Sources: John Gasper, Principles of Game Theory; Pedro Domingos, Decision Theory; Randall Munroe, XKCD; Wikipedia, various

Decision Making in Technology

A painfully brief intro to game theory, decision theory, utility and preferences, and equilibria

RJ10, 3D Printing: Sunday
Paper Topic: Tuesday

Bookkeeping

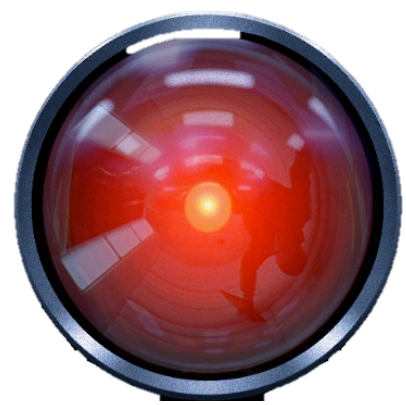
- Paper and paper topic
  - Paper topic: November 4th (Tuesday) at midnight
    - Topic; primary ethical concern or sub-area; proposed title
  - Paper due: last day of class
    - There will be some in-between steps posted Tuesday
  - You MAY work with ONE partner
    - Shared presentation
    - Recommend 50% longer paper (6 instead of 4)
    - Written statements about what each person did
- Last ~3 class periods: presentations
- Participation Portfolio 2 will be due Nov 8th

Bookkeeping

- Midterms
  - Don't be alarmed by essay grades – always curved
    - Essay grading: there's always something to improve
  - Specific concerns? 1-week window
    - Please do come ask questions or object!
- Ethical Analysis 2 will be posted Monday
  - Due Nov. 17th – two weeks later

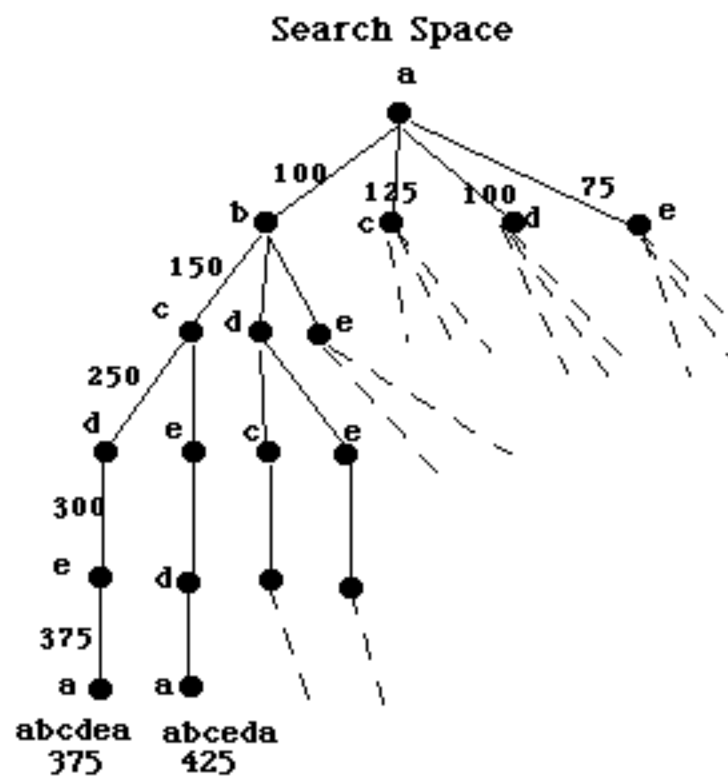
How Do Agents Make Decisions?

- What agents?
  - People: "You" ; "Everyone else"
  - Computer systems (real-time and not)
  - Artificial intelligences (many kinds)



When Do Computers Make Decisions?

- At every fork in a search problem
- Every time a credit card is run
- Every time a prescription is filled
- …a book is recommended
- …a dossier is flagged
- …a command is interpreted
- And so on.



