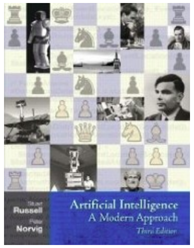
Decision Making

- Rational preferences
- Utilities
- Money
- Multiattribute utilities
- Decision networks
- Value of information

Material from Russell & Norvig, chapter 16



Many slides taken from Russell & Norvig's slides *Artificial Intelligence: A Modern Approach*

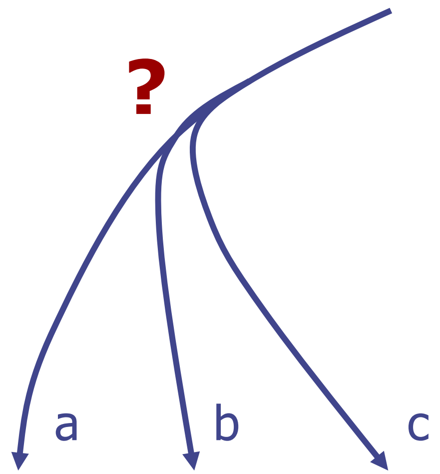
Some based on Slides by Lise Getoor, Jean-Claude Latombe and Daphne Koller

Decision Making under Uncertainty

- Many environments are uncertain in the sense that it is not clear what state an action will lead to
  - **Uncertainty:** Some states may be likely, others may be unlikely
  - **Utility:** Some states may be desirable, others may be undesirable
- Still, an agent has to make a decision which action to choose

→ **Decision Theory** is concerned with making rational decisions in such scenarios

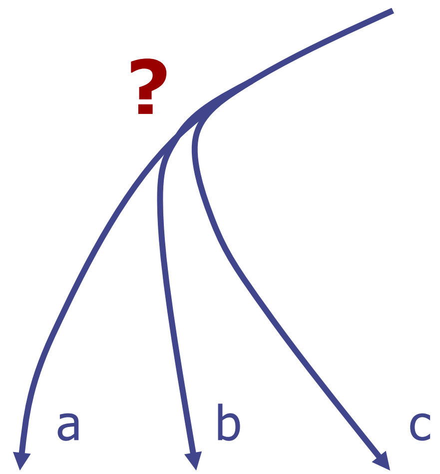
Non-Deterministic vs. Probabilistic Uncertainty



{a,b,c}

→ decision that is best for worst case

Non-deterministic model



[P(a),a; P(b),b; P(c),c]

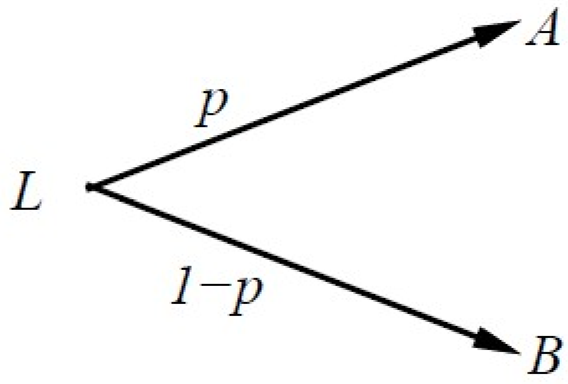
→ decision that maximizes expected utility value

Probabilistic model

~ Adversarial search

Lotteries and Preferences

- In the following, we call such probabilistic events lotteries
  - A lottery consists of a set of events (prizes) with their probabilities



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

- Preferences:
  - An agent likes certain prizes better than others
  - An agent therefore also likes certain lotteries better than others

Notation:

$A \succ B$	A preferred to B
$A \sim B$	indifference between A and B
$A \succsim B$	B not preferred to A

Preferences and Rational Behavior

- Preferences between prizes may, in principle, be arbitrary
- For example, preferences may be cyclic

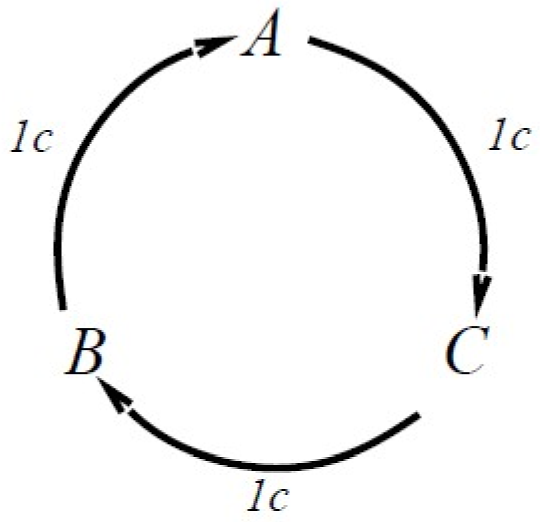
$$A \succ B \,, B \succ C \,, C \succ A$$

- However, cyclic preferences lead to irrational behavior:

If $B \succ C$, then an agent who has C would pay (say) 1 cent to get B

If $A \succ B$, then an agent who has B would pay (say) 1 cent to get A

If $C \succ A$, then an agent who has A would pay (say) 1 cent to get C



→ Eventually the agent will give away all its money

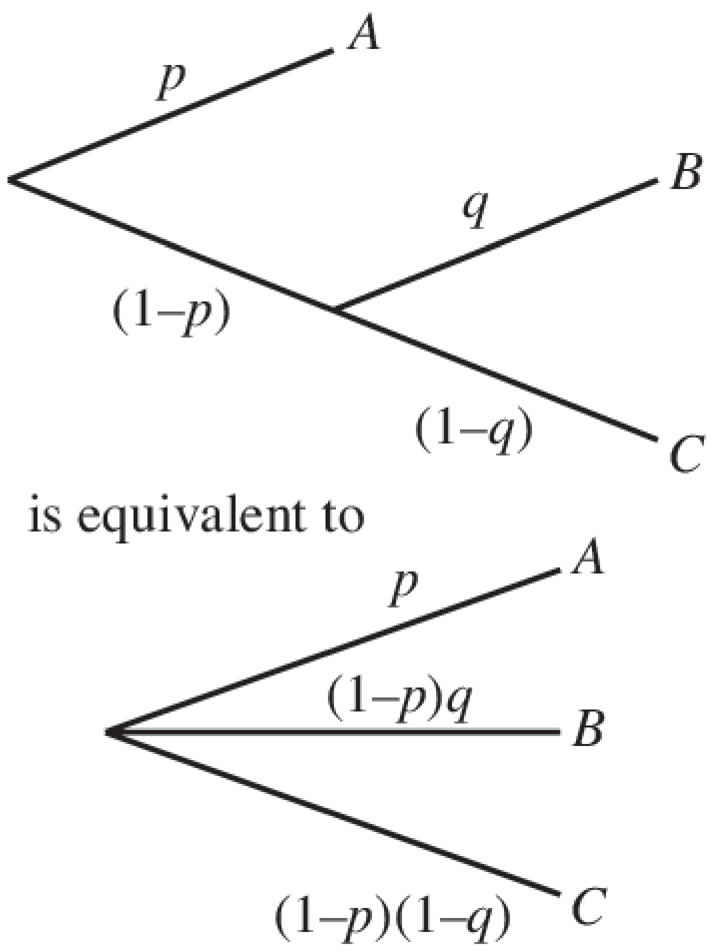
Preferences and Rational Behavior

- Another property that should be obeyed is that lotteries are decomposable
- Therefore, no rational agent should have a preference between the two equivalent formulations

$$[p,\mathrm{A};\ 1\text{-}p,\ [q,\mathrm{B};\ 1\text{-}q,\ \mathrm{C}]] \sim$$
$$[p,\mathrm{A};\ (1\text{-}p)q,\ \mathrm{B};\ (1\text{-}p)(1\text{-}q),\ \mathrm{C}]$$

- Such properties be formulated as constraints on preferences

- Decomposability



Other Constraints for Rational Behavior

- Together with Decomposobality, these constraints define a rational behavior:

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

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

Utility functions

- A natural way for measuring how desirable certain prizes are is using a utility function U
  - A utility function assigns a numerical value to each prize
- Utility function naturally lead to preferences

$$A \succ B \Leftrightarrow U(A) > U(B)$$

- The Expected Utility of an Event is the expected value of the utility function in a lottery

$$EU(X) = \sum_{x \in X} P(x) \cdot U(x)$$

- A utility function in a deterministic environment (no lotteries) is also called a value function

Maximizing Utilities

- It has been shown that acting according to rational preferences corresponds to maximizing a utility function U

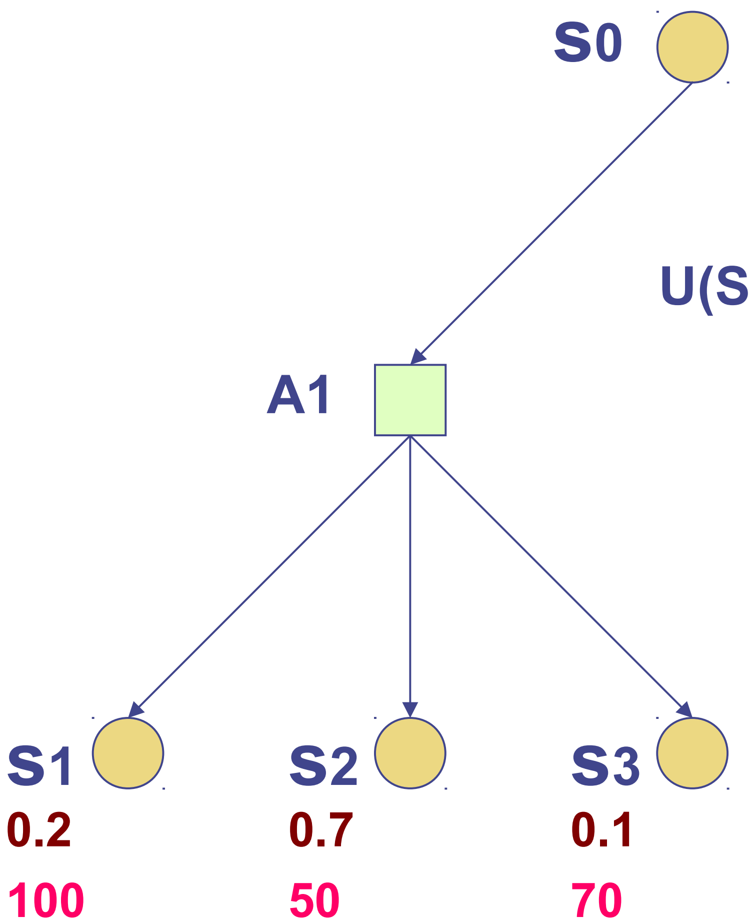
Theorem (Ramsey, 1931; von Neumann and Morgenstern, 1944):
Given preferences satisfying the constraints
there exists a real-valued function U such that

