Decision Making
in Robots and Autonomous Agents

Dynamic Programming Principles and Decision Theory

Subramanian Ramamoorthy
School of Informatics

8 February, 2019

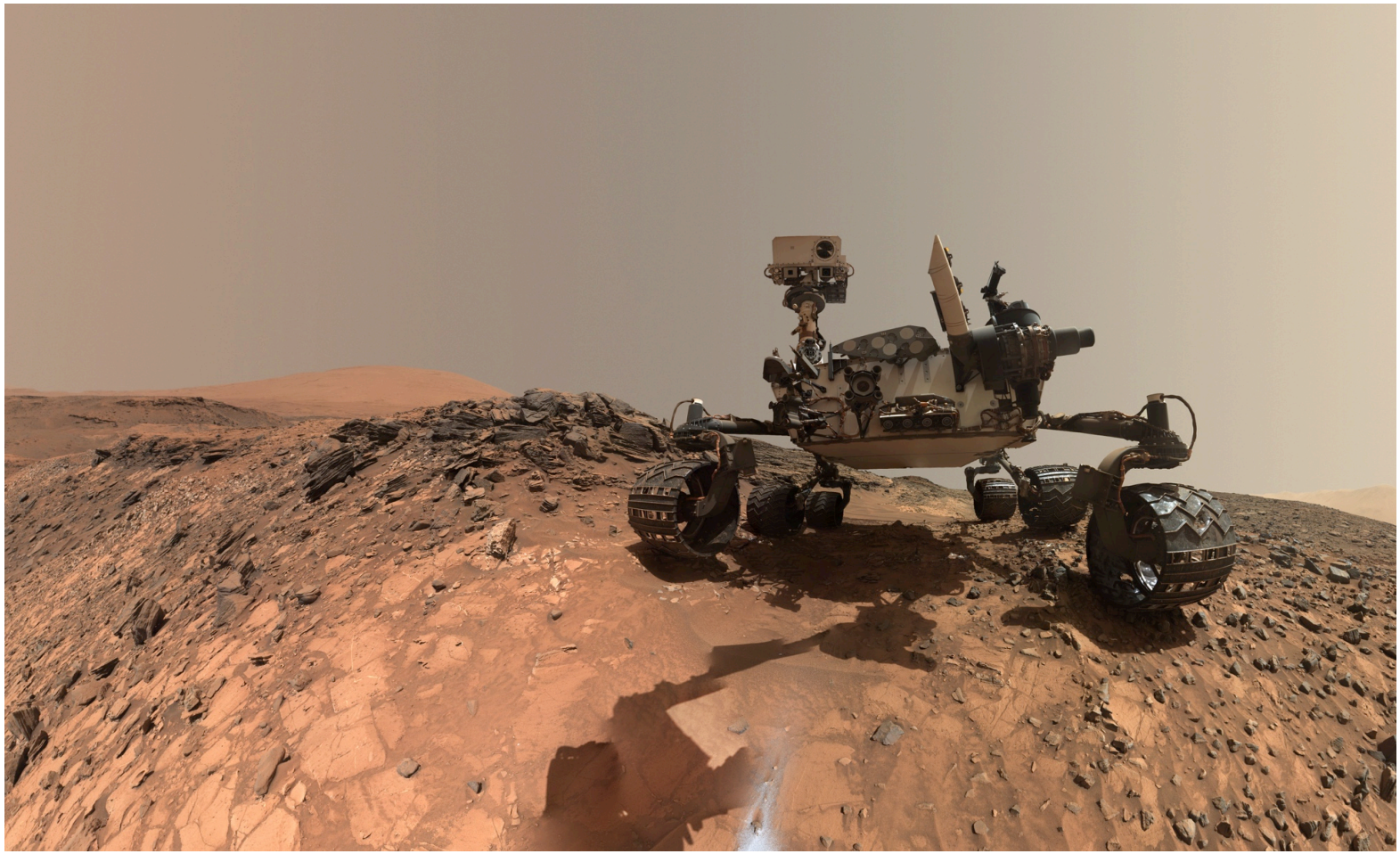
Objectives of this Lecture

- Introduce the dynamic programming principle, a way to solve sequential decision problems (such as path planning)

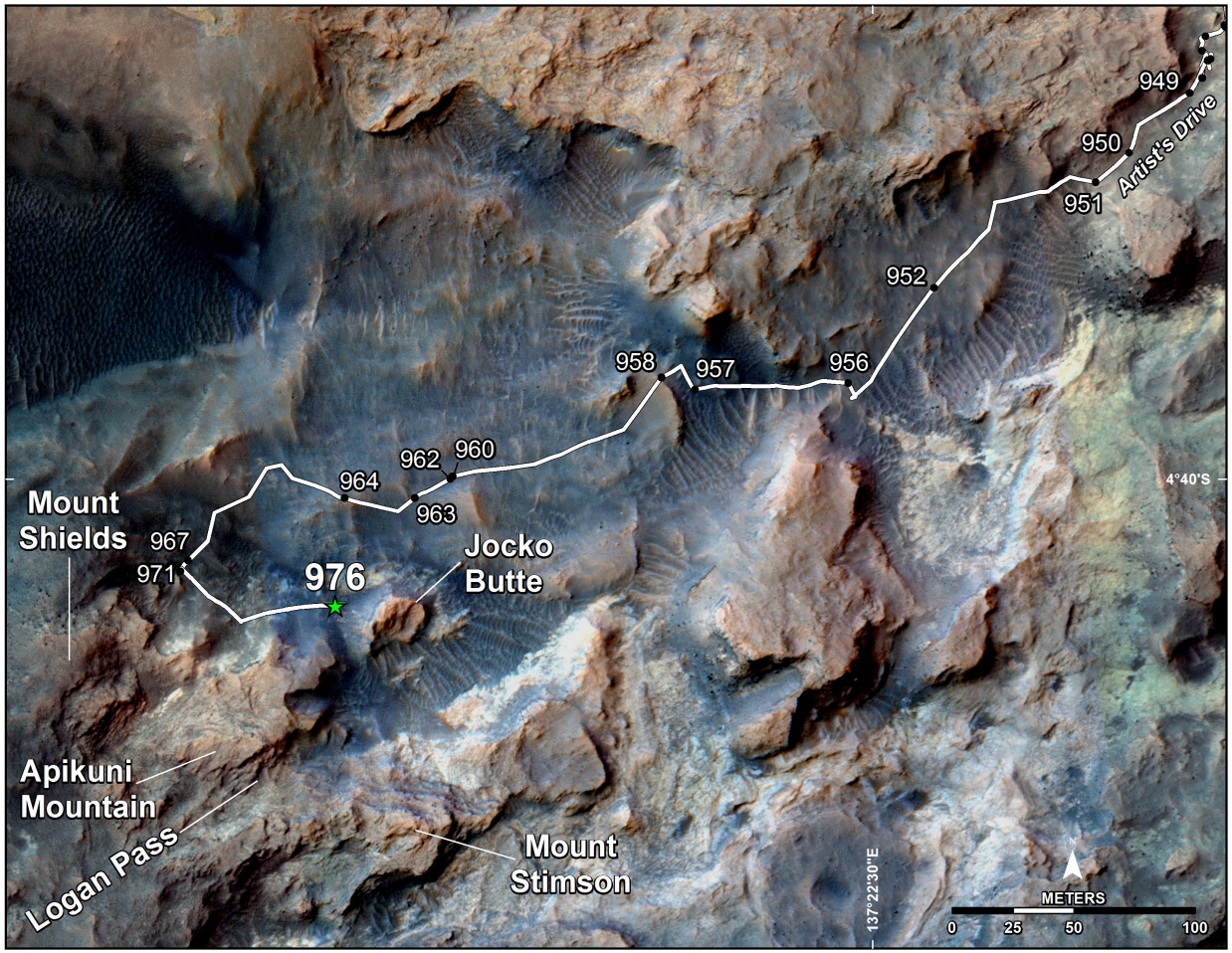
- Introduce the Markov Decision Process model, and discuss the nature of the policy arising in a similar sequential decision problem with probabilistic transitions
  - Includes recap of the notion of Markov Chains

