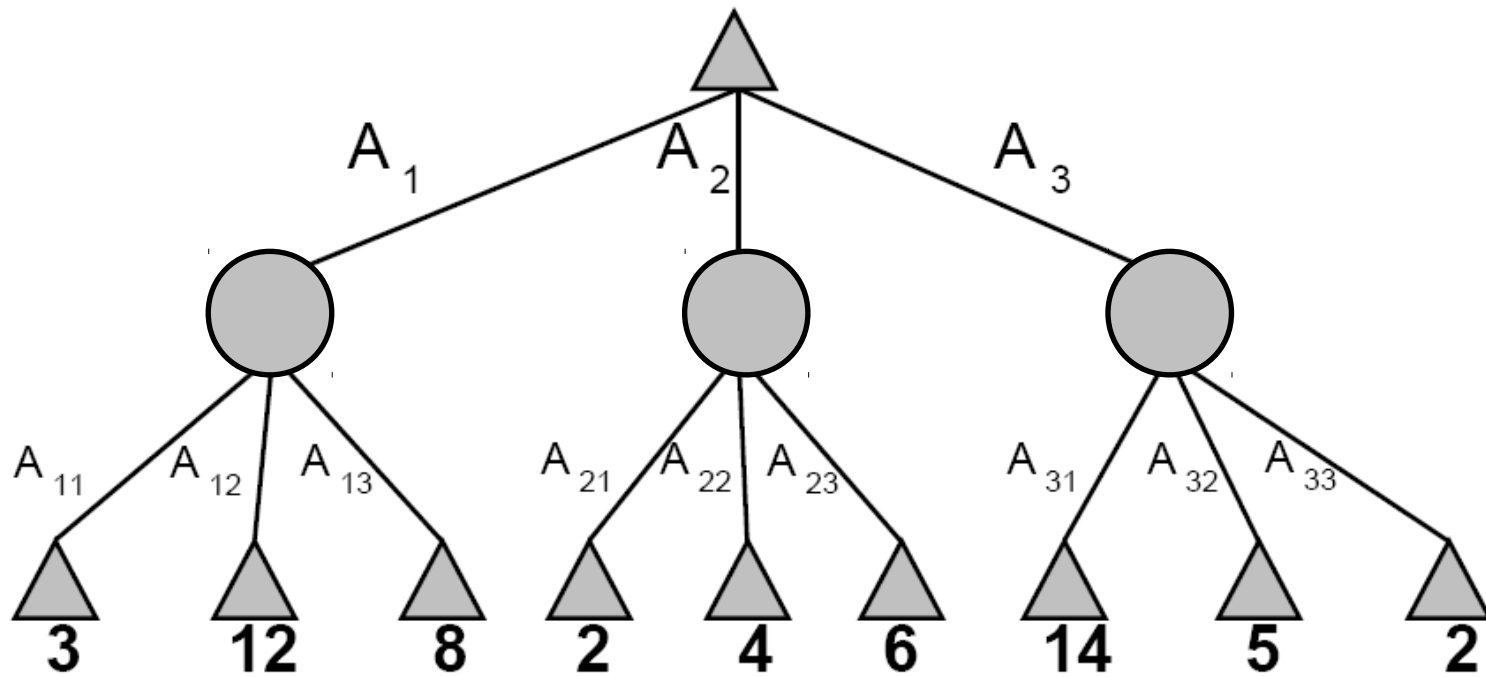
Expectimax for Pacman

- Notice that we've gotten away from thinking that the ghosts are trying to minimize pacman's score
- Instead, they are now a part of the environment
- Pacman has a belief (distribution) over how they will act

- Quiz: Can we see minimax as a special case of expectimax?
- Quiz: what would pacman's computation look like if we assumed that the ghosts were doing 1-ply minimax and taking the result 80% of the time, otherwise moving randomly?

- If you take this further, you end up calculating belief distributions over your opponents' belief distributions over your belief distributions, etc...
 - Can get unmanageable very quickly!

Expectimax Pruning?



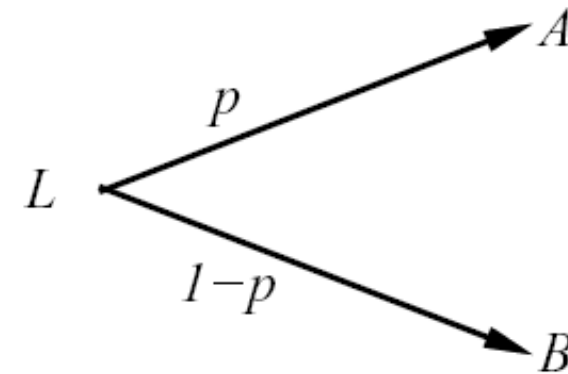
Expectimax Evaluation

- For minimax search, evaluation function insensitive to monotonic transformations
 - We just want better states to have higher evaluations (get the ordering right)
- For expectimax, we need the magnitudes to be meaningful as well
 - E.g. must know whether a 50% / 50% lottery between A and B is better than 100% chance of C
 - 100 or -10 vs 0 is different than 10 or -100 vs 0