$$U(A) \geq U(B) \quad \Leftrightarrow \quad A \succsim B$$

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

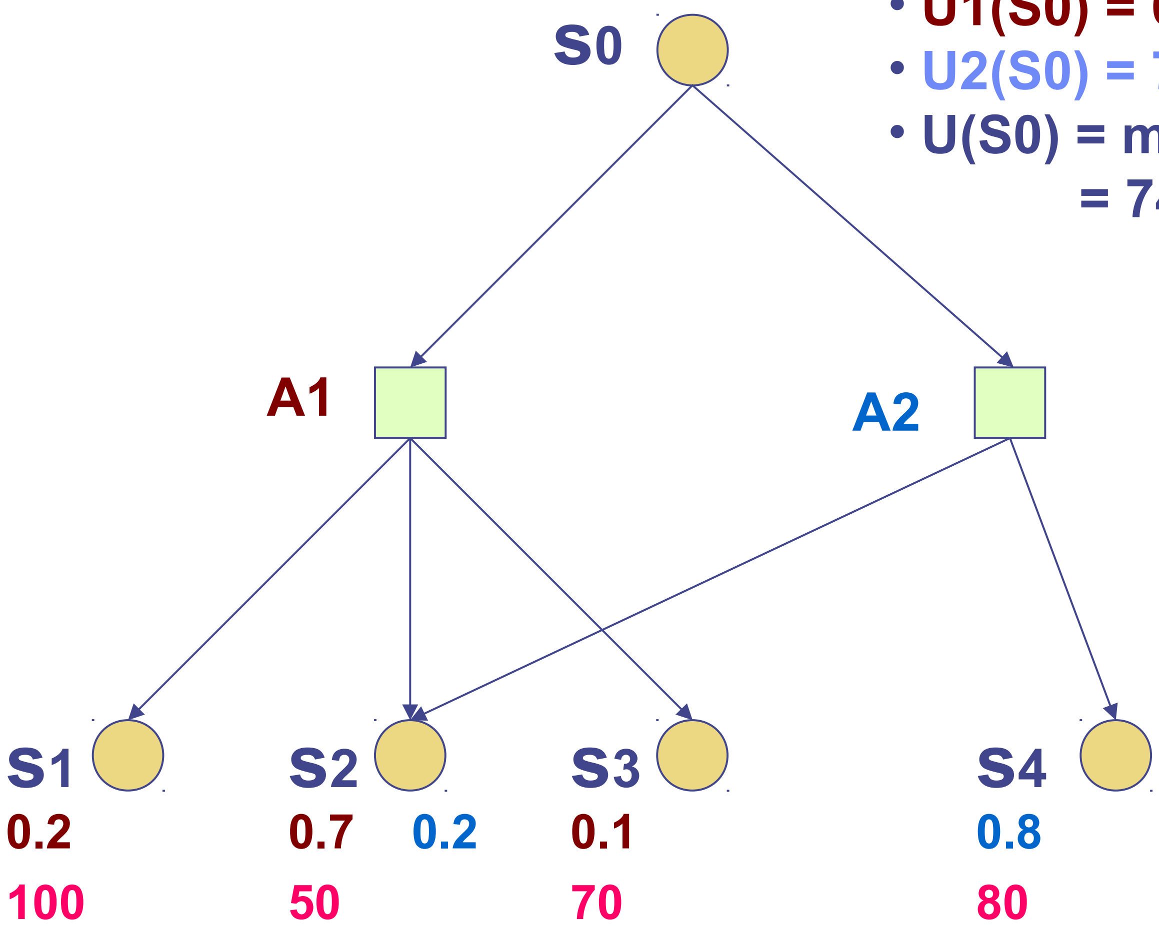
- Maximizing Expected Utility (MEU) principle
  - An agent acts rationally if it selects the action that promises the highest expected utility
- Note that an agent may act rationally without knowing U or the probabilities!
  - e.g., according to pre-compiled look-up tables for optimal actions