- In the second half, introduce different ways of posing decision problems in terms of utilities, motivating principles of Bayesian choices

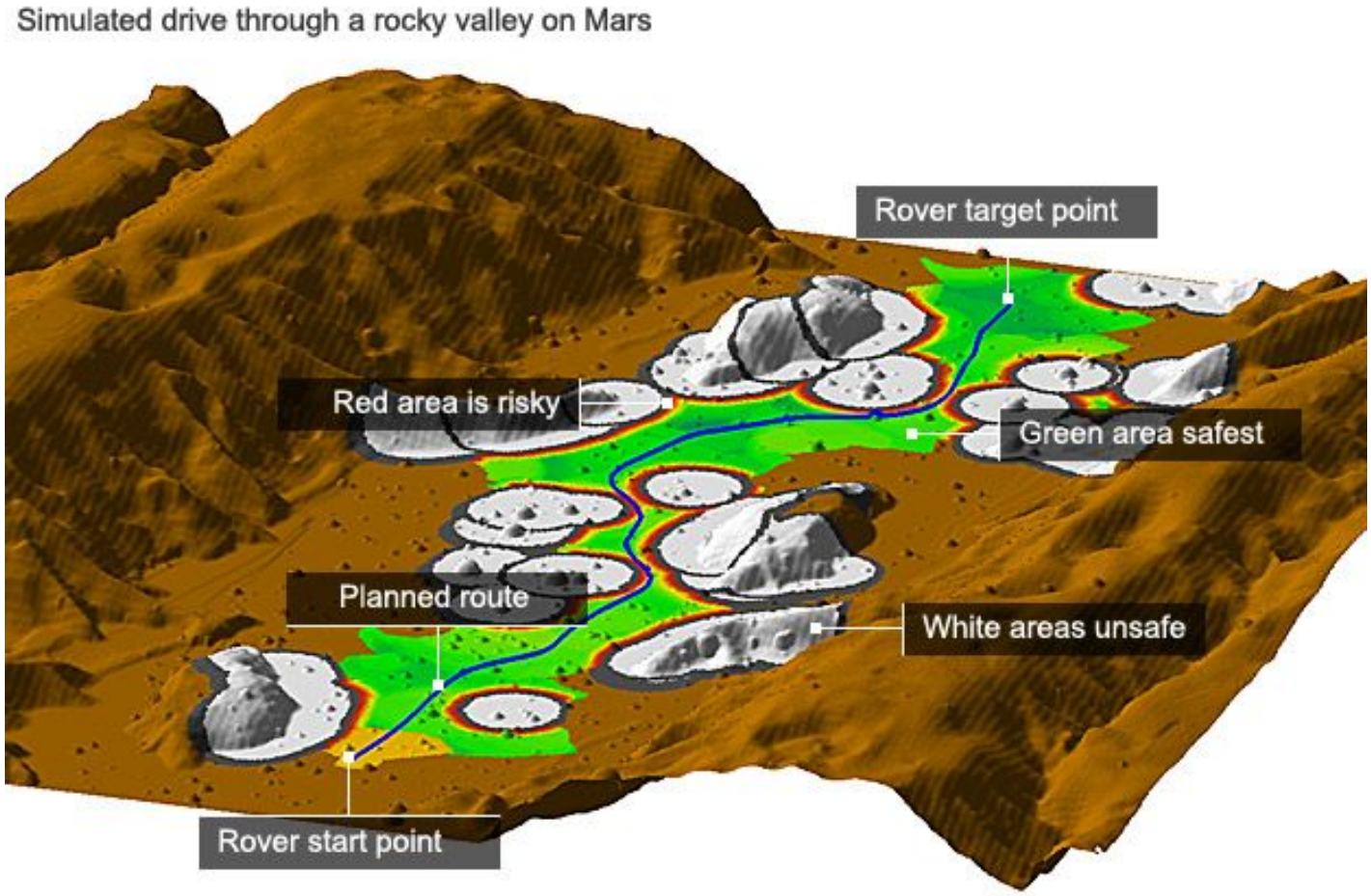
Problem of Determining Paths



Getting from “A to B”: Bird’s Eye View



Getting from "A to B": Local View



How could we calculate the best path?

Dynamic Programming (DP) Principle

- Mathematical technique often useful for making a sequence of inter-related decisions
- Systematic procedure for determining the combination of decisions that maximize overall effectiveness
- There may not be a "standard form" of DP problems, instead it is an approach to problem solving and algorithm design

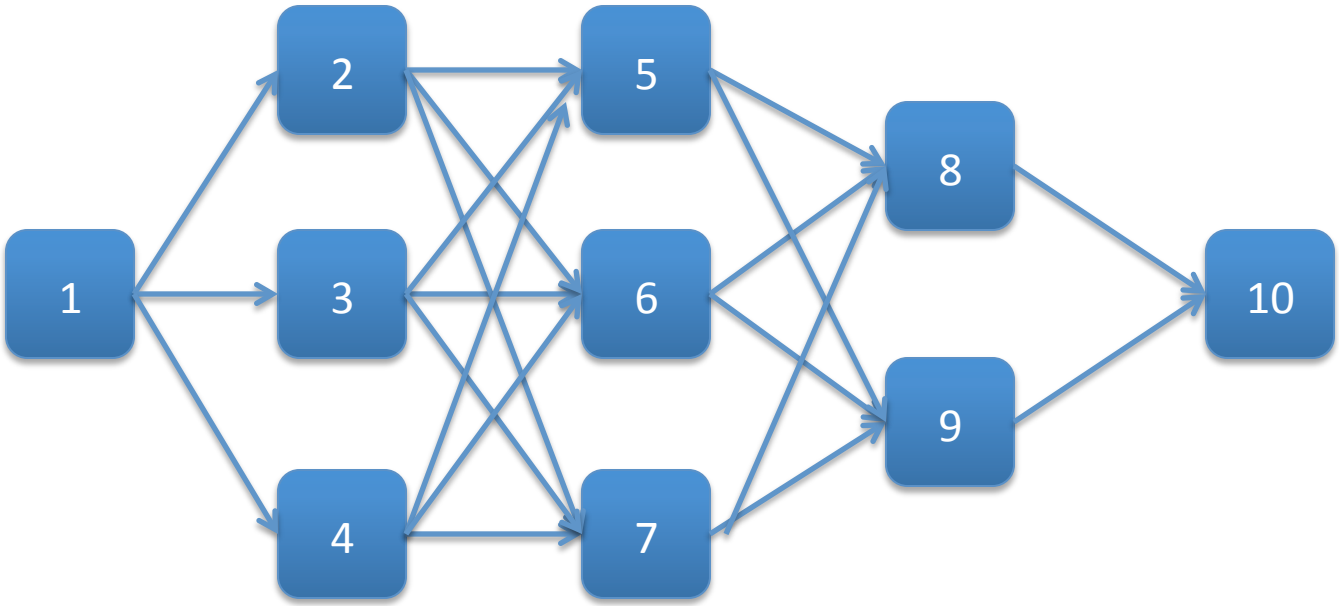
- We will try to understand this through a few example models, solving for the "optimal policy" (the notion of which will become clearer as we go along)

