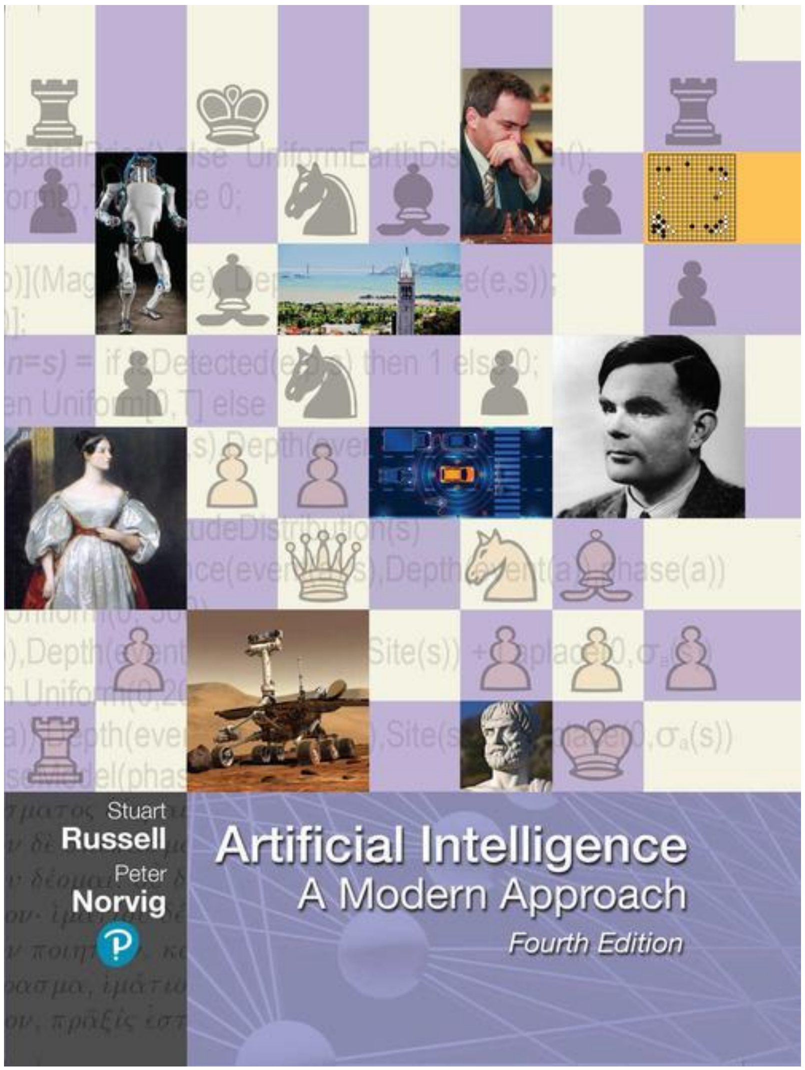
Artificial Intelligence: A Modern Approach

Fourth Edition



Chapter 16

Making Simple Decisions

Outline

- Combining Beliefs and Desires under Uncertainty Utilities
- The Basis of Utility Theory
- Utility Functions
- Multiattribute Utility Functions
- Decision networks
- Value of information
- Unknown Preferences

Pearson

Combining Beliefs and Desires under Uncertainty

An agent assigns a probability $P(s)$ to each possible current state s. There may also be uncertainty about the action outcomes; the transition model is given by $P(s^t \mid s, a)$,

$$P(\text{RESULT}(a) = s^t) = \sum_{s'} P(s)P(s^t \mid s, a) \ .$$

The **expected utility** of an action given the evidence, $EU(a)$, is just the average utility value of the outcomes, weighted by the probability that the outcome occurs:

$$EU(a) = \sum_{s'} P(\text{RESULT}(a) = s^t)U(s^t) \ .$$

The principle of **maximum expected utility (MEU)** says that a rational agent should choose the action that maximizes the agent's expected utility:

