# Model-Based Reasoning: Preferences, Utilities and Decisions

AI for Computer Games

The University of Aberdeen

# Preference

- Often have situation where an agent is making a decision and it has a **preference** between the objects of choice.
- Example
  - Choices: Ice cream, Stilton
  - Ice Cream preferred to Stilton
- Can have many options in any one decision.
- If an agent prefers ice cream to stilton, and is faced with the above choice, then presumably it chooses ice cream rather than stilton.
- Some people take the fact that an agent makes choices that are consistent with its preferences to be the definition of rationality.

# Preference Notation

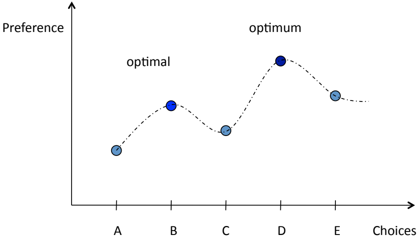
- $a \succ b$ means *$a$ is preferred to $b$.*
- $a \sim b$ means *$a$ is exactly as preferable as $b$.*
- $a \succeq b$ means *$a$ is preferable to $b$, or exactly as preferable as $b$.*

# Constraints on Rational Preference

- Must have
  - $a \sim a$
  - If $a \sim b$, then $b \sim a$
  - If $a \sim b$ and $b \sim c$, then $a \sim c$.
  - If $a \succ b$ and $b \succ c$, then $a \succ c$
  - At most one of $a \succ b$, $b \succ a$ or $a \sim b$ holds.
  - *etc.*

- Other possibilities (?)
  - (Totality) At least one of $a \sim b$ or $a \succ b$ or $b \succ a$ for all $a$, $b$.

# Decision as Optimization




- If there is an option that is known to be preferred to all others, then it should certainly be chosen.

# Choice of Action

- Sometimes the decision is between various choices of action.
- If an action is **deterministic** then it leads to a *single* **outcome**.
- Example: NPC deciding which room to visit next: $A$ or $B$.
  - Suppose deterministic actions: NPC chooses to go to room $A$, then finds itself in room $A$, etc.
- If the agent knows which outcomes it prefers, then it can use this to decide which action to choose.
  - Induces preference on actions from preference on outcomes.
  - If NPC prefers room $A$ to room $B$, then choose action '*go to room A*', rather than '*go to room B*'.

# Scoring outomes

- Sometimes preference is captured by scoring in some way:
  - Various words used: utility, payoff, loss, ...
- Suppose that an agent is faced with a decision over a set $A$ of choices. A **utility function** is a function

$$U : A \longrightarrow \mathbb{R} \ .$$

  That is, it assigns a value

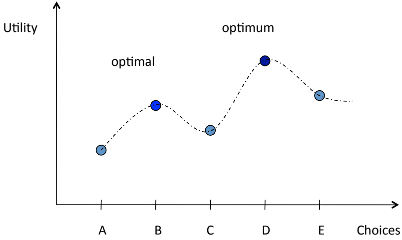
$$U(a) \in \mathbb{R}$$

  to each choice $a \in A$. This is called the **utility** of $a$.

# Fundamental Constraint on Utility

- The fact that the single utility score captures an agent's preferences means that:
  - $a \succ b$ if and only if $U(a) > U(b)$
  - $a \sim b$ if and only if $U(a) = U(b)$

# Maximizing the Utility Function




- Decision process becomes: take choice with the largest utility.

# Example Optimization

- Non-human player of strategy game has two actions:
  - build army (cost 1000 groats), and raid neighbouring country by land (revenue = 3000 groats)
  - build fleet (cost 4000 groats), and raid distant country by sea (revenue = 9000 groats)
- Game based around simplified economic model to make it fun and sufficiently realistic.
- Use simple utility function

$$U(\text{action}) = \text{revenue from action} - \text{cost of action} \ .$$

- Since $U(\text{fleet}) = 5000 > 2000 = U(\text{army})$, the second option is chosen.

# Loss

- An equivalent alternative decision procedure.
- Score each choice with a **loss function**.
- Choose the option with the **minimum** loss.