Stagecoach Problem

- Simple thought experiment due to H.M. Wagner at Stanford
- Consider a mythical American salesman from over a hundred years ago. He needs to travel west from the east coast, through unfriendly country with bandits.
- He has a well defined start point and destination, but the states he visits en route are up to his own choice

- Let us visualize this, using numbered blocks for states

Stagecoach Problem: Possible Routes



Each box is a state (generically indexed by an integer, i)
Transitions, i.e., edges, can be annotated with a "cost"

Stagecoach Problem: Setup

- The salesman needs to go through four stages to travel from his point of departure in state 1 to destination in state 10
- This salesman is concerned about his safety – does not want to be attacked by bandits
- One approach he could take (as envisioned by Wagner):
  - Life insurance policies are offered to travellers
  - Cost of each policy is based on evaluation of safety of path
  - Safest path = cheapest life insurance policy

Stagecoach Problem: Costs

The cost of the standard policy on the stagecoach run from state i to state j denoted by c_{ij} is

	2	3	4
1	2	4	3

	5	6	7
2	7	4	6
3	3	2	4
4	4	1	5

	8	9
5	1	4
6	6	3
7	3	3

	10
8	3
9	4

Which route minimizes the total cost of the policy?

Myopic Approach

- Making the decision which is best for each successive stage need not yield the overall optimal decision
- WHY?

- Selecting the cheapest run offered by each successive stage would give the route 1 -> 2 -> 6 -> 9 -> 10.
- What is the total cost?

- **Observation**: Sacrificing a little on one stage may permit greater savings thereafter.
  - e.g., a cheaper alternative to 1 -> 2 -> 6 is 1 -> 4 -> 6

Is Trial and Error Useful?

- What does it mean to solve the problem (finding the cheapest cost path) by trial and error?
  - What are the trials over? What is the error?

- How many possible routes do we have in this problem?

  Ans: 18

- Is exhaustive enumeration always an option? How does the number of routes scale?

Dynamic Programming Principle

- Start with a small portion of the problem and find optimal solution for this smaller problem
- Gradually enlarge the problem – finding the current optimal solution from the previous one

  … until original problem is solved in its entirety

- This general philosophy is the essence of the DP principle
  - The details are implemented in many different ways in different specialised scenarios

Solving the Stagecoach Problem

- At stage n, consider the decision variable x_n ($n = 1,2,3,4$).
- The selected route is: $1 \rightarrow x_1 \rightarrow x_2 \rightarrow x_3 \rightarrow x_4$

Which state is implied by x_4?

- Total cost of the overall best *policy* for the *remaining* stages, given that the salesman is in state s and selects x_n as the immediate destination: $f_n(s, x_n)$

$$x_n^* = \arg\min f_n(s, x_n)$$

$$f_n^*(s) = \text{minimum value of } f_n(s, x_n)$$

$$f_n^*(s) = f_n(s, x_n^*)$$

Solving the Stagecoach Problem

- The objective is to determine $f_1^*(1)$
  and the corresponding optimal policy achieving this
- DP achieves this by successively finding $f_4^*(s), f_3^*(s), f_2^*(s)$
  which will lead us to the desired $f_1^*(1)$

- When the salesman has only one more stage to go, his route is entirely determined by his final destination. Therefore,

s	$f_4^*(s)$	x_4^*
8	3	10
9	4	10

Solving the Stagecoach Problem

- What about when the salesman has two more stages to go?

- Assume salesman is at stage 5 – he must next go either to stage 8 or 9 at cost of 1 or 4 respectively
  - If he chooses stage 8, minimum additional cost after reaching there is 3 (table in earlier slide)
  - So, cost for that decision is 1 + 3 = 4
  - Total cost if he chooses stage 9 is 4 + 4 = 8

- Therefore, he should choose state 8

The Two-stage Problem

$$f_3(s, x_3) = c_{sx_3} + f_4^*(x_3)$$

$s \backslash x_3$	8	9	$f_3^*(s)$	x_3^*
5	4	8	4	8
6	9	7	7	9
7	6	7	6	8

Likewise, Three-stage Problem

$$f_2(s, x_2) = c_{sx_2} + f_3^*(x_2)$$

$s \backslash x_2$	5	6	7	$f_2^*(s)$	x_2^*
2	11	11	12	11	5 or 6
3	7	9	10	7	5
4	8	8	11	8	5 or 6

Finally, the Four-stage Problem

$$f_1(s, x_1) = c_{sx_1} + f_2^*(x_1)$$

$s \backslash x_1$	2	3	4	$f_1^*(s)$	x_1^*
1	13	11	11	11	3 or 4

Optimal Solution:
Salesman should first go to either 3 or 4
Say, he chooses 3, the three-stage problem result is 5
Which leads to the two-stage problem result of 8
And, of course, finally 10

Characteristics of DP Problems

The stagecoach problem might have sounded strange, but it is the literal instantiation of key DP terms

DP problems all share certain features:

1. The problem can be divided into **stages**, with a **policy decision** required at each stage
2. Each stage has several **states** associated with it
3. The effect of the policy decision at each stage is to transform the current state into a state associated with the next stage (could be according to a probability distribution, as we'll see next).

Characteristics of DP Problems, contd.

5. Given the current state, an optimal policy for the remaining stages is **independent** of the policy adopted in previous stages
6. The solution procedure begins by finding the optimal policy for each state of the last stage.
7. Recursive relationship identifies optimal policy for each state at stage n, given optimal policy for each state at stage $n+1$:

$$f_n^*(s) = \min_{x_n}\{c_{sx_n} + f_{n+1}^*(x_n)\}$$

8. Using this recursive relationship, the solution procedure moves backward stage by stage – until finding optimal policy from initial stage

Let us now consider a problem where the transitions may not be deterministic:

(A little bit about) Markov Chains and Decisions

Stochastic Processes

- A *stochastic process* is an indexed collection of random variables $\{X_t\}$
  - e.g., collection of weekly demands for a product
- One type: At a particular time t, labelled by integers, system is found in exactly one of a finite number of mutually exclusive and exhaustive categories or **states**, labelled by integers too
- Process could be *embedded* in that time points correspond to occurrence of specific events (or time may be equi-spaced)
- Random variables may depend on others, e.g.,

$$X_{t+1} = \begin{cases} \max\{(3 - D_{t+1}), 0\}, if X_t < 0 \\ \max\{(X_t - D_{t+1}), 0\}, if X_t \geq 0 \end{cases}$$

Markov Chains

- The stochastic process is said to have a **Markovian** property if

$$P\{X_{t+1} = j | X_0 = k_0, X_1 = k_1, ..., X_{t-1} = k_{t-1}, X_t = i\} = P\{X_{t+1} = j | X_t = i\}$$

$$\text{for} \quad t \quad = \quad 0, 1, ... \quad \text{and} \quad \text{every} \quad \text{sequence} \quad i, j, k_0, ..., k_{t-1}.$$