$$action = \underset{a}{\text{argmax}}\ EU(a) \ .$$

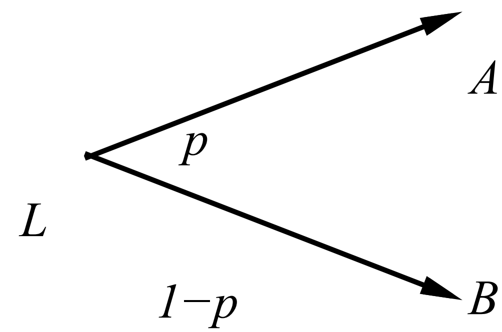
If agent acts so as to maximize a utility function that correctly reflects the performance measure, then the agent will achieve the highest possible performance score (averaged over all the possible environments).

The Basis of Utility Theory

Constraints on rational preferences

An agent chooses among prizes (*A*, *B*, etc.) and lotteries, i.e., situations with uncertain prizes

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



Notation:

A <; *B* — *A* preferred to *B*

A ~ *B* — indifference between *A* and *B*

A <; ~ *B* — *B* not preferred to *A*

Rational preferences

Idea: preferences of a rational agent must obey constraints.

Rational preferences $\Rightarrow$

behavior describable as maximization of expected utility

Constraints:

Orderability

$$(A <; B) \vee (B <; A) \vee (A \sim B)$$

Transitivity

$$(A <; B) \wedge (B <; C) \Rightarrow (A <; C)$$

Continuity

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

Substitutability

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

Monotonicity

$$A <; B \quad \Rightarrow (p \geq q \;\Leftrightarrow\; [p, A; \; 1 - p, B] \underset{\sim}{<;} [q, A; \; 1 - q, B])$$

Pearson

Rational preferences contd.

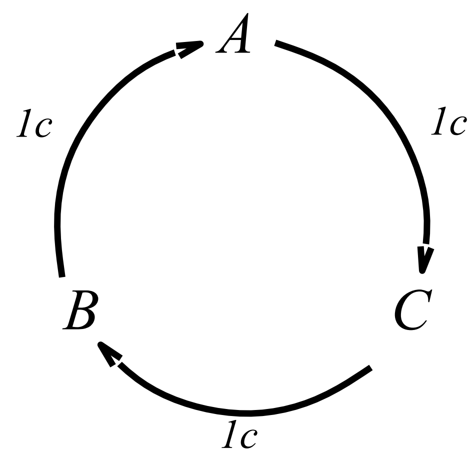
Violating the constraints leads to self-evident irrationality

For example: an agent with intransitive preferences can be induced to give away all its money

If B <; C, then an agent who has C would pay (say) 1 cent to get B

If A <; B, then an agent who has B would pay (say) 1 cent to get A

If C <; A, then an agent who has A would pay (say) 1 cent to get C



Maximizing expected utility

Theorem (Ramsey, 1931; von Neumann and Morgenstern, 1944):
Given preferences satisfying the constraints
there exists a real-valued function U such
that

$$U(A) \geq U(B) \quad \Leftrightarrow \quad A \succsim B$$
$$U([p_1, S_1; \; \ldots \; ; \; p_n, S_n]) = \Sigma_i \; p_i U(S_i)$$

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

Utility Functions

Utilities map states to real numbers. Which numbers?

Standard approach to assessment of human utilities:

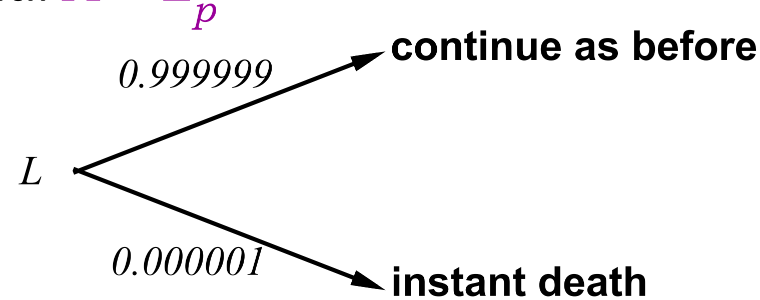
compare a given state A to a standard lottery L_p that has