Terminology and Concepts

- Optimizing: obtaining best possible outcome
  - Maximizing the value of some *objective function*
- Objective Function
  - A mathematical expression whose output you want to maximize/minimize
  - Optimization is finding the right *input parameters*
- Multi-objective optimization
  - Finding the best possible parameters given *multiple* objective functions

Terminology and Concepts

- Expected Value
  - The predicted value of a variable, calculated as:
  - The sum of all possible values, *each* multiplied by the probability of its occurrence

A \$1000 bet for a 20% chance to win \$10,000

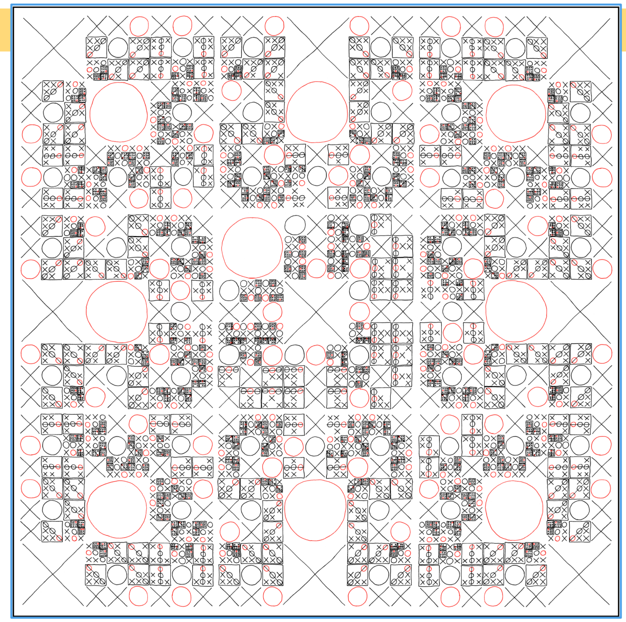
[20%(\$10,000)+80%(\$0)] = \$2000

- Satisficing: achieving a goal *sufficiently*
  - You can win a baseball game by one point now, or by two points in another inning
  - You can have a search function that finishes in one second, or spend another 2 hours to make it half a second; full credit is 3 seconds or less

Terminology and Concepts

- Game Theory
  - *Mathematical* models of interaction
  - Between intelligent, rational decision-makers
- Decision Theory
  - Normative: how *should* agents make decisions?
  - Descriptive: how *do* agents make decisions?
- Utility and utility functions
  - Something's perceived ability to satisfy needs or wants
  - A mathematical function that ranks alternatives by utility

What is Game Theory?



- Study of rational behavior in interactive situations
  - Everyone is self-interested and selfish
  - Or at least rational (weaker than selfish)
- Problems:
  - We aren't "rational" (agents *can* be)
  - Knowing theory doesn't guarantee success
- Goal: optimize chances of success
  - Achieving some target state
  - Optimizing some value

More Terminology

- Rationality (an overloaded word).
- A rational agent…
  - Behaves according to a *ranking over possible outcomes* that is:
    - Complete (covers all situations) and consistent
    - Optimizes over strategies to best serve a desired interest
- Has logical implications of knowledge
  - Assume that players have logical omniscience
    - If player 1 knows A, then 1 knows all of the logical implications of A
  - Assume players know all possible implications
    - If 2 knows that 1 knows that 2 knows that …

Classifying games

1. Sequential or simultaneous move?
   - Does a player get to change actions based on others' behavior?
     - Tic-tac-toe ; rock paper scissors
2. Zero-sum?
   - Gain or loss is exactly balanced by opponent's gain or less
     - Chess ; tic-tac-toe ; soccer
3. One shot or repeated?
   - Either one is not necessarily easier or harder
   - Opportunity to build *reputation* (for good or bad)

Classifying games

4. Full, partial, or asymmetric information?

- Perfect information:
- Each player has information of all previous events

  chess ; battleship

- Complete information:
- Every player knows payoffs for all possible actions

  tic-tac-toe ; Prisoner's Dilemma ; car buying

5. Non-cooperative or cooperative?

- Are agreements enforceable?