Example: Expected Utility of an Action



U(S0) = 100 x 0.2 + 50 x 0.7 + 70 x 0.1
= 20 + 35 + 7
= 62

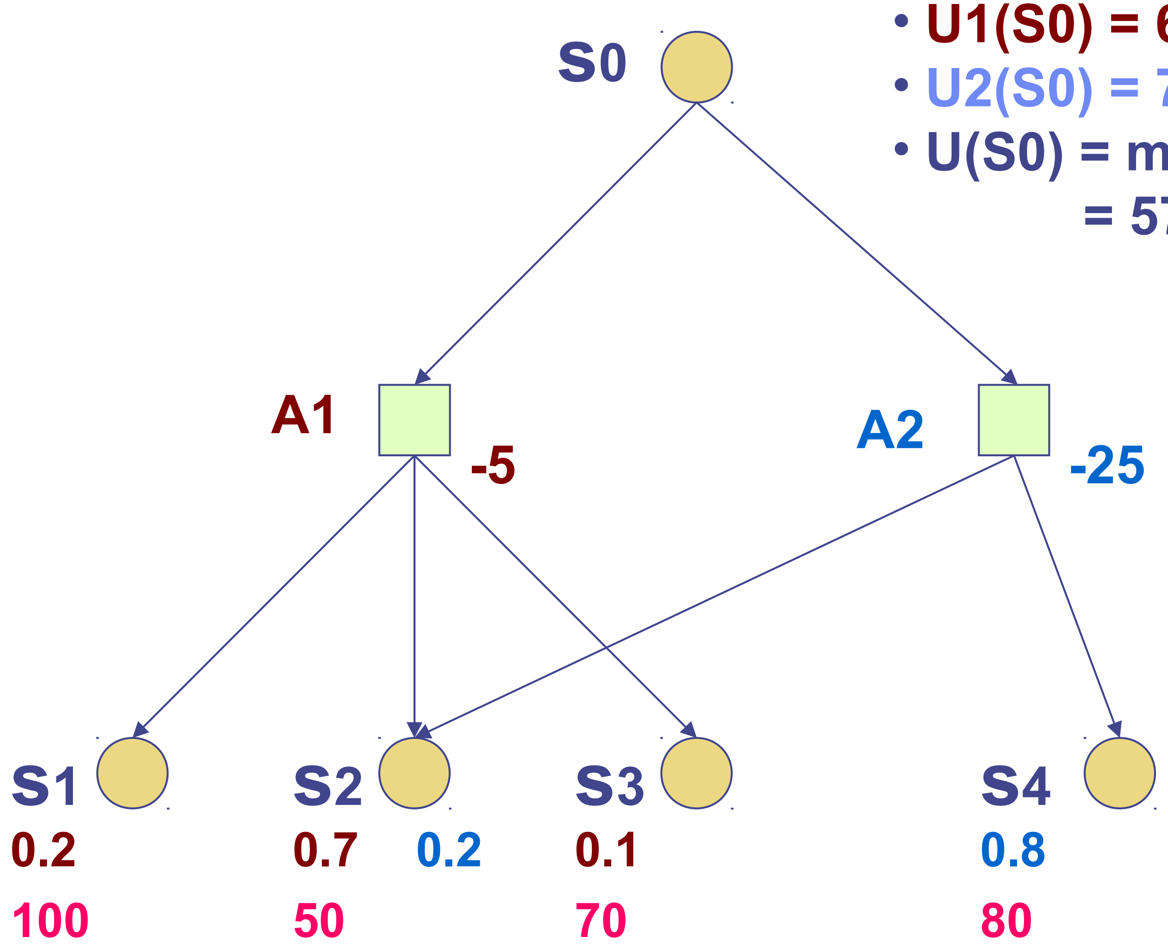
Example: Choice between 2 Actions



- U1(S0) = 62
- U2(S0) = 74
- U(S0) = max{U1(S0),U2(S0)} = 74

→ The agent should pick action 2.

Example: Adding Action Costs



- U1(S0) = 62 – 5 = 57
- U2(S0) = 74 – 25 = 49
- U(S0) = max{U1(S0),U2(S0)} = 57

→ The agent should pick action 1.

MEU Principle

- A **rational agent** should choose the action that maximizes agent's expected utility
- This is the basis of the field of **decision theory**
- The MEU principle provides a **normative criterion** for rational choice of action

Do we now have a working definition of rational behavior?
And therefore solved AI?

Not quite…

- Must have **complete** model of:
  - Actions
  - Utilities
  - States
- Even if you have a complete model, it will be computationally **intractable**
- In fact, a truly rational agent takes into account the utility of reasoning as well – **bounded rationality**
- Nevertheless, great progress has been made in this area recently, and we are able to solve much more complex decision-theoretic problems than ever before

Decision Theory vs. Reinforcement Learning

- Simple decision-making techniques are good for selecting the best action in simple scenarios
- → Reinforcement Learning is concerned with selecting the optimal action in Sequential Decision Problems
  - Problems where a sequence of actions has to be taken until a goal is reached.

How to measure Utility?

An obvious idea: Money

- However, Money is not the same as utility

Example:

- If you just earned 1,000,000$, are you willing to bet them on a double-or-nothing coin flip?
- How about triple or nothing?

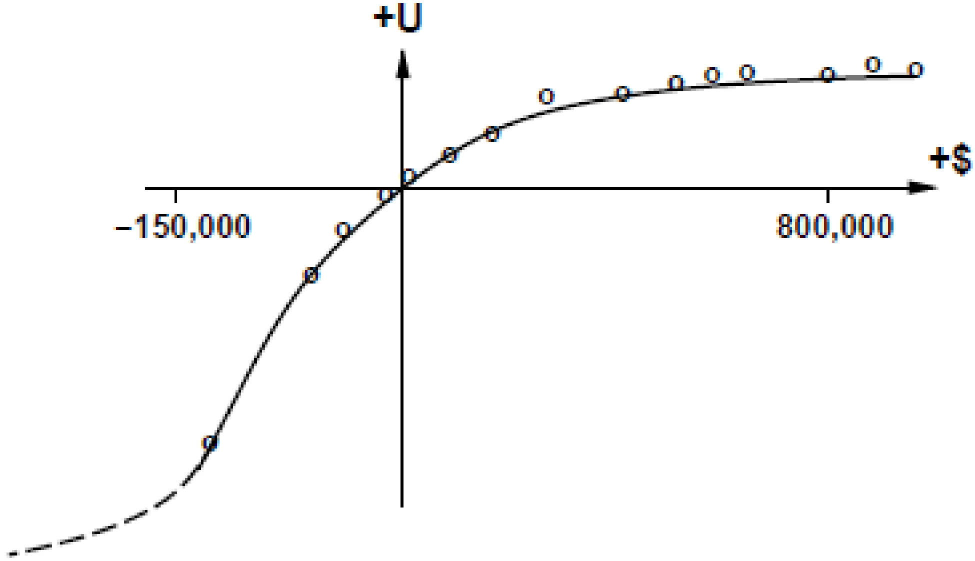
$$U(1{,}000{,}000) > EU([0.5, 0 ; 0.5, 3{,}000{,}000])?$$