"best possible prize" $u_\top$ with probability p

"worst possible catastrophe" $u_\bot$ with probability $(1 - p)$

adjust lottery probability p until $A \sim L_p$

pay $30 ~



Utility scales

Normalized utilities: $u_{\top} = 1.0$, $u_{\perp} = 0.0$

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

Note: behavior is invariant w.r.t. +ve linear transformation

$$U'(x) = k_1 U(x) + k_2 \text{ where } k_1 > 0$$

With deterministic prizes only (no lottery choices), only ordinal utility can be determined, i.e., total order on prizes

Pearson

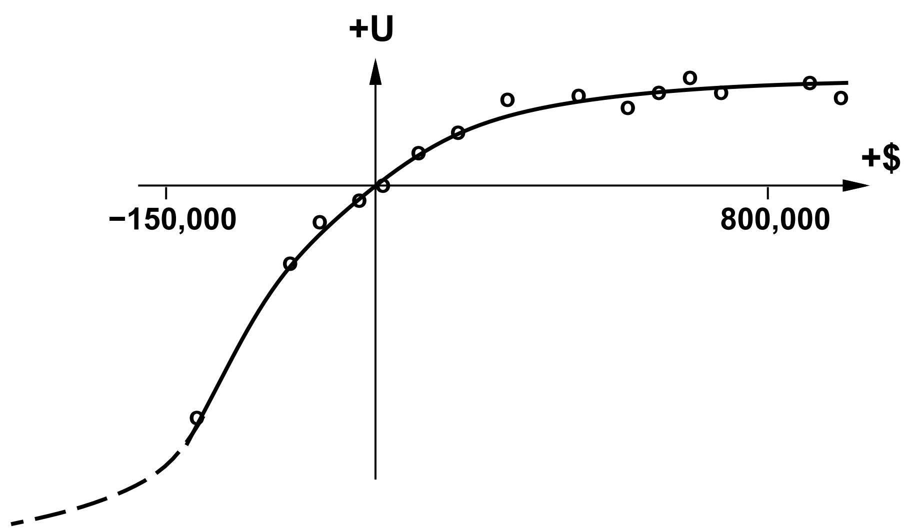
Money

Money does not behave as a utility function

Given a lottery L with 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 and a lottery $[p, \$M; (1-p), \$0]$ for large M ?

Typical empirical data, extrapolated with risk-prone behavior:



Multiattribute Utility Functions

How can we handle utility functions of many variables $X_1 \ldots X_n$? E.g., what is $U(Deaths, Noise, Cost)$?

How can complex utility functions be assessed from preference behaviour?

Idea 1: identify conditions under which decisions can be made without com- plete identification of $U(x_1, \ldots, x_n)$

Idea 2: identify various types of independence in preferences and derive consequent canonical forms for $U(x_1, \ldots, x_n)$