Preferences

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

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



- Notation:

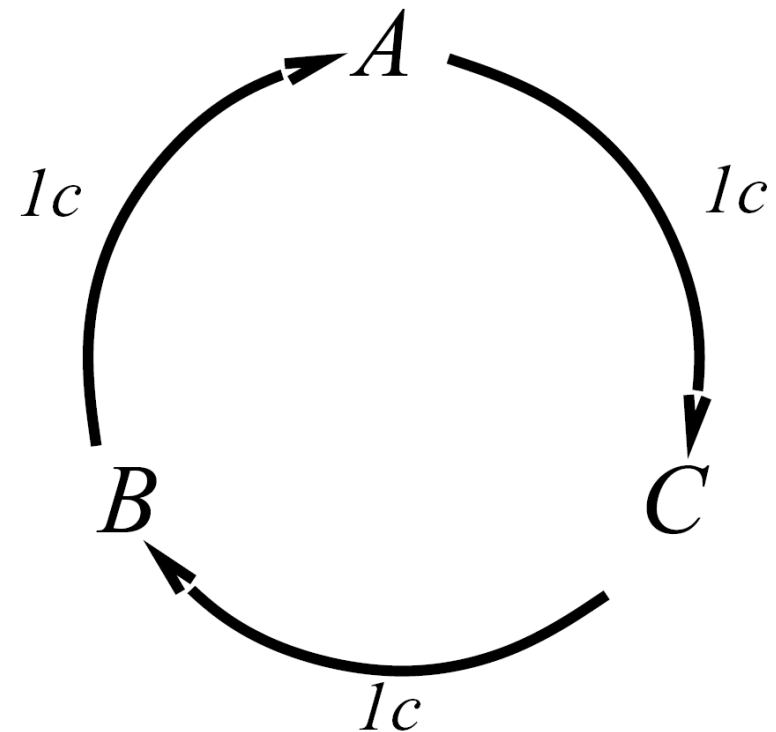
$A \succ B$ A preferred over B

$A \sim B$ indifference between A and B

$A \succeq B$ B not preferred over A

Rational Preferences

- We want some constraints on preferences before we call them rational
- For example: an agent with intransitive preferences can be induced to give away all 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

- Preferences of a rational agent must obey constraints.
- 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

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

Substitutability

$$A \sim B \Rightarrow [p, A; 1 - p, C] \sim [p, B; 1 - p, C]$$

Monotonicity

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

- 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:

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

$$U([p_1, S_1; \dots ; p_n, S_n]) = \sum_i p_i U(S_i)$$

- Maximum expected likelihood (MEU) principle:
 - Choose the action that maximizes expected utility
 - Note: an agent can be entirely rational (consistent with MEU) without ever representing or manipulating utilities and probabilities
 - E.g., a lookup table for perfect tictactoe, reflex vacuum cleaner

Pascal's Wager (d 1662)

- A “proof” that it is a good idea to believe in God

	God Exists	God Doesn't Exist
Believe	+infinity	-10
Don't Believe	-infinity	0

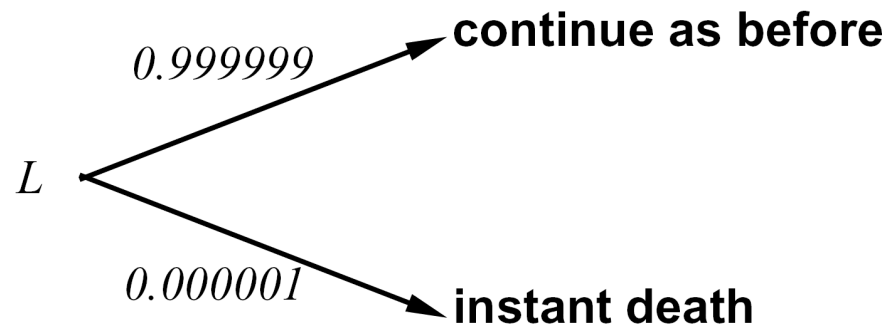
- Problems with this argument (mathematically)?
- Problems with this argument (theologically)?
- Exists in many cultures:
 - Islam: al-Juwayni (d 1085)
 - Sanskrit: Sarasamuccaya

Human Utilities

- Utilities map states to real numbers. Which numbers?
- Standard approach to assessment of human utilities:
 - Compare a state A to a **standard lottery** L_p between
 - “best possible prize” u_+ with probability p
 - “worst possible catastrophe” u_- with probability $1-p$
 - Adjust lottery probability p until $A \sim L_p$
 - Resulting p is a utility in $[0,1]$

pay \$30

~



Utility Scales

- **Normalized utilities:** $u_+ = 1.0$, $u_- = 0.0$
- **Micromorts:** one-millionth chance of death, useful for paying to reduce product risks, etc.
- **QALYs:** quality-adjusted life years, useful for medical decisions involving substantial risk
- Note: behavior is invariant under positive linear transformation

$$U'(x) = k_1 U(x) + k_2 \quad \text{where } k_1 > 0$$
- With deterministic prizes only (no lottery choices), only **ordinal utility** can be determined, i.e., total order on prizes

Example: Insurance

- Consider the lottery $[0.5, \$1000; 0.5, \$0]$
 - What is its **expected monetary value**? (\$500)
 - What is its **certainty equivalent**?
 - Monetary value acceptable in lieu of lottery
 - \$400 for most people
 - Difference of \$100 is the **insurance premium**
 - There's an insurance industry because people will pay to reduce their risk
 - If everyone were risk-prone, no insurance needed!

