Lecture 6: Rational Decisions (Simple Decisions)

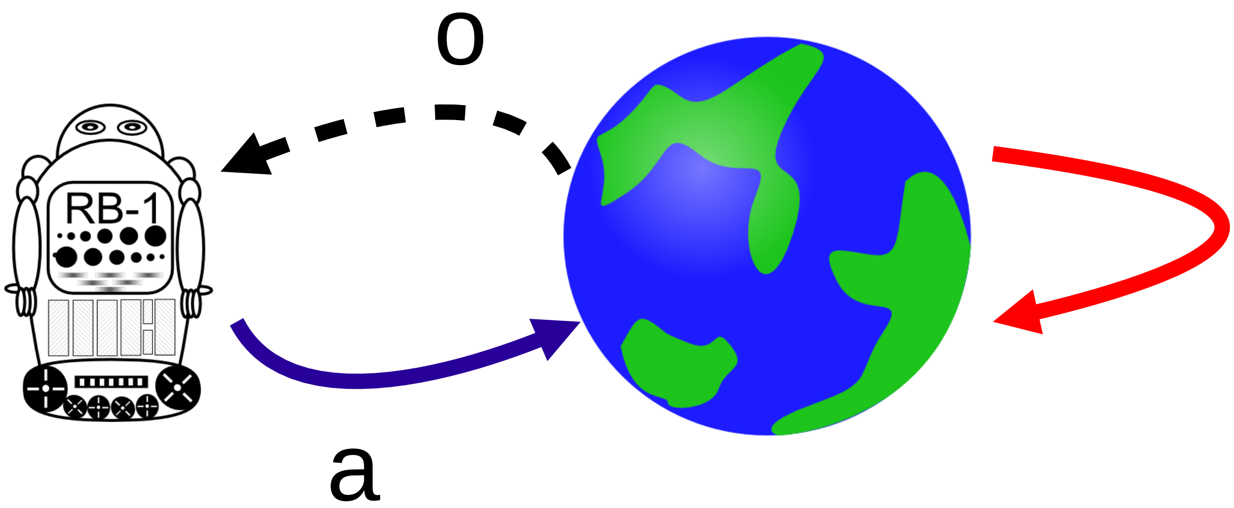
Matthijs Spaan

(based on slides by Catholijn M. Jonker)

Russell & Norvig
Chapter 15 (sections 1-3, 5, 6)

21/09/2023

Intelligent decision making



- What do I want?
- How do I make that happen?

- Decision theory = utility theory + probability theory

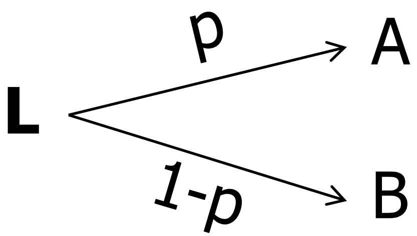
Contents of these slides

Making Simple Decisions (Chapter 15)

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

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

Rational preferences

- Idea: preferences of a rational agent must obey constraints.
- Rational preferences ⇒ behavior describable as maximization of expected utility

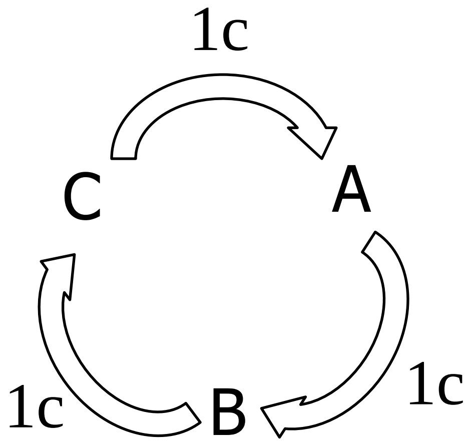
Constraints:

- 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: $A \succ B \succ C \Rightarrow \exists p\, [p, A; 1\text{-}p, C] \sim B$
- Substitutability: $A \sim B \Rightarrow [p, A; 1\text{-}p, C] \sim [p, B; 1\text{-}p, C]$
- Monotonicity: $A \succ B \Rightarrow (p \geq q \Leftrightarrow [p, A; 1\text{-}p, B] \succsim [q, A; 1\text{-}q, B])$

Rational preferences

- 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 $B \succ C$, 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



Lottery - Utilities

S_n: the state of possessing n€,
Current wealth is k€
p the chance of winning the lottery:

EU(accept) = (1-p) * U(S_k) + p * U($S_{k+3M€}$)
EU(decline) = U($S_{k+1M€}$)

U: first million is worth more than later millions.