- Markovian property means that the conditional probability of a future event given any past events and current state, i*s independent* of past states and depends only on present
- The conditional probabilities are **transition probabilities**,

$$P\{X_{t+1} = j | X_t = i\}$$

- These are stationary if time invariant, called p_{ij},

$$P\{X_{t+1} = j | X_t = i\} = P\{X_1 = j | X_0 = i\}, \forall t = 0, 1, ...$$

Markov Chains

- Looking forward in time, n-step **transition probabilities**, $p_{ij}^{(n)}$

$$P\{X_{t+n} = j | X_t = i\} = P\{X_n = j | X_0 = i\}, \forall t = 0, 1, ...$$

- One can write a transition matrix,

$$\mathbf{P}^{(n)} = \begin{bmatrix} p_{00}^{(n)} & \cdots & p_{0M}^{(n)} \\ \vdots & & \\ p_{M0}^{(n)} & \cdots & p_{MM}^{(n)} \end{bmatrix}$$

- A stochastic process is a finite-state Markov chain if it has,
  - Finite number of states
  - Markovian property
  - Stationary transition probabilities
  - A set of initial probabilities $P\{X_0 = i\}$ for all i

Markov Chains

- n-step transition probabilities can be obtained from 1-step transition probabilities recursively (Chapman-Kolmogorov)

$$p_{ij}^{(n)} = \sum_{k=0}^{M} p_{ik}^{(v)} p_{kj}^{(n-v)}, \forall i, j, n; 0 \leq v \leq n$$

- We can get this via the matrix too

$$P^{(n)} = P.P \ldots P = P^n = PP^{n-1} = P^{n-1}P$$

- **First Passage Time**: number of transitions to go from i to j for the first time
  - If $i = j$, this is the **recurrence time**
  - In general, this itself is a random variable

Markov Chains

- n-step recursive relationship for first passage time

$$
\begin{aligned}
f_{ij}^{(1)} &= p_{ij}^{(1)} = p_{ij}, \\
f_{ij}^{(2)} &= p_{ij}^{(2)} - f_{ij}^{(1)} p_{jj}, \\
&\vdots \\
f_{ij}^{(n)} &= p_{ij}^{(n)} - f_{ij}^{(1)} p_{jj}^{(n-1)} - f_{ij}^{(2)} p_{jj}^{(n-2)} \ldots - f_{ij}^{(n-1)} p_{jj}
\end{aligned}
$$

- For fixed i and j, these $f_{ij}^{(n)}$ are nonnegative numbers so that

$$\sum_{n=1}^{\infty} f_{ij}^{(n)} \leq 1$$

What does <1 signify?

- If, $\sum_{n=1}^{\infty} f_{ii}^{(n)} = 1$, state is **recurrent**; If n=1 then it is **absorbing**

Markov Chains: Long-Run Properties

- Consider this transition matrix of an inventory process:

$$P^{(1)} = P = \begin{bmatrix} 0.08 & 0.184 & 0.368 & 0.368 \\ 0.632 & 0.368 & 0 & 0 \\ 0.264 & 0.368 & 0.368 & 0 \\ 0.08 & 0.184 & 0.368 & 0.368 \end{bmatrix}$$

- This captures the evolution of inventory levels in a store
  - What do the 0 values mean?
  - Other properties of this matrix?

Markov Chains: Long-Run Properties

The corresponding 8-step transition matrix becomes:

$$P^{(8)} = P^8 = \begin{bmatrix} 0.286 & 0.285 & 0.264 & 0.166 \\ 0.286 & 0.285 & 0.264 & 0.166 \\ 0.286 & 0.285 & 0.264 & 0.166 \\ 0.286 & 0.285 & 0.264 & 0.166 \end{bmatrix}$$

Interesting property: probability of being in state j after 8 weeks appears independent of *initial* level of inventory.

- For an irreducible ergodic Markov chain, one has limiting probability

$$\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j$$

Reciprocal gives you recurrence time

$$\pi_j = \sum_{i=0}^{M} \pi_i p_{ij}, \forall j = 0, ..., M$$

Markov **Decision** Model

- Consider the following application: machine maintenance
- A factory has a machine that deteriorates rapidly in quality and output and is inspected periodically, e.g., daily
- Inspection declares the machine to be in four possible states:
  - 0: Good as new
  - 1: Operable, minor deterioration
  - 2: Operable, major deterioration
  - 3: Inoperable
- Let X_t denote this observed state
  - evolves according to some "law of motion", it is a stochastic *process*
  - Furthermore, assume it is a finite state Markov chain

Markov **Decision** Model

- Transition matrix is based on the following:

States	0	1	2	3
0	0	7/8	1/16	1/16
1	0	3/4	1/8	1/8
2	0	0	1/2	1/2
3	0	0	0	1

- Once the machine goes inoperable, it stays there until repairs
  - If no repairs, eventually, it reaches this state which is absorbing!

- Repair is an **action** – a very simple maintenance **policy**.
  - e.g., machine from from state 3 to state 0

Markov **Decision** Model

- There are costs as system evolves:
  - State 0: cost 0
  - State 1: cost 1000
  - State 2: cost 3000
- Replacement cost, taking state 3 to 0, is 4000 (and lost production of 2000), so cost = 6000
- The modified transition probabilities are:

States	0	1	2	3
0	0	7/8	1/16	1/16
1	0	3/4	1/8	1/8
2	0	0	1/2	1/2
3	1	0	0	0

Markov **Decision** Model

- Simple question (a behavioural property):
  What is the average cost of this maintenance policy?

- Compute the steady state probabilities:

$$\pi_0 = \frac{2}{13}; \pi_1 = \frac{7}{13}; \pi_2 = \frac{2}{13}; \pi_3 = \frac{2}{13}$$

How?

- (Long run) expected average cost per day,

$$0\pi_0 + 1000\pi_1 + 3000\pi_2 + 6000\pi_3 = \frac{25000}{13} = 1923.08$$

Markov **Decision** Model

- Consider a slightly more elaborate policy:
  - When it is inoperable or needing major repairs, replace
- Transition matrix now changes a little bit
- Permit one more possible action: overhaul
  - Go back to minor repairs state (1) for the next time step
  - Not possible if truly inoperable, but can go from major to minor
- Key point about the system behaviour. It evolves according to
  - "Laws of motion"
  - Sequence of decisions made (actions from {1: none,2:overhaul,3: replace})
- Stochastic process is now defined in terms of $\{X_t\}$ and $\{\Delta_t\}$
  - Policy, R, is a rule for making decisions
    - Could use all history, although popular choice is (current) state-based

Markov **Decision** Model

- There is a space of potential policies, e.g.,

Policies	$d_0(R)$	$d_1(R)$	$d_2(R)$	$d_3(R)$
R_a	1	1	1	3
R_b	1	1	2	3
R_c	1	1	3	3
R_d	1	3	3	3

- Each policy defines a transition matrix, e.g., for R_b

States	0	1	2	3
0	0	7/8	1/16	1/16
1	0	3/4	1/8	1/8
2	0	1	0	0
3	1	0	0	0

Which policy is best?
Need costs….

Markov **Decision** Model

- C_{ik} = expected cost incurred during next transition if system is in state i and decision k is made

State	Dec.	1	2	3
0		0	4	6
1		1	4	6
2		3	4	6
3		∞	∞	6

- The long run average expected cost for each policy may be computed using

$$E(C) = \sum_{i=0}^{M} C_{ik}\pi_i$$

R_b is best

So, What is a Policy?

- A "program"
- Map from states (or situations in the decision problem) to actions that could be taken
  - e.g., if in 'level 2' state, call contractor for overhaul
  - If less than 3 DVDs of a film, place an order for 2 more

- A probability distribution $\pi(s,a)$
  - A joint probability distribution over states and actions
  - If in a state s_1, then with probability defined by π, take action a_1

Utility and Decision Theory:
How should a robot incorporate notions of choice?