Hal's Lottery, revisited

Friendly game	\$100	\$4.28
Millionaire	\$1,000,000	\$10.95
Billionaire	\$1,000,000,000	\$15.93
Bill Gates (2008)	\$58,000,000,000	\$18.84
US GDP (2007)	\$13.8 trillion	\$22.79
World GDP (2007)	\$54.3 trillion	\$23.77
Googolaire	10^{100}	\$166.50

- How much would you pay (\$M) to enter my lottery?

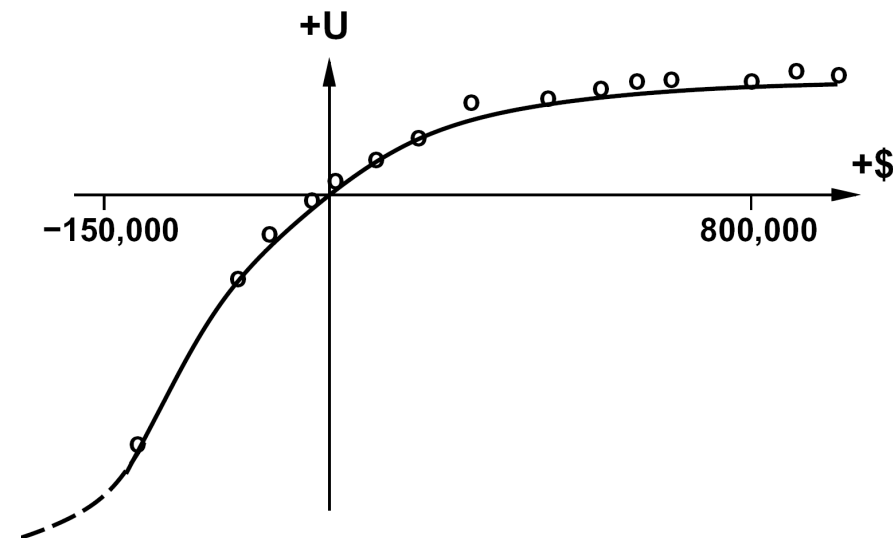
$$\begin{aligned}
 E[\text{payoff}] &= (1/2) 1 + (1/4) 2 + (1/8) 4 + (1/16) 8 + \dots \\
 &= (1/2) + (1/2) + (1/2) + (1/2) + \dots \\
 &= \text{infinity!}
 \end{aligned}$$

- Why weren't we willing to pay \$1m to enter?

- Non-linearity of utility of \$\$\$ (don't fix the problem)
- People underestimate small probabilities (not true!)
- No one would ever offer this lottery (I did!)
- Game can't be played infinitely long (no infinite resources)

Money

- Money does **not** behave as a utility function
- Given a lottery L :
 - Define **expected monetary value** $EMV(L)$
 - Usually $U(L) < U(EMV(L))$
 - I.e., people are **risk-averse**
- Utility curve: for what probability p am I indifferent between:
 - A prize x
 - A lottery $[p, \$M; (1-p), \$0]$ for large M ?
- Typical empirical data, extrapolated with **risk-prone** behavior:



Example: Human Rationality?

- Famous example of Allais (1953)
 - A: [0.8, \$4k; 0.2, \$0]
 - B: [1.0, \$3k; 0.0, \$0]
 - C: [0.2, \$4k; 0.8, \$0]
 - D: [0.25, \$3k; 0.75, \$0]
- Most people prefer $B > A$, $C > D$
- But if $U(\$0) = 0$, then
 - $B > A \Rightarrow U(\$3k) > 0.8 U(\$4k)$
 - $C > D \Rightarrow 0.8 U(\$4k) > U(\$3k)$

Reinforcement Learning

- [DEMOS]
- Basic idea:
 - Receive feedback in the form of **rewards**
 - Agent's utility is defined by the reward function
 - Must learn to act so as to **maximize expected rewards**
 - **Change the rewards, change the learned behavior**
- Examples:
 - Playing a game, reward at the end for winning / losing
 - Vacuuming a house, reward for each piece of dirt picked up
 - Automated taxi, reward for each passenger delivered

Markov Decision Processes

- An MDP is defined by:
 - A **set of states** $s \in S$
 - A **set of actions** $a \in A$
 - A **transition function** $T(s,a,s')$
 - Prob that a from s leads to s'
 - i.e., $P(s' | s,a)$
 - Also called the model
 - A **reward function** $R(s, a, s')$
 - Sometimes just $R(s)$ or $R(s')$
 - A **start state** (or distribution)
 - Maybe a **terminal state**

- MDPs are a family of non-deterministic search problems
 - Reinforcement learning: MDPs where we don't know the transition or reward functions