Lottery - Utilities

- Accept means play lottery L = [p, 3M€; (1 – p), 0€]
- Decline means you walk away with 1M€

- Suppose U(Sk) = 0.5
- U(Sk+1M€) = 0.8
- U(Sk+3M€) = 1

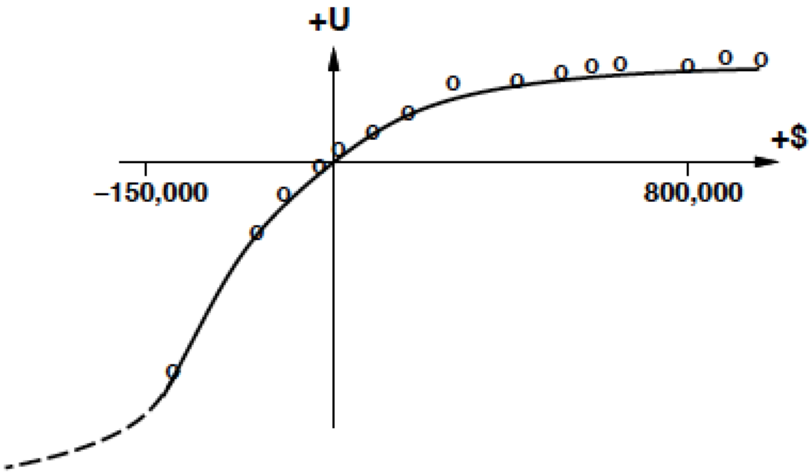
- Then EU(accept) = 0.75
- EU(decline) = 0.8

Lottery - Utilities

Suppose $U(S_k) = 0.5$ $U(S_{k+1M\text{€}}) = 0.8$ $U(S_{k+3M\text{€}}) = 1$ Then EU(accept) = 0.75 EU(decline) = 0.8	Suppose $U(S_k) = 0.8$ $U(S_{k+1M\text{€}}) = 0.9$ $U(S_{k+3M\text{€}}) = 1$ Then EU(accept) = 0.9 EU(decline) = 0.9

The impact of money on people

- 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:



Maximizing expected utility (MEU)

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

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

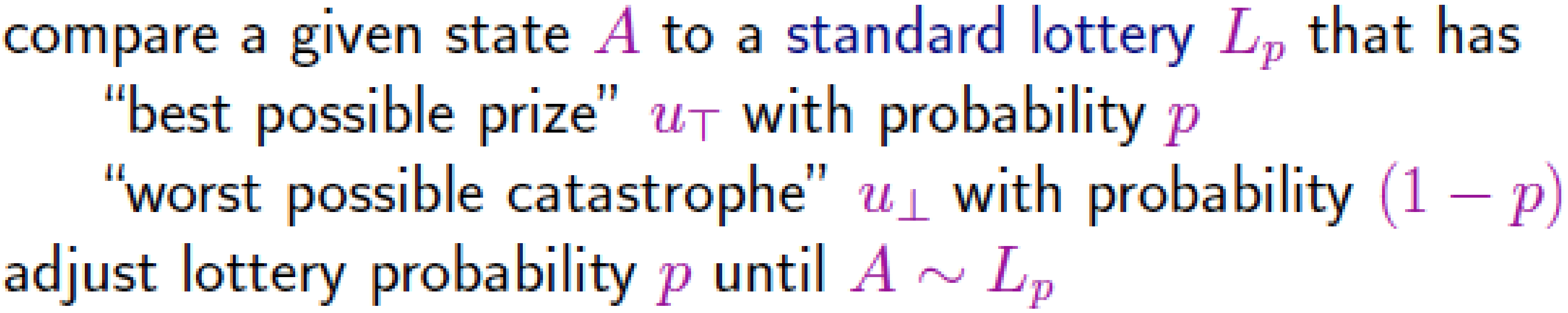
- MEU principle:

Choose the action that maximizes expected utility!