Prisoner's Dilemma

		Player 1	
		Silence	Testify
Player 2	Silence	P1: 6 months P2: 6 months	P1: Free P2: 10 years
	Testify	P1: 10 years P2: Free	P1: 5 years P2: 5 years

- One-shot or repeated?
  - Actually, either.
- Simultaneous?
- Zero-sum?
- Any uncertainty?
- Fixed rules?
- Cooperative?

Elements of a Game

Conceptual elements:

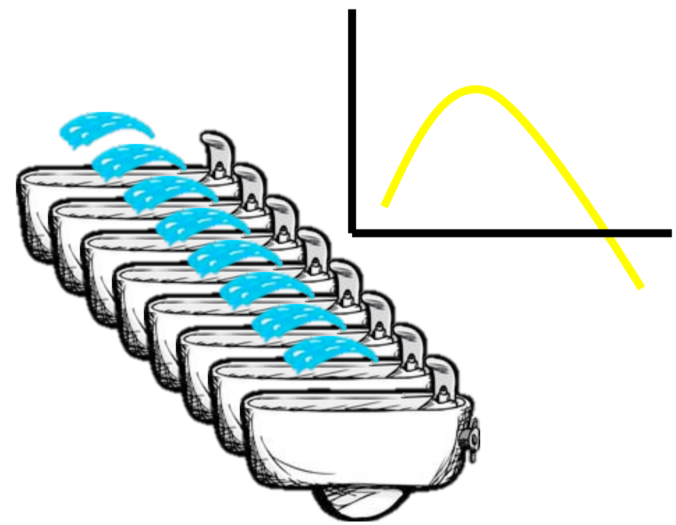
1. Actions and Strategies
   - *Actions*: the available actions any singular point
   - *Strategy*: complete plan for deciding actions
     - Different from "tactics and strategies"
2. Payoffs
   - An objective isn't necessarily "winning"!
     - See: non-zero-sum games; car buying
   - Represent preferences with a payoff *function*
   - Can be monetary but will be represented by *utility*

What is Decision Theory?

- Mathematical study of strategies for optimal decision-making
  - Options involve different *risks* or *expectations of gain or loss*
- The study of identifying:
  - The *values*, *uncertainties* and other issues relevant to a decision
  - The resulting *optimal decision*
- What does decision mean?

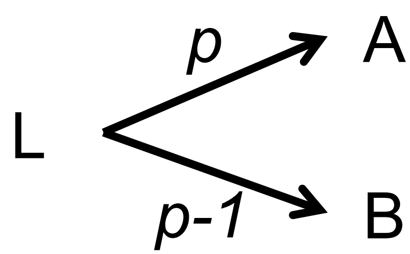
Utility

- Utility: perceived ability to satisfy needs or wants
- Utility function: Mathematical f that ranks alternatives
- Marginal utility: utility of subsequent iterations of thing
- Total utility: Utility of consuming ALL THE THINGS



Preferences

- An agent chooses among:
  - *Prizes* (A, B, etc.)
  - *Lotteries* (situations with uncertain prizes and probabilities)



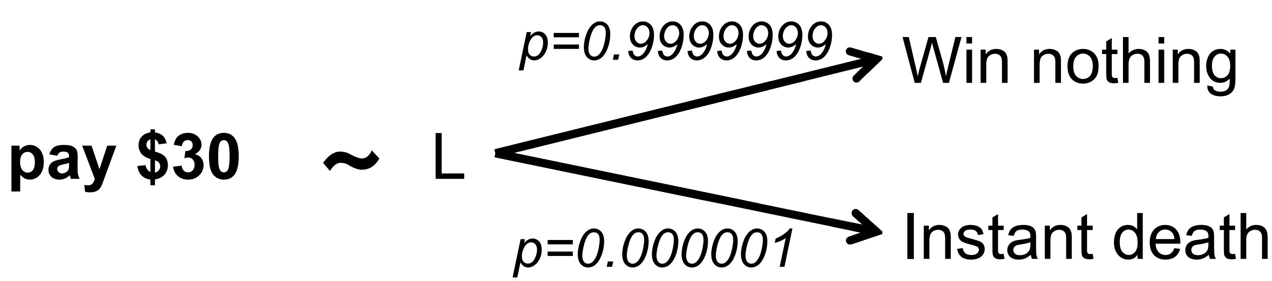
- Notation:
  - $A \succ B$ — A preferred to B
  - $A \sim B$ — Indifference between A and B
  - $A \succsim B$ — B not preferred to A