$$U(1{,}000{,}000) > 0.5 \cdot U(0) + 0.5 \cdot U(3{,}000{,}000)?$$

Most people would grab a million and run, although the expected value of the lottery is 1.5 million

The Utility of Money

- Grayson (1960) found that the utility of money is almost exactly proportional to its logarithm
- One way to measure it:
  - Which is the amount of money for which your behavior between „grab the money“ changes to „play the lottery“?
  - Obviously, this also depends on the person i
    - if you already have 50 million, you are more likely to gamble...
- Utility of money for a certain Mr. Beard:



Risk-Averse vs. Risk-Seeking

- People like Mr. Beard are risk-averse
  - Prefer to have the expected monetary value of the lottery ($EMV(L)$) handed over than to play the lottery L

$$U(L) < U(S_{EMV(L)})$$

- Other people are risk-seeking
  - Prefer the thrill of a wager over secure money

$$U(L) > U(S_{EMV(L)})$$

- For risk-neutral people, the Utility function is the identity

$$U(L) = U(S_{EMV(L)})$$

- The difference $U(L) - U(S_{EMV(L)})$ is called the **insurance premium**. This is the business model of insurances

General Approach to Assessing Utilities

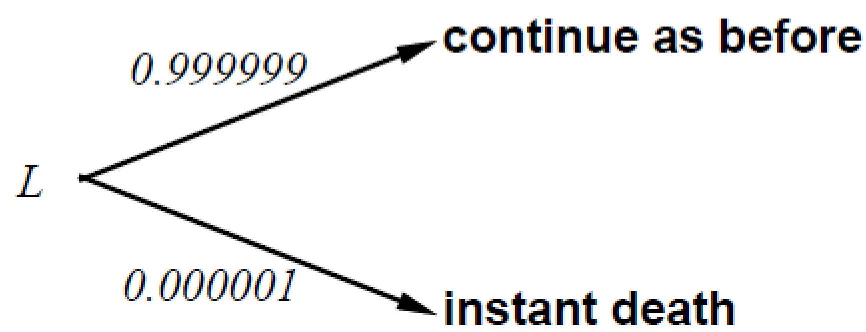
- Find probablity p so that the expected value of a lottery between two extremes corresponds to value of the prize A

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 $A \sim L_p$

- Normalized utility scales interpolate $u_\top = 1.0$, $u_\bot = 0.0$
  - Normalization does not change the behavior of an agent, because (positive) linear transformations $U'(S) = k_1 + k_2 \cdot U(S)$ leave the ordering of actions unchanged
  - If there are no lotteries, any monotonic transformation leaves the preference ordering of actions unchanged

Other Units of Measurements for Utilities

- In particular for medicine and safety-critical environments, other proposals have been made (and used)
- Micromorts:
  - A micromort is the lottery of dying with a probability of one in a million



  - It has been established that a micromort is worth about $50.
    - Does not mean that you kill yourself for $50,000,000 (we have already seen that utility functions are not linear...)
  - Used in safety-critical scenarios, car insurance, ...
- Quality-Adjusted Life Year (QALY)
  - A year in good health, used in medical applications

Multi-Attribute Utilities

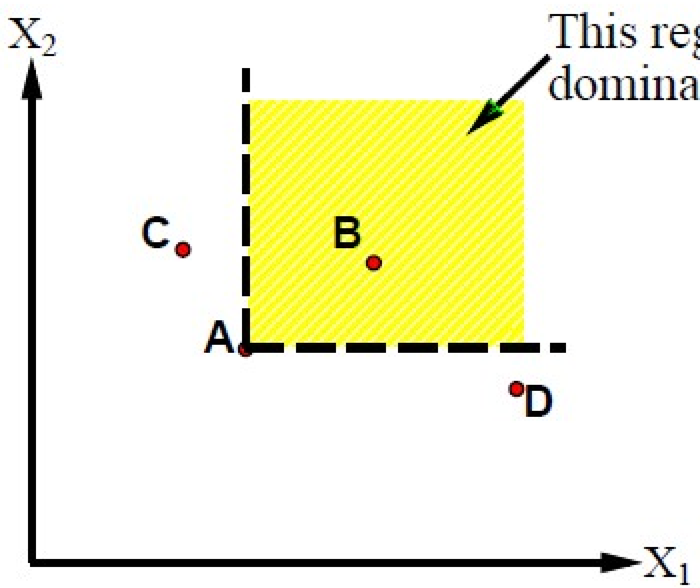
- Often, the utility does not depend on a single value but on multiple values simultaneously
- Example: Utility of a car depends on
  - Safety
  - Horse-Power
  - Fuel Consumption
  - Size
  - Price
- How can we reason in this case?
  - It is often hard to define a function that maps multiple dimensions X_i to a single utility value $U(X_1, X_2, \ldots, X_n)$