- Note: an agent can be entirely rational (consistent with MEU) without ever representing utilities and probabilities (E.g., a lookup table for perfect tictactoe)

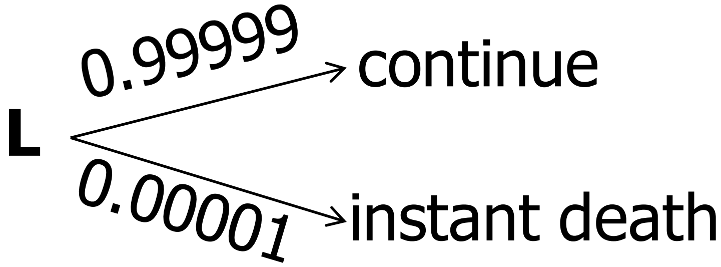
Utilities

- 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: $u_\top = 1.0,\ u_\perp = 0.0$

- Note: behavior is invariant w.r.t. positive linear transformation

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

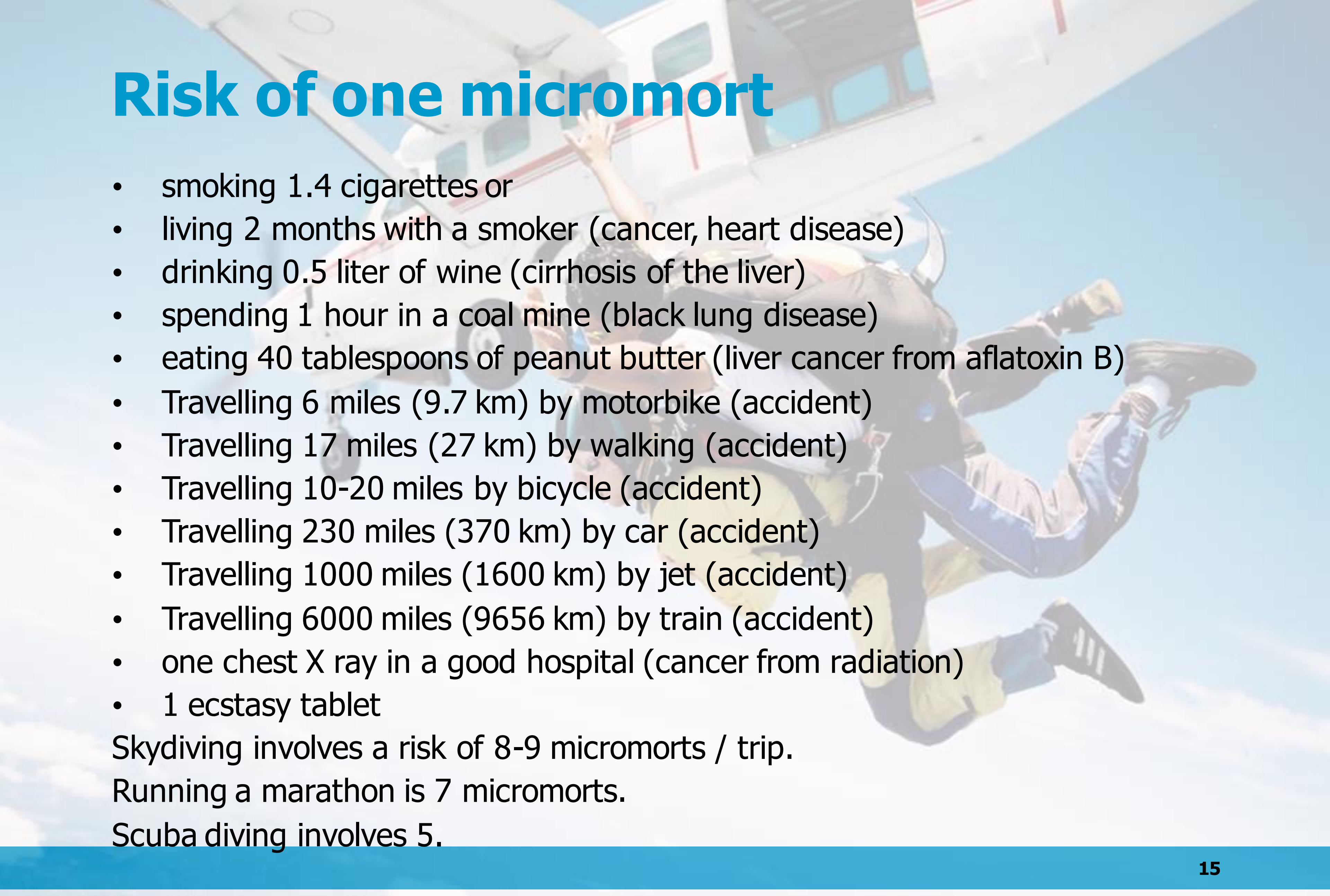
$$U'(x) = k_1 U(x) + k_2 \quad \text{where } k_1 > 0$$

- Micromorts: one-millionth chance of death, useful for Russian roulette, paying to reduce product risks, etc.