Types of Decisions

- Who makes it?
  - Individual
  - ‘Group’
- What are the conditions?
  - Certainty
  - Risk
  - Uncertainty

How to Model Decision under *Certainty*?

- Given a set of possible acts
- Choose one that maximizes some given index

If **a** is a generic act in a set of feasible acts **A**, f(**a**) is an index being maximized, then

Problem: Find **a*** in **A** such that f(**a***) > f(**a**) for all **a** in **A**.

The index f plays a key role, e.g., think of buying a painting.

Essential problem: How should the subject select an index function such that her choice reduces to finding maximizers?

Operational Way to Find *an* Index Function

- Observe subject's behaviour in restricted settings and predict purchase behaviour from that:
- Instruct the subject as follows:
  - Here are ten valuable reproductions
  - We will present these to you in pairs
  - You will tell us which one of the pair you prefer to own
  - After you have evaluated all pairs, we will pick a pair at random and present you with the choice you previously made (it is to your advantage to remember your true tastes)
- The subject's behaviour is **as though** there is a ranking over all paintings, so each painting can be summarized by a number

Some Properties of this Ranking

- *Transitivity*: Previous argument only makes sense if the rank is transitive – if A is preferred in (A, B) and B is preferred in (B, C) then A is preferred in (A, C); and this holds for all triples of alternatives A, B and C
- *Ordinal nature of index*: One is tempted to turn the ranking into a latent measure of 'satisfaction' but that is a mistake as utilities are non-unique.

  e.g., we could assign 3 utiles to A, 2 utiles to B and 1 utile to C to explain the choice behaviour

  Equally, 30, 20.24 and 3.14 would yield the same choice

While it is OK to compare indices, it is not OK to add or multiply

What Happens if we Relax Transitivity?

- Assume Pandora says (in the pairwise comparisons):
  - Apple < Orange
  - Orange < Fig
  - Fig < Apple
- Is this a problem for Pandora? Why?

- Assume a merchant who transacts with her as follows:
  - Pandora has an Apple at the start of the conversation
  - He offers to exchange Orange for Apple, if she gives him a penny
  - He then offers an exchange of Fig for Orange, at the price of a penny
  - Then, offers Apple for the Fig, for a penny
  - **Now, what is Pandora's net position?**

Decision Making under *Risk*

- Initially appeared as analysis of fair gambles, needed some notions of utility
- Gamble has n outcomes, each worth $a_1, \ldots, a_n$
- The probability of each outcome is $p_1, \ldots, p_n$
- How much is it worth to participate in this gamble?

$$b = a_1 p_1 + \ldots + a_n p_n$$

One may treat this monetary expected value as a fair price

Is this a sufficient description of choice behaviour under risk?

St. Petersburg Paradox of D. Bernoulli

- A fair coin is tossed until a head appears
- Gambler receives 2^n if the first head appears on trial n
- Probability of this event = probability of tail in first (n-1) trials and head on trial n, i.e., $(1/2)^n$

Expected value $= 2.(1/2) + 4.(1/2)^2 + 8.(1/2)^8 + \ldots = \infty$

- Are you willing to bet in this way? Is anyone?

Defining Utility

- Bernoulli went on to argue that people do not act in this way
- The thing to average is the 'intrinsic worth' of the monetary values, not the absolute values

  e.g., intrinsic worth of money may increase with money but at a *diminishing rate*

- Let us say utility of m is $\log_{10} m$, then expected value is,

  $$\log_{10} 2.(1/2) + \log_{10} 4.(1/2)^2 + \log_{10} 8.(1/2)^8 + \ldots = b < \infty$$

  Monetary fair price of the gamble is a where $\log_{10} a = b$.

Some Critiques of Bernoulli's Formulation

von Neumann and Morgenstern (vNM), who 'started' game theory, raised the following questions:

- The assignment of utility to money is arbitrary and *ad hoc*
  - There are an infinity of functions that capture 'diminishing rate', how should we choose?
  - The association may vary from person to person

- Why is the definition of the decision based upon expected value of this notion of utility?
  - Is this actually descriptive of a single gambler, in "one-shot" choice?

von Neumann & Morgenstern Formulation

- If a person is able to express preferences between every possible pair of gambles

  where gambles are taken over some basic set of alternatives

- Then one *can* introduce utility associations to the basic alternatives in such a manner that

- If the person is guided solely by the utility expected value, *he is acting in accord with his true tastes*.
  - provided his tastes are consistent in some way

Constructing Utility Functions

- Suppose we know the following preference order:
  - A < b ~ c < d < e

- The following are utility functions that capture this:

	a	b	c	d	E
U	0	1/2	1/2	3/4	1
V	-1	1	1	2	3
W	-8	0	0	1	8

  - So, in situations like St Petersburg paradox, the revealed preference of any realistic player may differ from the case of infinite expected value
  - Satisfaction at some large value, risk tolerance, time preference, etc.

Certainty Equivalents and Indifference

- The previous statement applies equally well to certain events and gambles or lotteries
- So, even attitudes regarding tradeoffs between the two ought to be captured
- Basic issue – how to compare?
- Imagine the following choice (A > B > C pref.) : (a) you get B for certain, (b) you get A with probability p and C otherwise
- If p is near 1, option b is better; if p is near 0, then option a: there is a single point where we switch
- Indifference is described as something like

$$(2/3)\,(1) + (1 - 2/3)\,(0) = 2/3$$

A C B

Caveats

- As before, we need to remember that the utility values should not be mis-interpreted
- The number 2/3 is determines by choices among risky alternatives and reflect attitude to ‘gambling’
- For instance, imagine a subject who would be indifferent to paying $9 and a 50-50 chance of paying $10 or nothing;
- This suggests utilities for $0, -$9, -$10 are 1, ½, 0.
- However, we can’t say it is just as enjoyable for him to go from -$10 to -$9 as it is to go from -$9 to $0!
- Subject’s preferences among alternatives or lotteries come prior to numerical characterization of them

Axiomatic Treatment of Utility

vNM and others formalize the above to define axioms for utility:

1) Any two alternatives shall be comparable, i.e., given any two, subject will prefer one over the other of be indifferent
2) Both preference and indifference relations for lotteries are transitive
3) In case a lottery has as one of its alternatives another lottery, then the first lottery is decomposable into the more basic alternatives through the use of the probability calculus
4) If two lotteries are indifferent to the subject then they are interchangeable as alternatives in any compound lottery

Axiomatic Treatment of Utility, contd.

vNM and others formalize the above to define axioms for utility:

5) If two lotteries involve the same two alternatives, then the one in which the more preferred alternative has a higher probability of occurring is itself preferred
6) If A is preferred to B and B to C, then there exists a lottery involving A and C (with appropriate probabilities) which is indifferent to B

Decision Making under *Uncertainty*

- A choice must be made from among a set of acts, $A_1, \ldots, A_m$.
- The relative desirability of these acts depends on which state of nature prevails, either $s_1, \ldots, s_n$.
- As decision maker we know that one of several things is true and this influences our choice but we **do not** have a probabilistic characterization of these alternatives

- Savage's omelet problem: Your friend has broken 5 good eggs into a bowl when you come in to volunteer and finish the omelet. A sixth egg lies unbroken (you must use it or waste it altogether). Your three acts: break it into bowl, break it into saucer – inspect and pour into bowl, throw it uninspected

Decision Making under *Uncertainty*

Table 1. Savage's example illustrating acts, states, and consequences

Act	State	
	Good	Rotten
Break into bowl	six-egg omelet	no omelet, and five good eggs destroyed
Break into saucer	six-egg omelet, and a saucer to wash	five-egg omelet, and a saucer to wash
Throw away	five-egg omelet, and one good egg destroyed	five-egg omelet

- To each outcome, we could assign a utility and maximize it
- What do we know about the state of nature?
  - We may act *as though* there is one true state and we just don't know it
  - If we assume a probability over s, this is decision under risk

- What criteria do we have for a decision problem under ***uncertainty*** (d.p.u.u.)?