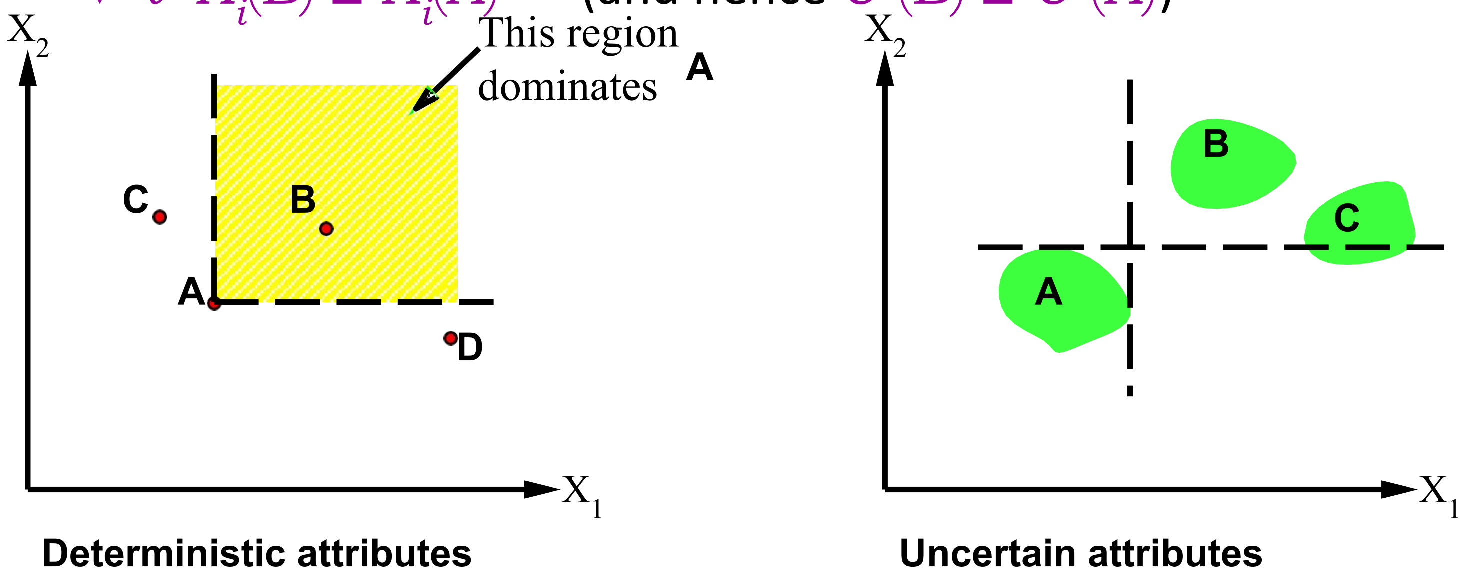
Pearson

Strict dominance

Typically define attributes such that U is monotonic in each

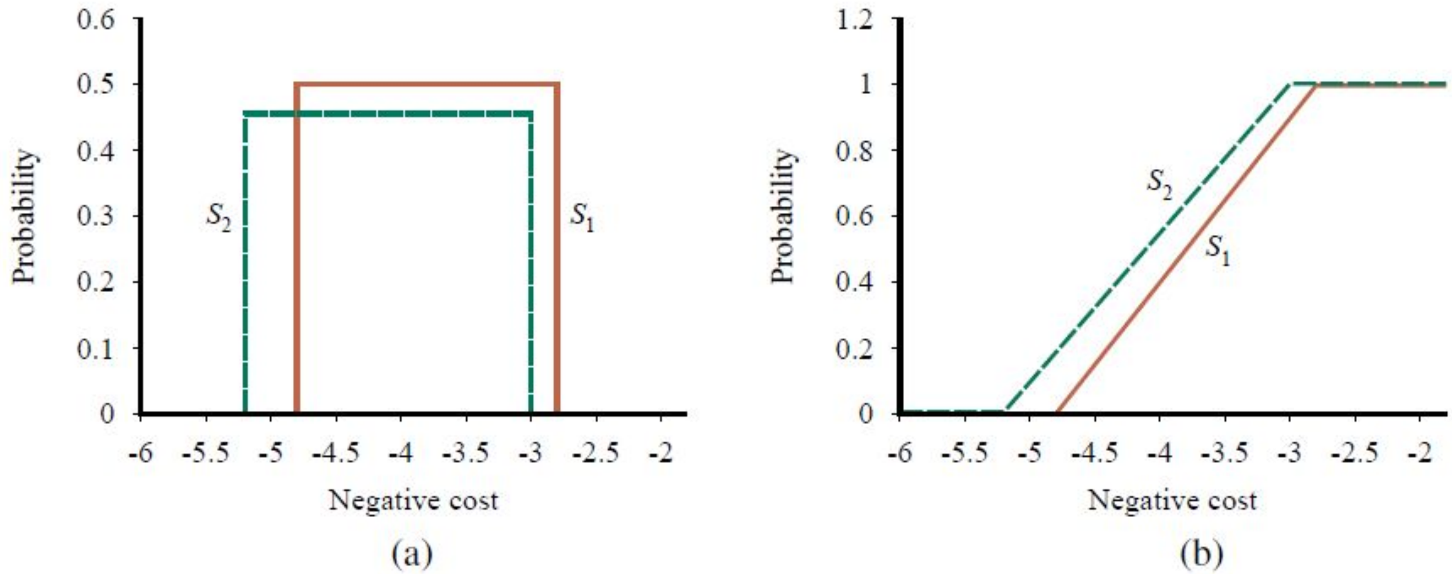
Strict dominance: choice B strictly dominates choice A iff