- QALYs: quality-adjusted life years

Micromorts

- A micromort is a unit of risk measuring a one-in-a-million probability of death.
- An application of micromorts is measuring the value that humans place on risk:
  - What is the amount of money one would have to pay a person to get him or her to accept a one-in-a-million chance of death?
  - Or what amount is someone willing to pay to avoid a one-in-a-million chance of death?
- When put thus people claim a high number but when inferred from their day-to-day actions (e.g., how much they are willing to pay for safety features on cars) a typical value is around $20.



Risk of one micromort

- smoking 1.4 cigarettes or
- living 2 months with a smoker (cancer, heart disease)
- drinking 0.5 liter of wine (cirrhosis of the liver)
- spending 1 hour in a coal mine (black lung disease)
- eating 40 tablespoons of peanut butter (liver cancer from aflatoxin B)
- Travelling 6 miles (9.7 km) by motorbike (accident)
- Travelling 17 miles (27 km) by walking (accident)
- Travelling 10-20 miles by bicycle (accident)
- Travelling 230 miles (370 km) by car (accident)
- Travelling 1000 miles (1600 km) by jet (accident)
- Travelling 6000 miles (9656 km) by train (accident)
- one chest X ray in a good hospital (cancer from radiation)
- 1 ecstasy tablet

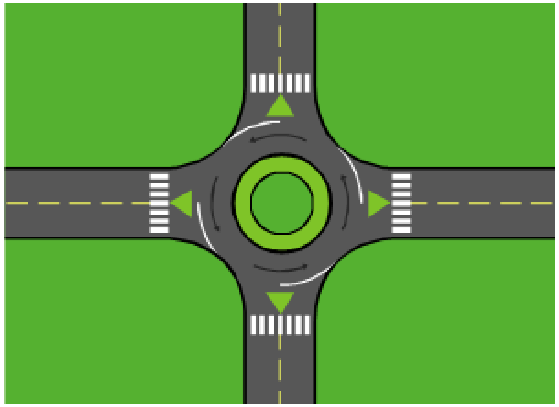
Skydiving involves a risk of 8-9 micromorts / trip.

Running a marathon is 7 micromorts.

Scuba diving involves 5.

Quality Adjusted Life Years (QALYs)

- The quality-adjusted life year (QALY) is a measure of disease burden, including both the quality and the quantity of life lived. It is used in assessing the value for money of a medical intervention.



Cost of construction

vs.

Nr. of deaths on crossing

Quality Adjusted Life Years (QALYs)

- The QALY model requires utility independent, risk neutral, and constant proportional tradeoff behaviour.

- The QALY is based on the number of years of life that would be added by the intervention.

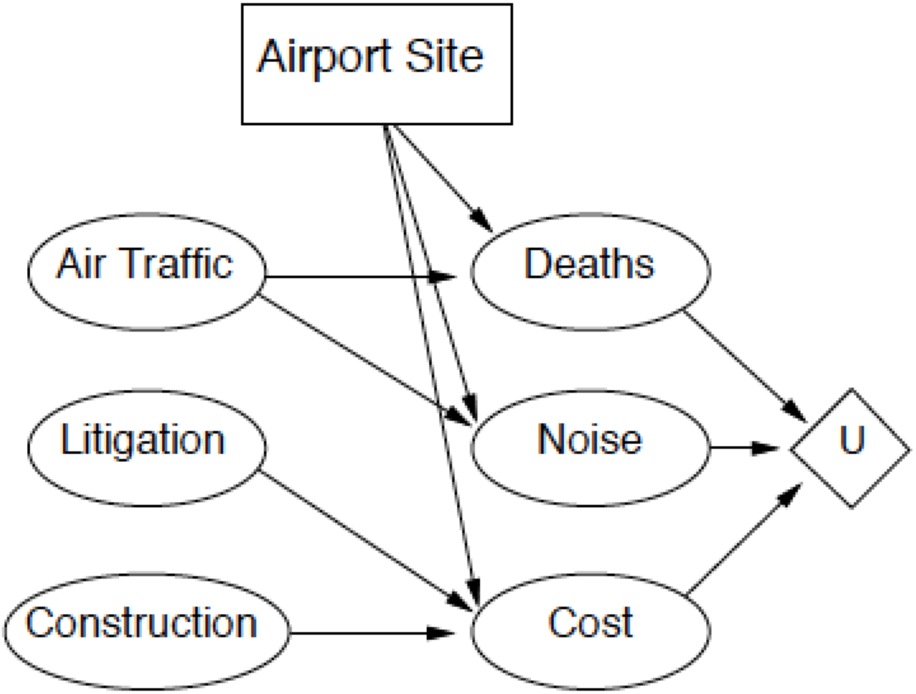
- Each year in perfect health is assigned the value of 1.0 down to a value of 0.0 for death. If the extra years would not be lived in full health, for example if the patient would lose a limb, or be blind or have to use a wheelchair, then the extra life-years are given a value between 0 and 1 to account for this.