Rational Preferences

- Preferences of a rational agent must obey constraints
  - Transitivity
  - Monotonicity
  - Orderability $(A > B) \lor (B > A) \lor (A \sim B)$
  - Substitutability $(A \sim B \Rightarrow [p,A;\ 1-p,C] \sim [p,B;1-p,C]\ )$
  - Continuity $(A > B > C \Rightarrow\ \exists\, p\ [p,A;\ 1-p,C] \sim B\ )$
- Rational preferences, when followed, give behavior that *maximizes expected utility.*
- Violating the constraints leads to irrationality
  - For example: an agent with intransitive preferences can be induced to give away all its money.

Maximizing Expected Utility

- Utilities map *states* to *real numbers*. Which numbers?
  - People are very bad at mapping their own preferences
- Standard approach to assessment of human utilities:
  - Compare a state *A* to a standard lottery *Lp* that has
    "best possible prize" $u_\top$ with probability p
    "worst possible catastrophe" $u_\perp$ with probability $(1-p)$
  - adjust lottery probability p until *A* ~ *Lp*



Money

- Money does not behave as a utility function
  - That is, people don't maximize expected value of *dollar assets.*
- People are risk-averse:
  - Given a lottery L with expected monetary value EMV(L), usually U(L) < U(EMV (L))

Want to bet $1000 for a 20% chance to win $10,000?

[20%($10,000)+80%($0)] = $2000 > [100%($1000)]

- Expected Utility Hypothesis
  - rational behavior maximizes the expectation of some function *u*, which in need not be monetary

Actual Utility Scales

- Micromorts: one-millionth chance of death
  - Useful for:
  - Russian roulette
  - Paying to reduce product risks, etc.
- QALYs: quality-adjusted life years
  - Useful for:
  - Medical decisions involving substantial risk

Equilibria

- *Nash Equilibrium:* a state where no party has an incentive to unilaterally change strategies

		Player 1	
		Silence	Testify
Player 2	Silence	P1: 6 months P2: 6 months	P1: Free P2: 10 years
	Testify	P1: 10 years P2: Free	P1: 5 years P2: 5 years

- P1 and P2 have both always testified.
  - Who has incentive to be silent next round?
- Are there other equilibria?
- What game element would make it possible to change?

More Terminology

- Multi-objective optimization
  - Finding the best possible parameters given *multiple* objective functions
  - Decisions need to be optimized given trade-offs between two or more conflicting objectives
    - Minimizing cost and maximizing comfort while buying a car
- Pareto-optimal: no function can improve without another one degrading
  - It's impossible to make anyone better off without making someone worse off

When Do We Care?

- Deciding Whether to Approve a CC Transaction

	Approve	**Don't Approve**
Fraudulent	Gain: None Cost: Lost value of transaction	Gain: Customer trust Cost: Minor customer inconvenience; reissuing fee
Not fraudulent	Gain: Improved customer trust; transaction fee Cost: None	Gain: None Cost: Major customer inconvenience; loss of trust; reissuing fee

When Do We Care?

- Recommend a Book

	They buy it	They don't
Show a few good options	Gain: $$ Cost: Calculating options; possible minimal annoyance	Gain: None Cost: Minimal annoyance
Show a whole bunch of options	Gain: $$ Cost: Calculating options; they may or may not be substantially annoyed	Gain: None Cost: Potentially substantial annoyance