Some Criteria for d.p.u.u.

Maximin criterion: To each act, assign its security level as an index. Index of A_i is the minimum of the utilities $u_{i1}, \ldots, u_{in}$

Choose the act whose associated index is maximum.

	s1	s2
A1	0	1
A2	1	0

- What is the security level for each act?
- What happens if we allow for mixed strategies (i.e., akin to a compound lottery, e.g., p = 0.5 for a1 and p = 0.5 for a2) ?
- Interpretation as game against nature: best response against nature's minimax strategy (least favourable a priori strategy)

Point to Ponder about Maximin

- Is nature a *conscious* adversary?!
- Consider:

	s1	s2
A1	0	100
A2	1	1

- What are the safety values for the actions?
  - If mixed strategies are allowed?

- What if 100 went up to 10^6 and 1 came down to 0.0001?

Some Criteria for d.p.u.u.

- **Minimax risk criterion** (Savage): Consider a setup as follows:

	s1	s2
A1	0	100
A2	1	1

If s1 is the true state, choosing A2 poses no ‘risk’ whereas if s2 is the true state then considerable ‘risk’ in A2.

Savage’s procedure: (i) Create new risk payoffs which are amounts to be added to utility to match maximum column utility, (ii) Choose act which minimizes maximum risk index

Minimax Risk Criterion

- Transform Utility Payoff to Risk Payoff:

	s1	s2
A1	0	100
A2	1	1

	s1	s2
A1	1	0
A2	0	99

- Take the u_{ij} and define r_{ij} so that it is the amount that has to be added to u_{ij} to equal maximum utility payoff in column j.
- Critique (due to Chernoff):
  - "Regret" may not be measured by utility difference
  - Different states of nature may not be traded off properly
  - **Taking away an irrelevant (obviously bad) action may change optimal decision!**

More Criteria for d.p.u.u.

- **Pessimism-optimism index criterion** of Hurwicz:

  Let m_i and M_i be minimum and maximum utility. Assume a fixed pessimism-optimism index, α. To each act, associate an α-index $\alpha m_i + (1 - \alpha) M_i$.

  Of two acts, the one with higher α-index is preferred.

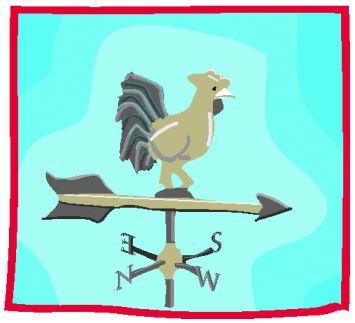
- **"Principle of insufficient reason"**: If one is completely ignorant, one should act as though all states are equally likely; so choice should be based on a utility index which is the average of utility for all possible states for any act

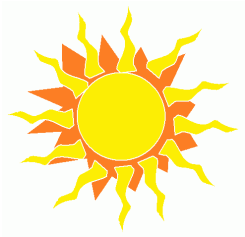
What is the effect of the way we enumerate possible states of nature?

Use of Bayesian Principles for Decisions: Simple Example

Bob observes the weather forecast before deciding whether to carry an umbrella to work. Bob wishes to stay dry, but carrying an umbrella around is annoying.

Forecast





Setup of Decision Theory

- Set ***A*** of actions
  - **Umbrella**=*{true, false}*
- Set ***E*** of (unobserved) events
  - **Weather**=*{rain, sun}*
- Set ***O*** of observations
  - **Forecast**=*{rain, sun}*
- Probability distribution over
  - events ***P(E)***
  - observations given events ***P(O | E)***
- Utility function from actions and events to real numbers.

	Weather
sun	0.7
rain	0.3

	Forecast	
Weather	*sun*	*rain*
sun	0.6	0.4
rain	0.4	0.6

Weather	Umbrella	Utility
sun	*TRUE*	-10
sun	*FALSE*	100
rain	*TRUE*	100
rain	*FALSE*	-10

Choosing the Best Action

Let $U^a(\text{Bob} \mid e)$ be Bob's reward for taking action $a \in \mathbf{A}$ after event $e \in \mathbf{E}$ has occurred.

The expected utility for Bob after observing $o \in \mathbf{O}$ is

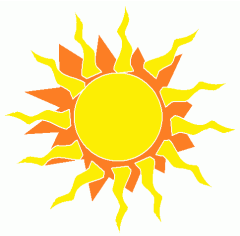
$$EU^a(\text{Bob} \mid o) = \sum_{e \in \mathbf{E}} P(e \mid o) \cdot U^a(\text{Bob} \mid e)$$

Optimal behavior — Given observation o choose the action that leads to maximal expected utility.

$$a^* = \text{argmax}_{a \in \mathbf{A}} EU^a(\text{Bob} \mid o)$$

Computing an Optimal Strategy for Bob

- A strategy for Bob must specify whether to take an umbrella for any possible value of the forecast.
- Suppose forecast predicts sun. What is Bob's expected utility for taking an umbrella ?



Computing Expected Utility for Bob for taking Umbrella

$$EU^{\text{UM}}(\text{Bob} \mid \text{F} = sun) = P(\text{W} = sun \mid \text{F} = sun) \cdot U^{\text{UM}}(\text{Bob} \mid \text{W} = sun) + P(\text{W} = rain \mid \text{F} = sun) \cdot U^{\text{UM}}(\text{Bob} \mid \text{W} = rain)$$

Weather	Umbrella	Utility
sun	*TRUE*	-10
sun	*FALSE*	100
rain	*TRUE*	100
rain	*FALSE*	-10

Marginal probability

$$\begin{aligned} P(\mathsf{F} = sun) =& P(\mathsf{F} = sun \mid \mathsf{W} = sun) \cdot P(\mathsf{W} = sun) + \\ & P(\mathsf{F} = sun \mid \mathsf{W} = rain) \cdot P(\mathsf{W} = rain) \\ =& 0.6 \cdot 0.7 + 0.4 \cdot 0.3 = 0.54 \end{aligned}$$

Bayes Rule

$$\begin{aligned} P(\mathsf{W} = sun \mid \mathsf{F} = sun) =& \frac{P(\mathsf{F} = sun \mid \mathsf{W} = sun) \cdot P(\mathsf{W} = sun)}{P(\mathsf{F} = sun)} \\ =& \frac{0.6 \cdot 0.7}{0.54} = 0.77 \end{aligned}$$

Computing Expected Cost

$$\begin{aligned} EU^{\mathsf{UM}}(\text{Bob} \mid \mathsf{F} = sun) =& P(\mathsf{W} = sun \mid \mathsf{F} = sun) \cdot U^{\mathsf{UM}}(\text{Bob} \mid \mathsf{W} = sun) + \\ & P(\mathsf{W} = rain \mid \mathsf{F} = sun) \cdot U^{\mathsf{UM}}(\text{Bob} \mid \mathsf{W} = rain) \\ =& 0.77 \cdot (-10) + 0.23 \cdot 100 = 15.3 \end{aligned}$$