QALYs use

- Cost-utility analysis : intervention cost / QALYs saved
- Used to allocate healthcare resources
- Controversial: some people will not receive treatment as it is calculated that cost of the intervention is not warranted by the benefit to their quality of life.
- Argument in favor: health care resources are limited, this allocation method is approximately optimal for society, including most patients.

Decision networks

- Add action nodes and utility nodes to belief networks to enable rational decision making

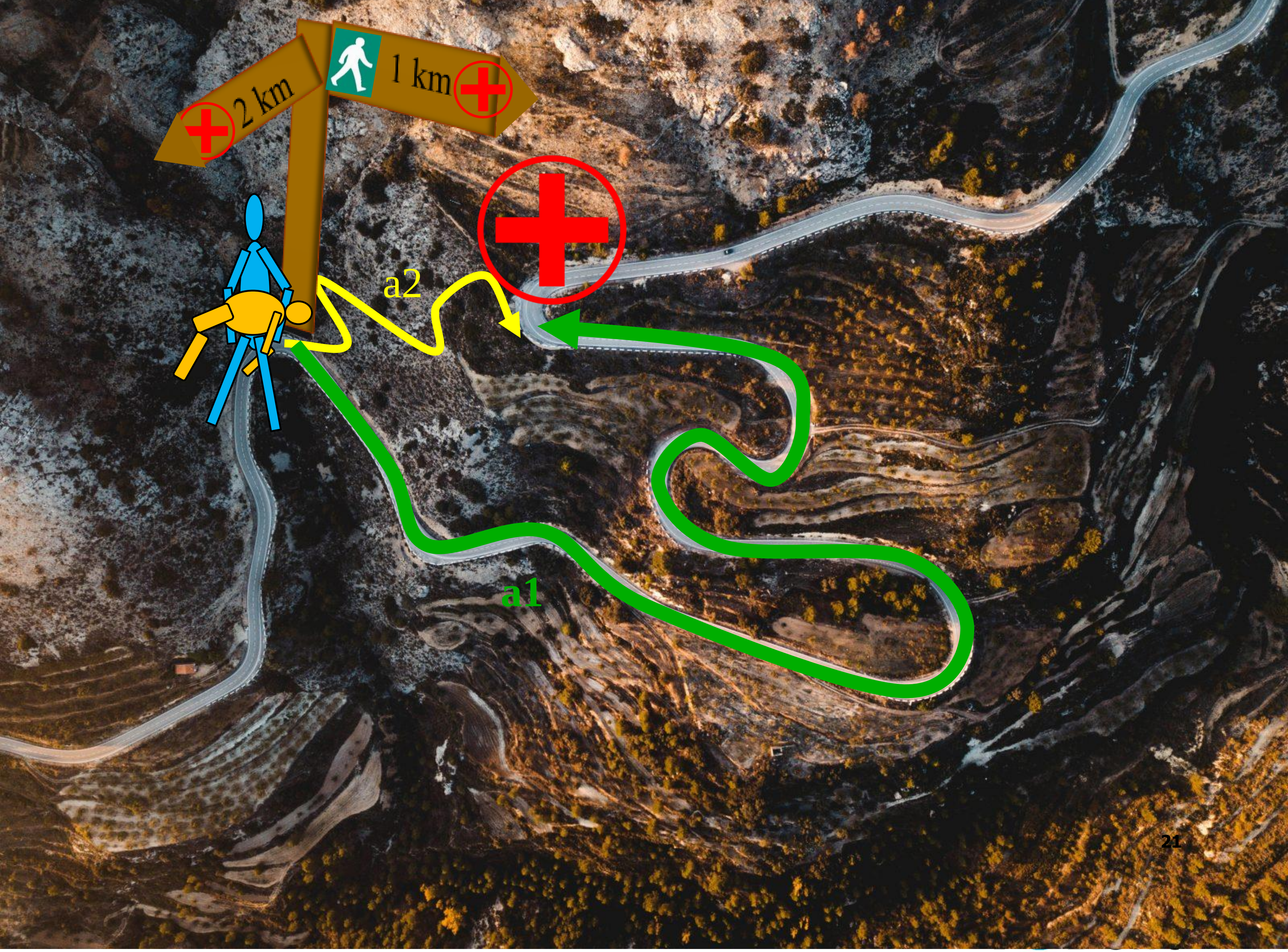


Algorithm:

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

Value of information

- Can we make better decisions if we gather more information?
  - Takes time, money, sometimes other consequences
  - Situation might have changed
  - Information might not be worth the cost of acquisition
- Additional information is only valuable if it is likely to alter your decision
  - How do you decide if gathering information is worthwhile?
  - What would you pay for information?


2 km
1 km
a2
a1

Value of information

- Idea: compute value of acquiring each possible piece of evidence
- Example:

There are two different routes through a mountain range: a1, a2.

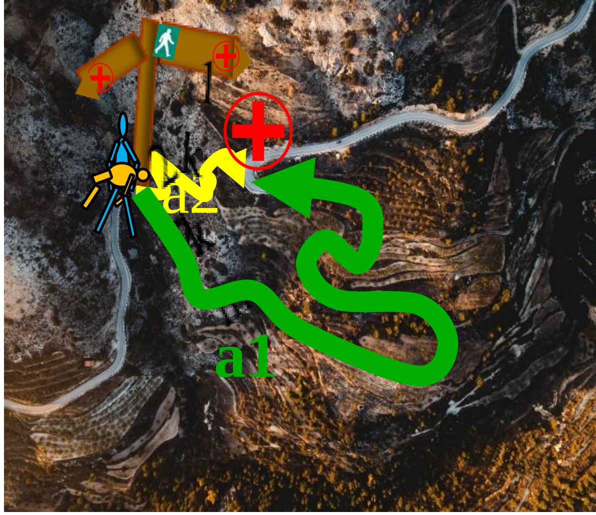
a1 is a long, straight way, and a2 is a short winding road.

It’s winter and snow or ice blockages can be anywhere.

It is possible to obtain satellite reports on the actual state of each road, that would give new expectations.

You carry an injured person. Do you pay for the report?

Actions, outcomes and observations



Actions:

a1: 2 km smooth walking, low chance of additional accidents, medium chance that the patient dies on route.