    → Dominance is a useful concept in such cases

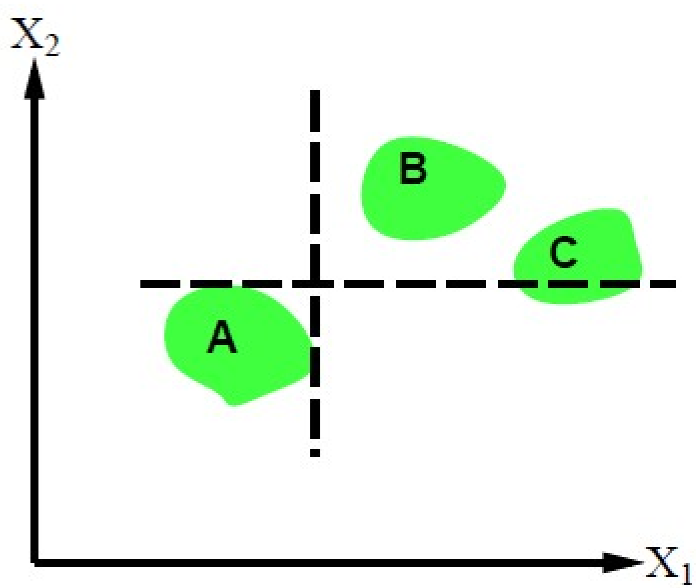
Strict Dominance

- Scenario A is better than Scenario B if it is better along all dimensions
- Example:
  - 2 dimensions, in both dimensions higher is better (utility grows monotonically with the value)

$$A \succ B \Leftrightarrow U(X_A, Y_A) \geq U(X_B, Y_B) \Leftrightarrow (X_A \geq X_B) \wedge (Y_A \geq Y_B)$$



Deterministic attributes

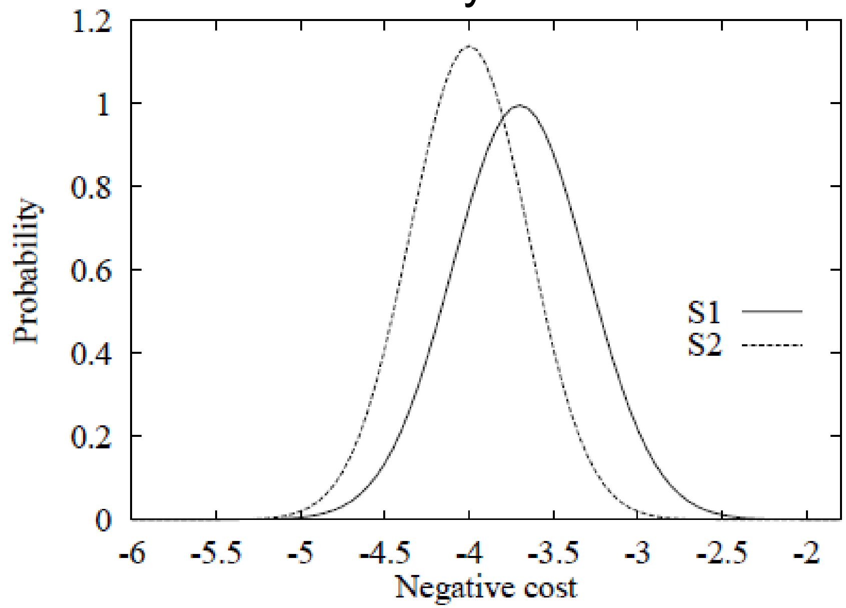


Uncertain attributes

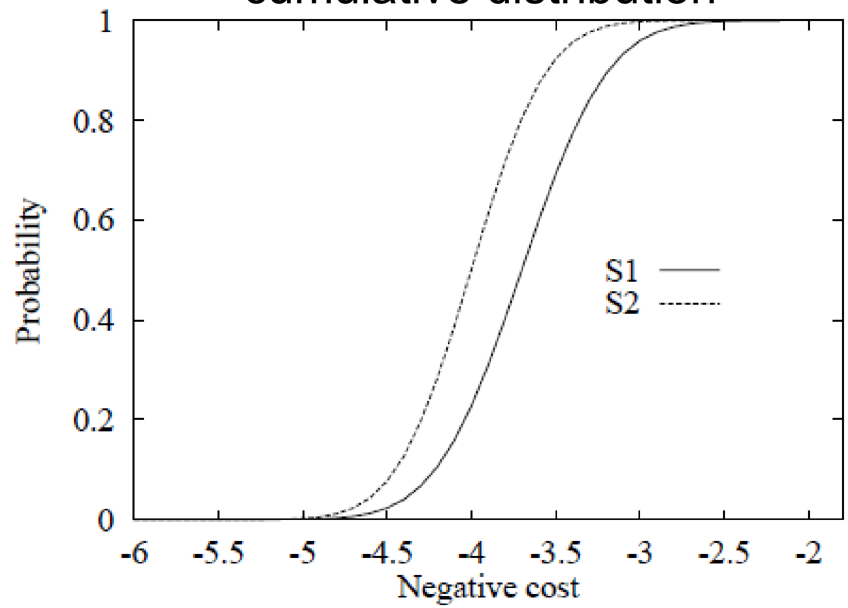
Stochastic Dominance

- Strict dominance rarely occurs in practice
  - The car that is better in horse-power is rarely also better in fuel consumption and price
- Stochastic dominance:
  - A utility distribution p_1 dominates utility distribution p_2 if the probability of having a utility less or equal a given threshold (cumulative probability) is always lower for p_1 than for p_2

density function



cumulative distribution



Stochastic Dominance

- If two actions A_1 and A_2 lead to probability distributions $p_1(\mathrm{x})$ and $p_2(\mathrm{x})$ on attribute X, then