$\forall\ i\ \ X_i(B) \geq X_i(A)$ (and hence $U(B) \geq U(A)$)



Strict dominance seldom holds in practice

Stochastic dominance



Distribution p_1 stochastically dominates distribution p_2 iff $\forall\, t \int_{-\infty}^{t} p_1(x)\,dx \leq \int_{-\infty}^{t} p_2(t)\,dt$

If U is monotonic in x, then A_1 with outcome distribution p_1 stochastically dominates A_2 with outcome distribution p_2:

$$\int_{-\infty}^{\infty} p_1(x)U(x)\,dx \geq \int_{-\infty}^{\infty} p_2(x)U(x)\,dx$$

Multiattribute case: stochastic dominance on all attributes $\Rightarrow$ optimal

Pearson

Stochastic dominance contd.

Stochastic dominance can often be determined without exact distributions using qualitative reasoning

E.g., construction cost increases with distance from city

S_1 is closer to the city than S_2

$\Rightarrow S_1$ stochastically dominates S_2 on cost

E.g., injury increases with collision speed

Can annotate belief networks with stochastic dominance information:

$X \xrightarrow{+} Y$ (X positively influences Y) means that For every value $\mathbf{z}$ of Y's other parents $\mathbf{Z}$

$\forall\; x_1, x_2\; x_1 \geq x_2 \Rightarrow \mathbf{P}(Y|x_1, \mathbf{z})$ stochastically dominates $\mathbf{P}(Y|x_2, \mathbf{z})$

Preference structure: Deterministic

X_1 and X_2 preferentially independent of X_3 iff
preference between (x_1, x_2, x_3) and (x'_1, x'_2, x_3) does not depend on x_3

E.g., $(Noise, Cost, Safety)$:
(20,000 suffer, \$4.6 billion, 0.06 deaths/mpm) vs.
(70,000 suffer, \$4.2 billion, 0.06 deaths/mpm)

Theorem (Leontief, 1947): if every pair of attributes is P.I. of its complement, then every subset of attributes is P.I of its complement: mutual P.I..

Theorem (Debreu, 1960): mutual P.I. $\Rightarrow$ $\exists$ additive value function:

$$V(S) = \Sigma_i V_i(X_i(S))$$

Hence assess n single-attribute functions; often a good approximation

Preference structure: Stochastic

Need to consider preferences over lotteries:

X is utility-independent of Y iff

preferences over lotteries in X do not depend on y

Mutual U.I.: each subset is U.I of its complement

$\Rightarrow \exists$ multiplicative utility function:

$$U = k_1U_1 + k_2U_2 + k_3U_3$$
$$+ k_1k_2U_1U_2 + k_2k_3U_2U_3 + k_3k_1U_3U_1$$
$$+ k_1k_2k_3U_1U_2U_3$$

Routine procedures and software packages for generating preference tests to identify various canonical families of utility functions