a2: 1 km rough walking, medium chance of additional accidents, low chance that the patient dies on route if the road is not blocked, chance that the route is blocked.

Possible outcomes:

- Ssafe: patient delivered safely to the red cross point
- Sbad: patient suffered more trauma but was delivered alive
- Sdied: patient died on route

Possible additional observations:

App for a satellite image: it takes time to wait for the image, which increases the chance of the patient dying. Possible results:

- no rockfall (yet)
- Road blocked by rockfall

Computing the Value of Information

For the Mountain road example:
Do you ask and wait for the satellite report?

To answer that question: compute expected value of information

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

Let’s call the currently available information E.
E_j: the satellite report
E_j can return: e_{j1} = no rockfall, e_{j2} = rockfall
Determine the chances of outcomes given the actions and the available information

The Mountain Example

Formally	Explanation
E	Currently available info (a1: 2 km & smooth, a2: 1 km & rough)
E_j	the satellite report
e_{j1}	Observation result: no rockfall
e_{j2}	Observation result: rockfall
a1	Take route a1
a2	Take route a2
Ssafe	patient safe
Sbad	patient suffered more trauma
Sdied	patient died

Some of the chance information, you need more:

$P(\text{Ssafe} \mid E, a1) > P(\text{Ssafe} \mid E, a2)$

$P(\text{Ssafe} \mid E, e_{j1}, a1) < P(\text{Ssafe} \mid E, e_{j1}, a2)$

$P(\text{Ssafe} \mid E, e_{j2}, a1) > P(\text{Ssafe} \mid E, e_{j2}, a2)$

Should we make the additional observation?