A_1 stochastically dominates A_2 on attribute X iff

$$\forall x \in X \int_{-\infty}^{x} p_1(x')\,\mathrm{d}x' \leq \int_{-\infty}^{x} p_2(x')\,\mathrm{d}x'$$

Definition of stochastic dominance for probability distributions

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

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

  - because high utility values have a higher probability in p_1
- Extension for Multiple attributes:
  - If there is stochastic dominance along all attributes, then action A_1 dominates A_2

Assessing Stochastic Dominance

- It may seem that stochastic dominance is a concept that is hard to grasp and hard to measure
- But actually it is often quite intuitive and can be established without knowing the exact distribution using qualitative reasoning
- Examples:
  - Construction costs for large building will increase with the distance from the city
    - For higher costs, the probability of such costs are larger for a site further away from the city than for a site that is closer to the city
  - Degree of injury increases with collision speed

Preference (In-)Dependence

- As with probability distribution, it may be hard to establish the utility for all possible value combinations of a multi-attribute utility function $U(X_1, X_2, \dots, X_n)$
- Again, we can simplify things by introducing a notion of dependency
  - Attribute X_1 is preference-independent of attribute X_2 if knowing X_1 does not influence our preference in X_2
- Examples:
  - Drink preferences depend on the choice of the main course
    - For meat, red wine is preferred over white wine
    - For fish, white wine is preferred over red wine
  - Table preferences do not depend on the choice of the main course
    - A quiet table is always preferred, no matter what is ordered

Mutual Preference Independence

- A set of variables is mutually preferentially independent if each subset of variables is preferentially independent of its complement
  - Can be established by checking only attribute pairs (Leontief, 1947)
- If variables are mutually preferentially independent, the value function can be decomposed

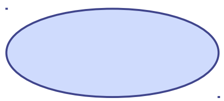
Theorem (Debreu, 1960): mutual P.I. $\Rightarrow$ $\exists$ additive value function:

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

- **Note:**
  - This only holds for deterministic environments (value functions). For stochastic environments (utility functions), establishing utility-independence is more complex

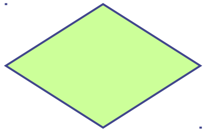
Decision Networks

- Extend BNs to handle actions and utilities and enable rational decision making



  - Chance nodes: random variables, as in BNs

  - Decision nodes: actions that decision maker can take



  - Utility/value nodes: the utility of the outcome state.

- Use BN inference methods to solve

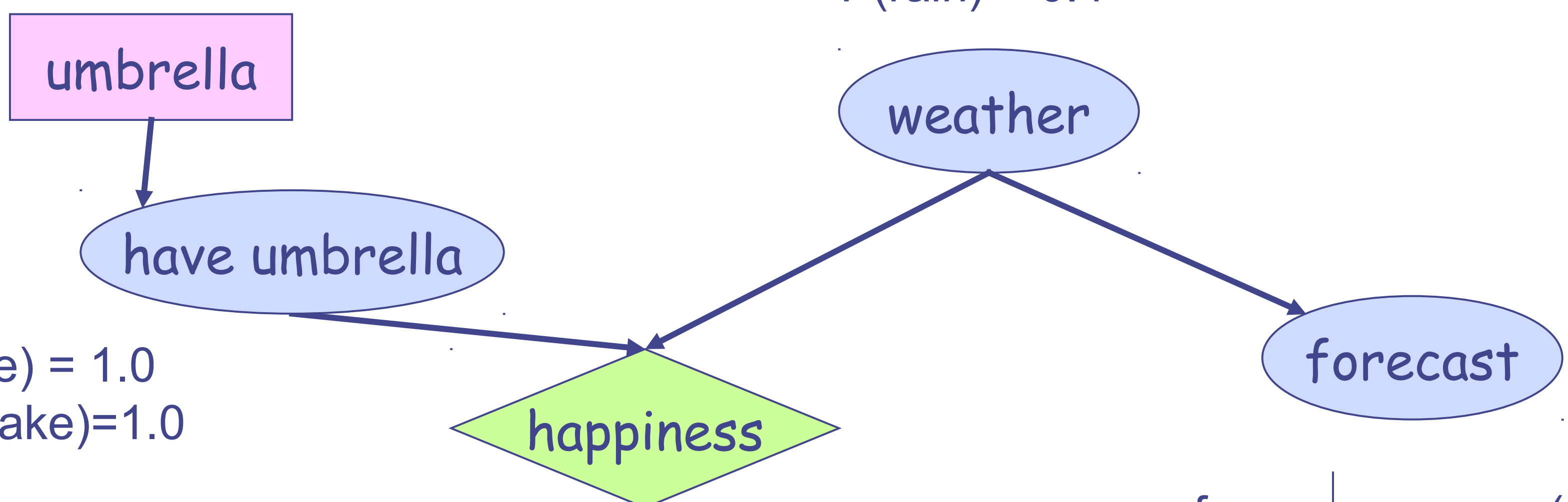
```
For each value of action node
    compute expected value of utility node given action, evidence
Return MEU action
```