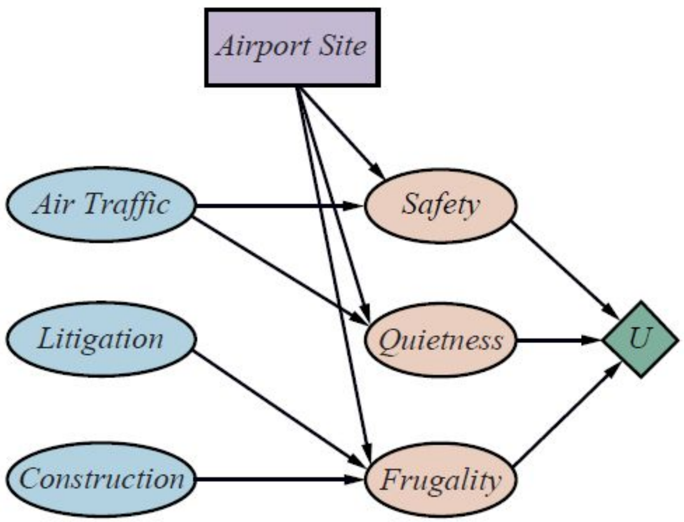
Decision networks

Decision network represents information about the agent's current state, its possible actions, the state that will result from the agent's action, and the utility of that state.

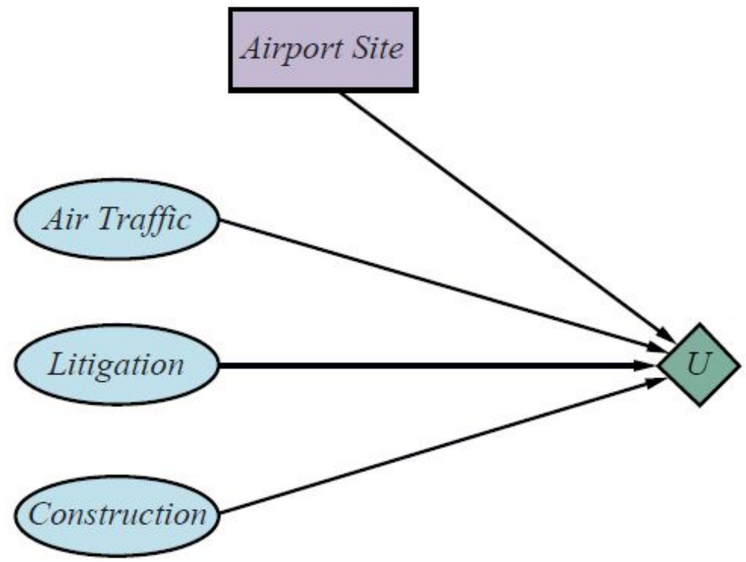
Three types of nodes used:

- **Chance nodes (ovals):** random variables, just as they do in Bayesian networks
- **Decision nodes (rectangles):** points where the decision maker has a choice of actions
- **Utility nodes (diamonds):** agent's utility function.

Decision networks



A decision network for the airport-siting problem.



A simplified representation of the airport-siting problem. Chance nodes corresponding to outcome states have been factored out.

Decision networks

The algorithm for **evaluating decision networks** is the following

1. Set the evidence variables for the current state.

2. For each possible value of the decision node:
(a) Set the decision node to that value.
(b) Calculate the posterior probabilities for the parent nodes of the utility node, using a standard probabilistic inference algorithm.
(c) Calculate the resulting utility for the action.

3. Return the action with the highest utility.

The Value of information

Idea: compute value of acquiring each possible piece of evidence Can be done directly from decision network

Example: buying oil drilling rights

Two blocks A and B, exactly one has oil, worth k

Prior probabilities 0.5 each, mutually exclusive

Current price of each block is $k/2$

"Consultant" offers accurate survey of A. Fair price?

Solution: compute expected value of information

= expected value of best action given the information
minus expected value of best action without
information

Survey may say "oil in A" or "no oil in A", prob. 0.5 each (given!)