To answer that question: compute expected value of information

= expected value of best action given the additional information

minus expected value of best action without the additional information

VPI: value of perfect information is thus defined as

$VPI_E(E_j) = EU(\alpha_{EEj} \mid E, E_j) - EU(\alpha_E \mid E)$, where

α_E : the best action given evidence E

α_{EEj} : the best action given evidence E and Ej

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

Note: the information you can get is typically not perfect

General formula for perfect info

Current evidence E, current best action α

Possible action outcomes S_i, potential new evidence E_j

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

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

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

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

$\Rightarrow$ must compute expected gain over all possible values:

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

(VPI = value of perfect information)

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 ⇒ evidence-gathering becomes a sequential decision problem

Qualitative behavior

Actions: a, b

Utilities: U_a, U_b

$P(U \mid E)$ ——

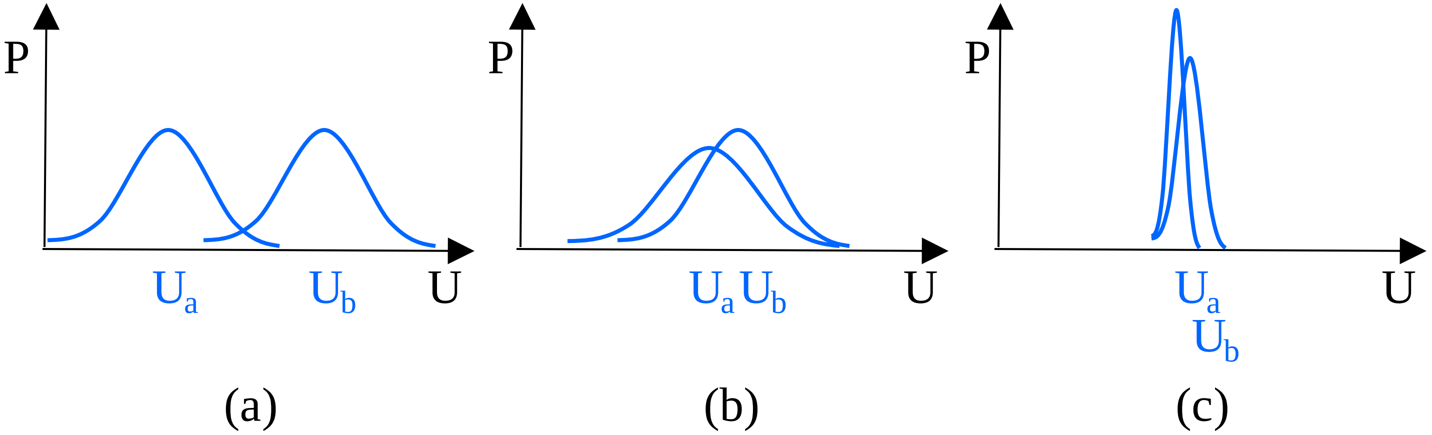
$P(U \mid E, E_j)$ - - - -

(a) Choice is obvious, information worth little

(b) Choice is nonobvious, information worth a lot

(c) Choice is nonobvious, information worth little

Before making observation E_j



Qualitative behavior

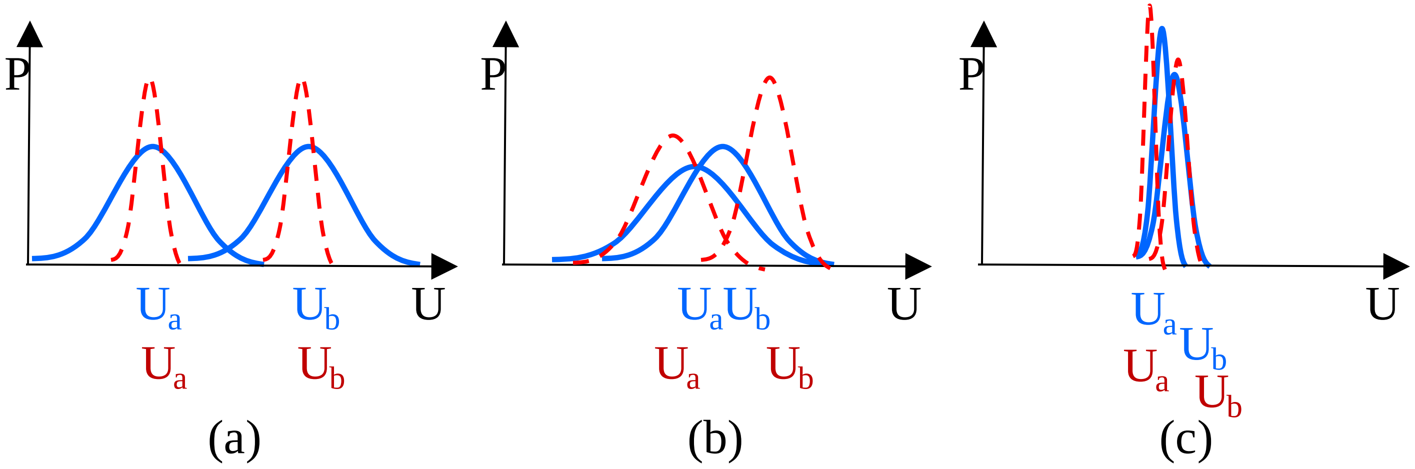
Actions: a, b

Utilities: U_a, U_b

P(U | E) ——

P(U | E, E_j) - - -

What might happen after making observation E_j



(a) (b) (c)

Qualitative behavior

Actions: a, b

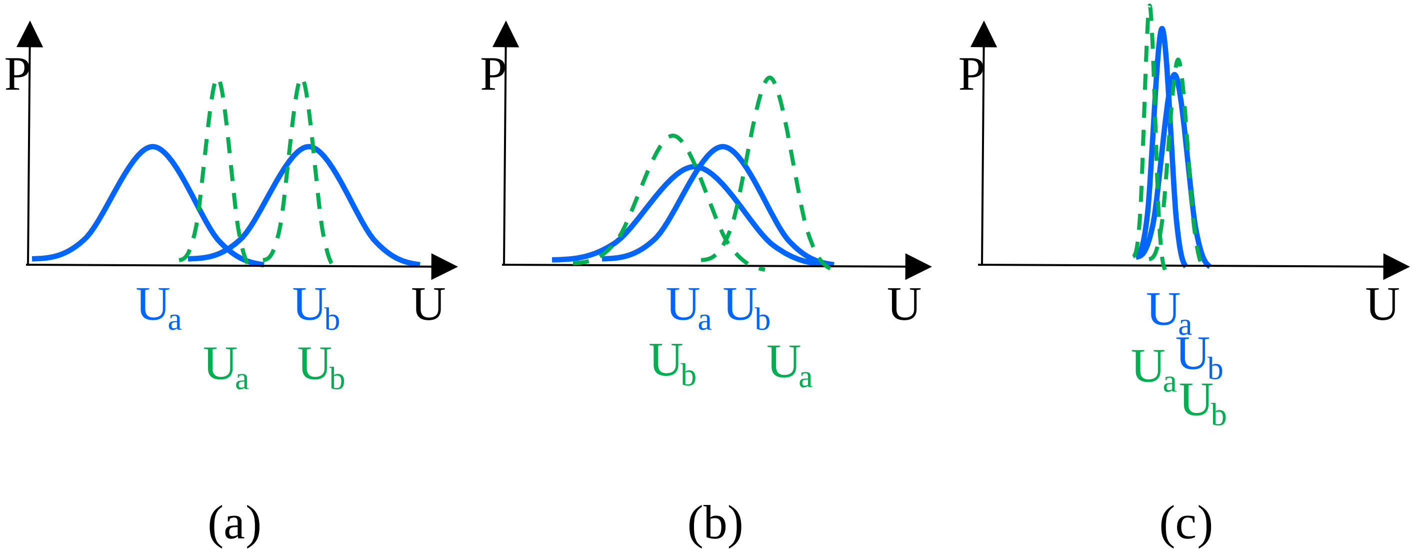
Utilities: U_a, U_b

P(U | E)

P(U | E, E_j)

What might ALSO happen after making observation E_j

Note that U_b and U_a swapped position in (b)



Looking ahead

- Sequential decision making
  - (Partially observable) Markov Decision processes
- Reinforcement Learning