# Nondeterministic actions

- Suppose that each choice of action does *not* lead to a single, known, fixed outcome. Suppose that there are random effects, so that instead each action $a$ leads to a probability distribution

$$p_a$$

  over the set $O$ of outcomes (so $p_a(o) \in [0, 1]$ for all $o \in O$).

- Assume that each outcome, $o$, has a utility $U(o)$.

- e.g. A **lottery**.
  - Agent *chooses* numbers (or draws a straw etc).
  - Random process decides how much agent wins or loses.

- e.g. NPC agent fires an arrow.
  - Agent chooses the particular angle at which to fire.
  - Randomly blowing wind determines whether-or-not the arrow strikes an enemy.

# Expectation

- Recall that the **expectation** of a random variable is the mathematical version of the notion of (weighted) average.
- For a random variable $X$ taking values $x_1, \ldots x_n$:

$$E[X] = \sum_{i=1}^{n} x_1 \; p(X = x_1)$$

- Example: You win \$$x$ if the (fair, 6-sided) die lands with the face $x$ upwards:

$$E[X] = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6} = 3\frac{1}{2}$$

Expect to win \$3.50.

# Expected Utility

- Induce utility of actions, from utility of outcomes, and their probabilities.
- The **expected utility** for action $a$ is:

$$E[U(a)] = \sum_{o \in O} U(o)\ p_a(o)\ .$$

- Standard decision procedure is to **maximize expected utility**:
  - Choose action $a$ that maximizes this function.

# Expected Utility Example

- NPC in a game decides between two options
  - attack dragon, and steal its 30 pieces of gold
  - attack vampire, and steal its 40 pieces of silver.
- The attack on the dragon is succesful with probability 0.25
- The attack on the vampire is successful with probability 0.5
- A gold piece is worth two pieces of silver.
- A sensible choice of utility of an outcome $o$, with amounts of gold $g(o)$ and silver $s(o)$ is:

$$U(o) = s(o) + 2 \times g(o) \ .$$

- NPC chooses to attack vampire, because

$$E[U(\text{drag})] = 0.25 \times (0 + 2 \times 30) + 0.75 \times (0 + 2 \times 0) = 15$$

$$E[U(\text{vamp})] = 0.5 \times (40 + 2 \times 0) + 0.5 \times (0 + 2 \times 0) = 20 \ .$$

# Subjective Expected Utility

- Preference is subjective.
- Utility is subjective
- Why not use subjective probability in the expected utility calculations?
- Then one can deal with uncertain decisions, given:
  - levels of belief about outcomes of actions
  - levels of belief about state of the world.
- Although some inputs to the decision process are subjective, we are making them explicit and transparent.

# SEU Example

- Two actions: choose heads or tails.
- Toss coin. Two outcomes $H$, $T$.
- If guess correctly win $\$1$, otherwise lose $\$1$.
- You are a Bayesian and having collected evidence, $\mathcal{E}$, have most recently updated to the posterior distribution to:

$$p(H \mid \mathcal{E}) = 0.4 \ .$$

- Then:

$$\begin{aligned} E[U(\text{heads}) \mid \mathcal{E}] &= 0.4 \times 1 + 0.6 \times (-1) &= -0.2 \\ E[U(\text{tails}) \mid \mathcal{E}] &= 0.4 \times (-1) + 0.6 \times 1 &= 0.2 \ . \end{aligned}$$

- So choose the action tails.

# State-dependent outcomes

- Sometimes outcomes of actions are **state-dependent**.
- Example 1: My satisfaction after eating ice cream or stilton depends on weather, what I have already eaten.
- Example 2: in dungeon, an NPC in a weak state may choose not to fight enemy, but if NPC is strong then attacks.
- Example 3: regard a decision tree (earlier lecture) as a single action.
  - Repeatedly query state to find out what *atomic actions* to do.
- Can view action as a function

$$\text{action} : S \longrightarrow O$$

where $S$ is the set of states, and $O$ is the set of outcomes.