= [$0.5 \times$ value of "buy A" given "oil in A"
+ $0.5 \times$ value of "buy B" given "no oil in A"]
– 0

= $(0.5 \times k/2) + (0.5 \times k/2) - 0 = k/2$

General formula

Current evidence E, current best action a

Possible action outcomes S_i, potential new evidence E_j

$$EU(a|E) = \max_a \Sigma_i \, U(S_i) \, P(S_i|E, a)$$

Suppose we knew $E_j = e_{jk}$, then we would choose a_{ejk} s.t.

$$EU(a_{ejk}|E, E_j = e_{jk}) = \max_a \Sigma_i \, U(S_i) \, P(S_i|E, a, E_j = e_{jk})$$

E_j is a random variable whose value is *currently* unknown

⇒ must compute expected gain over all possible values:

$$VPI_E(E_j) = \Sigma_k \, P(E_j = e_{jk}|E) EU(a_{ejk}|E, E_j = e_{jk}) - EU(a|E)$$

(VPI = value of perfect information)

Properties of VPI

Nonnegative—in expectation, not post hoc

$$\forall\ j, E \quad VPI_E(E_j) \geq 0$$

Nonadditive—consider, e.g., obtaining E_j twice

$$VPI_E(E_j, E_k) \ = VPI_E(E_j) + VPI_E(E_k)$$

Order-independent

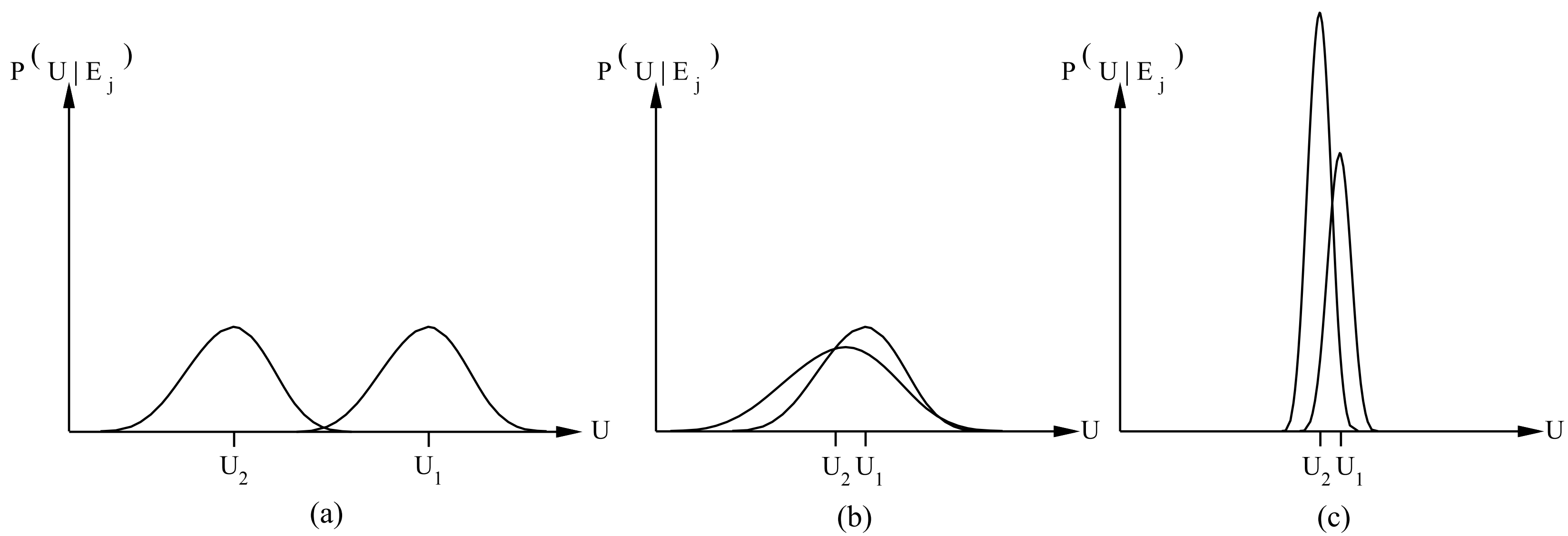
$$VPI_E(E_j, E_k) = VPI_E(E_j) + VPI_{E,E_j}(E_k) = VPI_E(E_k) + VPI_{E,E_k}(E_j)$$