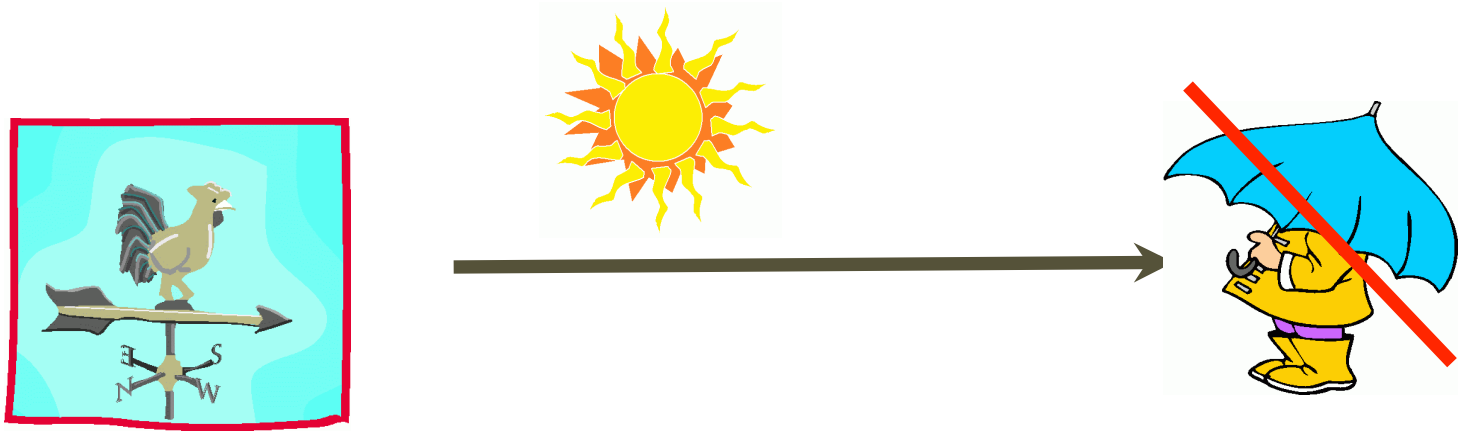
We now compute the expected utility for Bob for the case where Bob does not take an umbrella.

$$\begin{aligned} EU^{\overline{\mathsf{UM}}}(\text{Bob} \mid \mathsf{F} = sun) =& P(\mathsf{W} = sun \mid \mathsf{F} = sun) \cdot U^{\overline{\mathsf{UM}}}(\text{Bob} \mid \mathsf{W} = sun) + \\ & P(\mathsf{W} = rain \mid \mathsf{F} = sun) \cdot U^{\overline{\mathsf{UM}}}(\text{Bob} \mid \mathsf{W} = rain) \\ =& 0.77 \cdot 100 + 0.23 \cdot (-10) = 74.7 \end{aligned}$$

Computing Bob's Best Action

$$EU^{\mathsf{UM}}(\text{Bob} \mid \mathsf{F} = sun) \overset{(15.3)}{<} EU^{\overline{\mathsf{UM}}}(\text{Bob} \mid \mathsf{F} = sun) \overset{(74.7)}{}$$

If the forecast predicts sun, then Bob should not take the umbrella



Computing Bob's Best Action

We now compute Bob's decision for the case where the forecast predicts rain. We have that

$$EU^{\mathsf{UM}}(\text{Bob} \mid \mathsf{F} = rain) \overset{(34)}{<} EU^{\overline{\mathsf{UM}}}(\text{Bob} \mid \mathsf{F} = rain) \overset{(56)}{}$$

We get the following strategy for Bob

	Forecast	
	rain	*sun*
Umbrella	*FALSE*	FALSE

Making Sequential Decisions

The newspaper forecast is more reliable, but costs money, decreasing Bob's utility by 10 units. There are now two decisions:

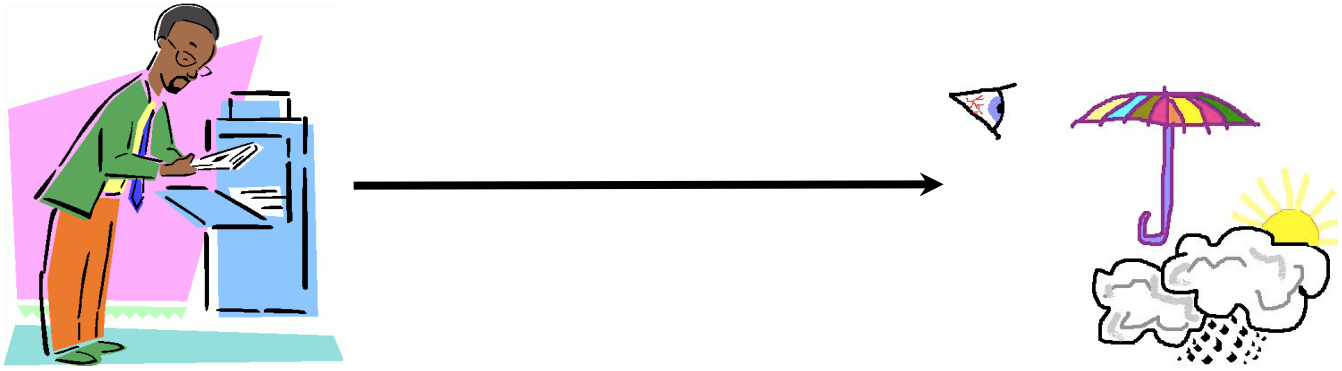
- Buying a newspaper
- Carrying an umbrella

	Forecast	
Weather	*sun*	*rain*
sun	0.8	0.2
rain	0.2	0.8

Weather	NP	Umbrella	Utility
sun	TRUE	*TRUE*	-20
sun	TRUE	*FALSE*	90
rain	TRUE	*TRUE*	90
rain	TRUE	*FALSE*	-20
…	….	….	….

Making Sequential Decisions

- Choosing the best action for one decision depends on the action for the other decision.
- How to weigh the tradeoff between these two decisions ?



Marginal probability

$$\begin{aligned} P^{\mathsf{NP}}(\mathsf{F} = sun) =& P^{\mathsf{NP}}(\mathsf{F} = sun \mid \mathsf{W} = sun) \cdot P(\mathsf{W} = sun) + \\ & P^{\mathsf{NP}}(\mathsf{F} = sun \mid \mathsf{W} = rain) \cdot P(\mathsf{W} = rain) \\ =& 0.8 \cdot 0.7 + 0.2 \cdot 0.3 = 0.62 \end{aligned}$$