# SEU Example 2

- Computer is playing a player it doesn't know. Assumes opponent has a skill level 1, 2, or 3.
- Bayesian computer player has updated given evidence $\mathcal{E}$ to
$$p(\text{skill} = 1 \mid \mathcal{E}) = 0.25 = p(\text{skill} = 3 \mid \mathcal{E}) \; .$$
- Two actions: hard, easy.
- Utility

| action / skill | 1 | 2 | 3 |
|---|---|---|---|
| easy | 9 | 4 | 1 |
| hard | 0 | 6 | 10 |

-

$$\begin{aligned} E[U(\text{easy})] &= 0.25 \times 9 + 0.5 \times 4 + 0.25 \times 1 &= 4.5 \\ E[U(\text{hard})] &= 0.25 \times 0 + 0.5 \times 6 + 0.25 \times 10 &= 5.5 \end{aligned}$$

- Choose hard-setting.

# Complications around optimization

- Not always computationally feasible to calculate utility values for all possible choices.
  - Imagine trying to calculate some 'score' for every possible chess strategy.
  - Imagine trying to score every possible path across complicated terrain.
- Sometimes the search itself has a cost.
  - In some game, can't sit and calculate all possible best reponses to attack, have to act soon or will be killed.
  - Commercial decisions, options in financial markets.
  - **Time-value** of decisions.
- **Satisficing** means searching for a choice that is good enough.
  - Common, heuristic AI technique
  - Also seems common in real creatures.

# Utility isn't always simple

- A lot of the above examples assume that utility can be defined in a simple way,
  - Often relative to some notion of money.
- But utility has to correspond to **all** aspects of preference.
- This means that it could be:
  - hard to design utilities that capture people's real preferences in real-life decision situations (lots of subtle context)
  - easy to design utilities in games that lead to 'unrealistic' behaviour
- The subject of Behavioural/Experimental Economics tries to get at the preferences of real agents in real decision situations.

# Pop-quiz, hot-shot

- I offer you a chance to participate in the following lottery: I toss a fair coin, and if it lands heads then you pay me \$2000, but if it lands tails then I pay you \$2000. Would you play?
  - No? Then you have **loss aversion**.

- I offer you a choice: take \$1000 guaranteed, or accept a shot at a lottery based on a single, fair coin-toss in which you get \$2000 if you win but nothing if you lose. What do you do?
- Note that the two choices have equal expectation value \$1000.
  - Either, doesn't matter: then you are **risk-neutral**
  - Go for the guaranteed amount: you are **risk-averse**
  - Take the shot: you are **risk-seeking**.

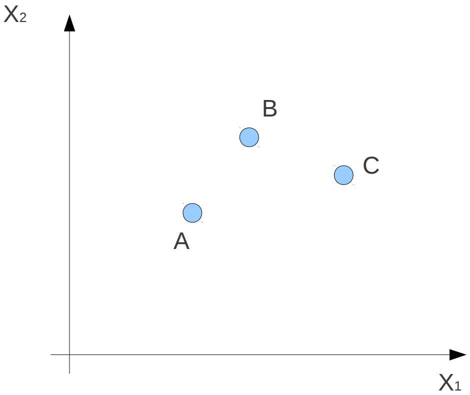
# Utility complications

- Loss and risk-aversion are very common in real people, in decisions that really matter to them.
- Just using the expectation value for a simple, monetary utility is not always a good guide to behaviour.
- Need to use more of distribution of outcomes than just expectations.
- Need to be very careful about situations where people/agents are motivated by things other than money.

# Attributes

- Objects of choice can have multiple **attributes**.
- Examples:
  - Would you prefer the fast, but uncomfortable sports car, or the slow, but luxurious saloon.
  - Would orc in dungeon prefer the room with a few potions and lots of gold, or the one with lots of potions and a little gold?
  - Would you sacrifice a pawn in order to get a bishop into a strong position?
- How can we compare these things, and make a rational decision?

# Three options with two attributes




- Do you prefer $A$, $B$, or $C$, with attributes $X_1$, $X_2$?