Example: Umbrella Network

- Should I take an umbrella to increase my happiness?

take/don’t take

P(rain) = 0.4



P(have|take) = 1.0
P(~have|~take)=1.0

U(have,rain) = -25
U(have,~rain) = 0
U(~have, rain) = -100
U(~have, ~rain) = 100

f	w	p(f|w)
sunny	rain	0.3
rainy	rain	0.7
sunny	no rain	0.8
rainy	no rain	0.2

Evaluating Decision Networks

- Set the evidence variables for current state
- For each possible value (action) of the decision node:
  - Set decision node to that value
  - Calculate the posterior probability of the parent nodes of the utility node, using BN inference
  - Calculate the resulting utility for action
- Return the action with the highest utility
- In the Umbrella example:

$$EU(\text{take}) = 0.4 \times -25 + 0.6 \times 0 = -10$$

$$EU(\neg\text{take}) = 0.4 \times -100 + 0.6 \times 100 = +20$$

→ My expected utility is higher if I don't take the umbrella

- Note that we did not take the weather Report into account!
  - Would it be worth to get that information?

Value of Information

- Decision Networks allow to measure the value of information

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

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

Value of Perfect Information (VPI)

- General Idea:
  - Compute the Expected Utility of an action without the evidence
  - Compute the Expected Utility of the action over all possible outcomes of the evidence
  - The difference is the value of knowing the evidence.
- More formally
  - current evidence E, α is best of actions A

$$EU(\alpha|E) = max_A \sum_i U(\text{Result}_i(A)) \cdot P(\text{Result}_i(A)|Do(A), E)$$

  - after obtaining new evidence E_j

$$EU(\alpha|E, E_j) = max_A \sum_i U(\text{Result}_i(A)) \cdot P(\text{Result}_i(A)|Do(A), E, E_j)$$

  - Difference between expected value over all possible outcomes e_{jk} of E_j and the expected value without E_j

$$VPI_E(E_j) = \left(\sum_k P(E_j = e_{jk}|E) \cdot EU(\alpha_{e_{jk}}|E, E_j = e_{jk})\right) - EU(\alpha|E)$$

Properties of VPI

Nonnegative—in **expectation**, not **post hoc**

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

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

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

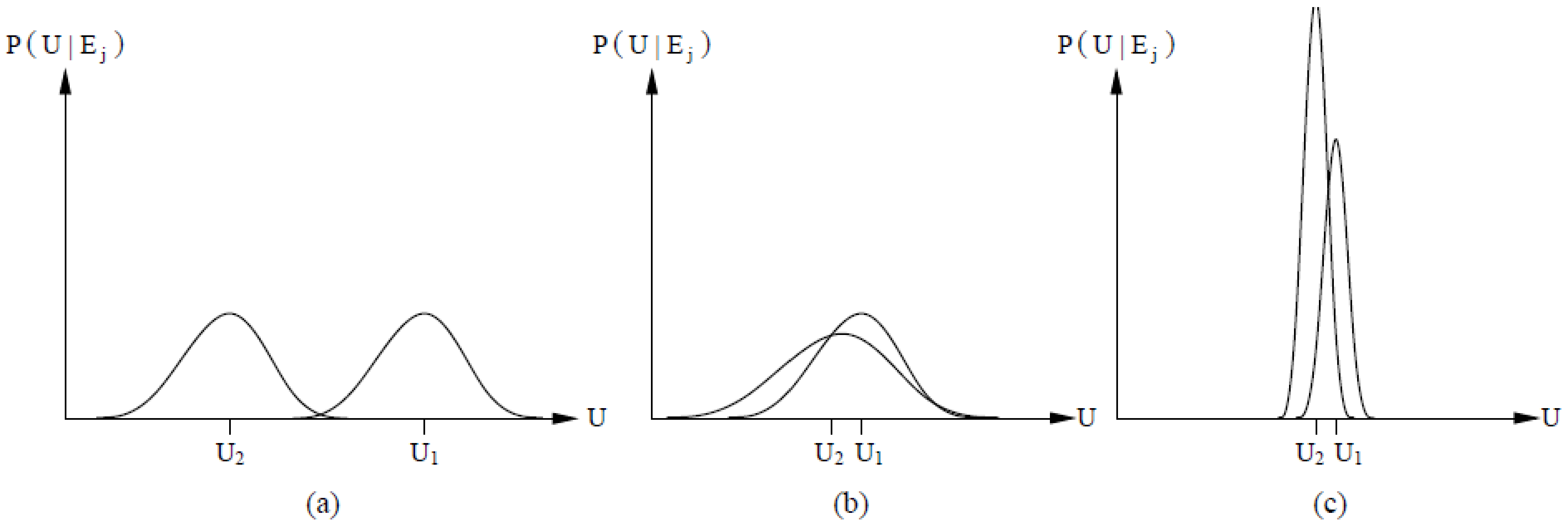
Order-independent

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

Note: when more than one piece of evidence can be gathered, maximizing VPI for each to select one is not always optimal
$\Rightarrow$ evidence-gathering becomes a **sequential** decision problem

Qualitative Behaviors

- The value of information depends on the distribution of the new utility values in dependence of their old estimates



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

Real-World Decision Network