Note: when more than one piece of evidence can be gathered, maximizing VPI for each to select one is not always optimal

⇒ evidence-gathering becomes a sequential decision problem

Qualitative behaviors

a) Choice is obvious, information worth little
b) Choice is nonobvious, information worth a lot
c) Choice is nonobvious, information worth little



Unknown Preferences

Imagine that you are at an ice-cream shop in Thailand and they have only two flavors left: vanilla and durian. Both cost $2. You know you have a moderate liking for vanilla and you'd be willing to pay up to $3 for a vanilla ice cream on such a hot day, so there is a net gain of $1 for choosing vanilla. On the other hand, you have no idea whether you like durian or not, but you've read on Wikipedia that the durian elicits different responses from different people: some find that "it surpasses in flavour all other fruits of the world" while others liken it to "sewage, stale vomit, skunk spray and used surgical swabs."

Unknown Preferences

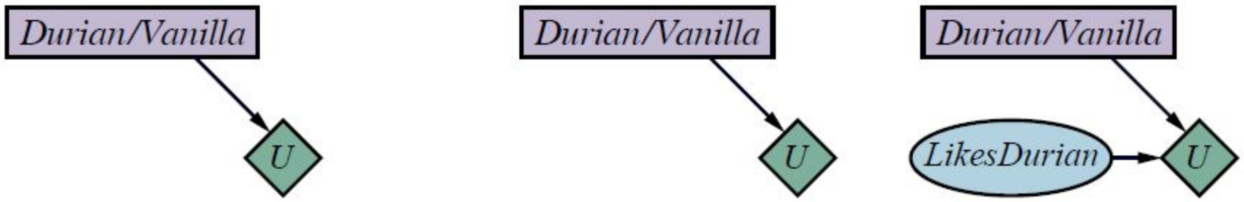
let’s say there’s a 50% chance you’ll find it sublime (+$100) and a 50% chance you’ll hate it (-$80 if the taste lingers all afternoon).

We can simply replace the uncertain value of the durian with its expected net gain below however the decision will still not changed.

$$(0.5 \times \$100) - (0.5 \times \$80) - \$2 = \$8$$

Rather than saying there is uncertainty about the utility function, we move that uncertainty “into the world,” so to speak. That is, we create a new random variable LikesDurian with prior probabilities of 0.5 for true and false in (c)

Unknown Preferences



D/V	⋯	−\$82	⋯	+\$0	+\$1	⋯	+\$98	⋯
durian	⋯	0.5	⋯	0.0	0.0	⋯	0.5	⋯
vanilla	⋯	0.0	⋯	0.0	1.0	⋯	0.0	⋯

(a)

D/V	*U*
durian	+\$8
vanilla	+\$1

(b)

D/V	*LikesDurian*	U
durian	*true*	+\$98
durian	*false*	−\$82
vanilla	*true*	+\$1
vanilla	*false*	+\$1

(c)

(a) A decision network for the ice cream choice with an uncertain utility function.
(b) The network with the expected utility of each action.
(c) Moving the uncertainty from the utility function into a new random variable

Pearson

Summary

- **Probability theory** describes what an agent should believe on the basis of evidence, **utility theory** describes what an agent wants, and **decision theory** puts the two together to describe what an agent should do.

- **Utility theory** shows that an agent **whose preferences** between lotteries are consistent with a set of simple axioms can be described as possessing a utility function

- **Multiattribute utility theory** deals with utilities that depend on several distinct attributes of states.

- **Stochastic dominance** is a particularly useful technique for making unambiguous decisions, even without precise utility values for attributes.

- **Decision networks** provide a simple formalism for expressing and solving decision problems.

- The **value of information** is defined as the expected improvement in utility compared with making a decision without the information