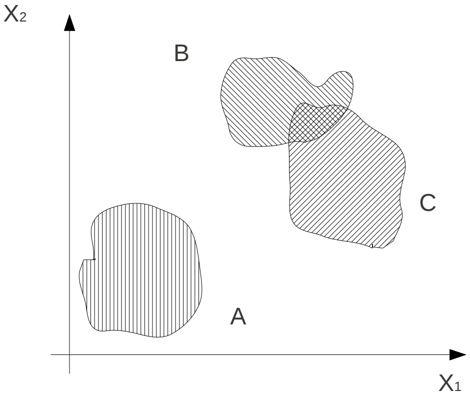
# Utility for multiattributes

- Multiattribute preferences are sometimes expressed through utilities.
- In example above, outcomes involved getting various quantities of gold (one attribute) and silver (other attribute).
- We were able to **trade-off** one attribute against the other, by saying that one piece of gold was worth two pieces of silver.
- Led us to

$$U(a) = s + 2 \times g$$

where $(g, s)$ were the attributes of the outcome of choosing $a$.

# Multiattributes with Uncertainty




- Outcomes of action choices can be smeared out as probability distributions.
- Just use expectations?

- Again, have a decision process with subjective components.
- However, subjective components and tradeoffs are all made explicit.

# Sequential Choices

- Have to decide between sequences of actions.
  - e.g. Sequences of chess moves, sequnces of rooms in dungeon to visit.
- If actions are deterministic can use standard search algorithms (like you learned in pathfinding)
- Suppose now that actions are nondeterministic.

# Markov Decision Processes

A **Markov Decision Process** (MDP) consists of the following components

- A set $A$ of actions
- A set of states, $S$
- An initial state $s_0$
- A **transition model** $T : S \times A \times S \longrightarrow [0, 1] \subseteq \mathbb{R}$
- A **reward function** $R : S \longrightarrow \mathbb{R}$

# Transition Model

- Basically, a directed graph, with edges between states $s$, $s'$ labelled by actions $a$ and probabilities $T(s, a, s')$.
- e.g.

$$\text{EnemyPresent} \xrightarrow{\text{FireArrow} \quad (0.4)} \text{EnemyDead}$$

- Describes the dynamics of the world: how actions cause state to change.
- For any $s$, $a$, have: $1 = \sum_{s' \in S} T(s, a, s')$.
- Example: Look back at the graph of the probabilistic automaton in the previous lecture/tutorial.

# Reward Function

- For each state $s$, the reward $R(s)$ is how much we like being in that state.
- A bigger reward is better.
- Instantaneous, collect it as we go. Different to utility.

# Combining Rewards

- The usual thing is to choose a **discount factor** $0 < \gamma \leq 1$.
- For a sequence of states $\langle s_0, s_1, \ldots \rangle$, define

$$U(\langle s_0, s_1, \ldots \rangle) = R(s_0) + \gamma R(s_1) + \ldots + \gamma^n R(s_n) + \ldots$$

- Discount factor tells us how much we prefer being rewarded now, compared to in the future.
- If **finite horizon**, i.e. stop after fixed, finite number of steps, can take $\gamma = 1$.
- Otherwise, choosing $\gamma < 1$ ensures that $U(\langle s_0, s_1, \ldots \rangle) < \infty$ (provided there is some bound $K$ such that $R(s) < K$ for all $s$).

# Policy

- A **policy**, $\pi$, consists of a choice of action for every reachable state $s$.
- Recall the original problem, to construct the best sequence.
- We will see that it suffices to construct an **optimal policy**, $\pi^*$.
- You have already looked at how to approximate optimal policies via Q-learning.

# Utility and Goal-seeking

- Utility maximization can be regarded as a simple form of **goal-oriented** behaviour
  - Goal: Maximize utility.
- Uncertainty: Handled by using probabilities and expectations.
- Multiple goals: multiple attributes wrapped up in one big utility function.
- Some games people, e.g. Millington, mean this when they refer to g.o.b., but others mean some form of logical AI **planning**.

# Warning

- Recall the warning I gave you about the difficulties of prediction when playing against intelligent agents who can second-guess your prediction of their choices.
- Again, need Game Theory if we have to deal with this level of complexity.

# References


- Russell and Norvig, chapter 16, 'Making Simple Decisions'
- Russel and Norvig, chapter 17 'Making 'Complex Decisions'
- Millington, Section 5.6 (unfortunately, calls it 'Goal-oriented Behaviour')
- Any good introduction to Decision Theory or Microeconomics, e.g. Kreps, *Notes on the Theory of Choice*