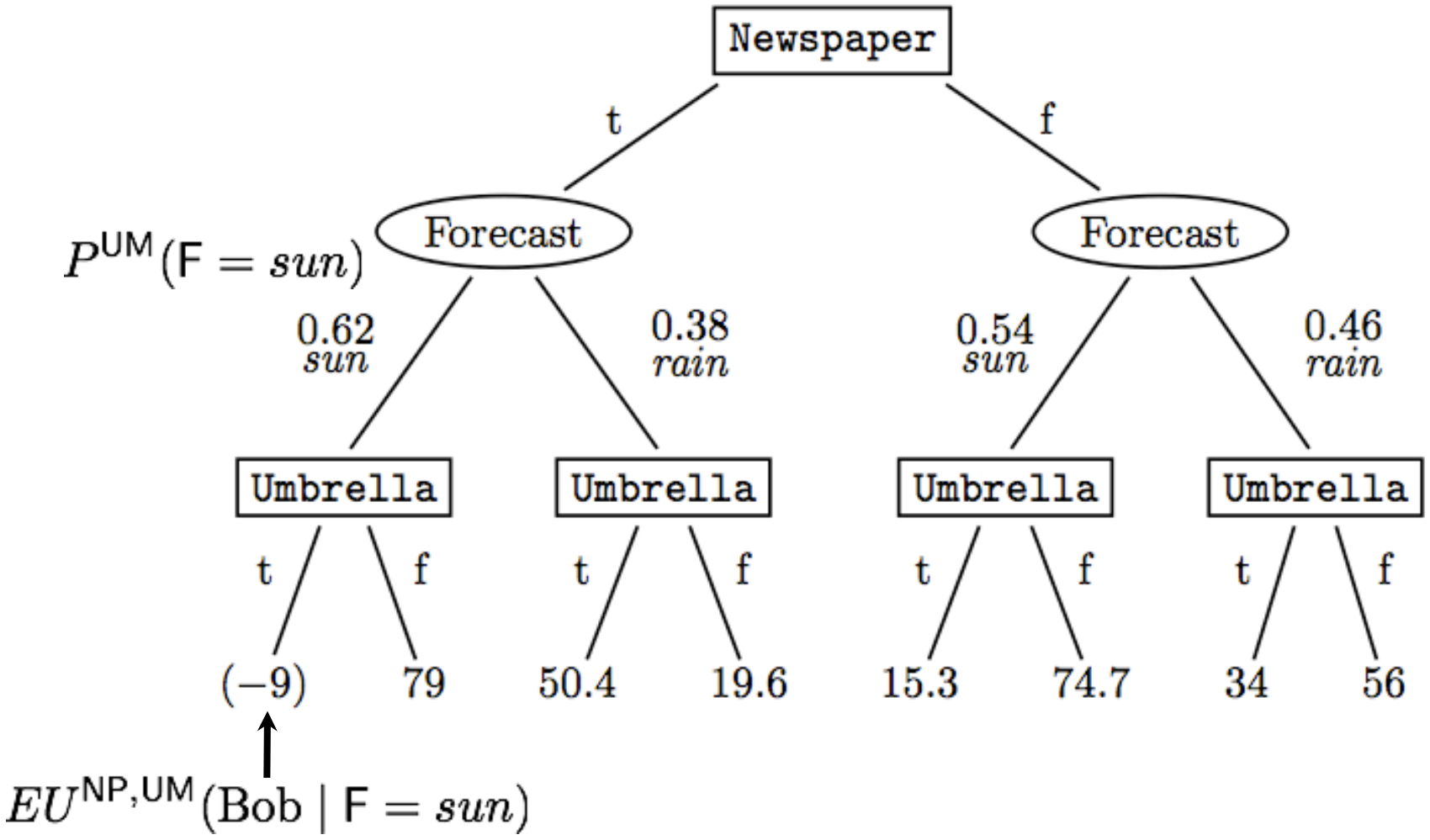
Bayes Rule

$$\begin{aligned} P^{\mathsf{NP}}(\mathsf{W} = sun \mid \mathsf{F} = sun) =& \frac{P^{\mathsf{NP}}(\mathsf{F} = sun \mid \mathsf{W} = sun) \cdot P(\mathsf{W} = sun)}{P^{\mathsf{NP}}(\mathsf{F} = sun)} \\ =& \frac{0.8 \cdot 0.7}{0.62} = 0.90 \end{aligned}$$

Expected utility

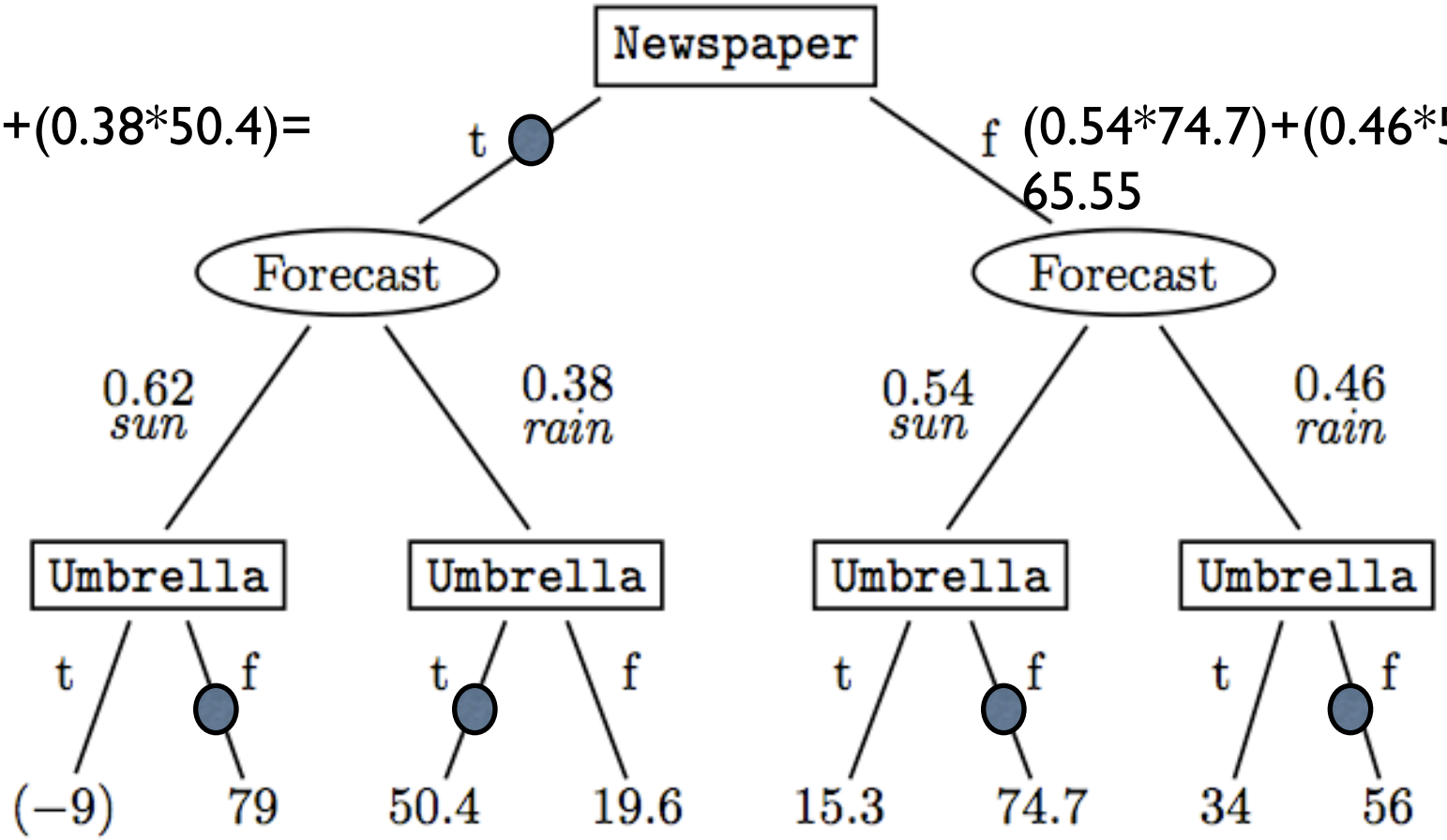
$$\begin{aligned} EU^{\mathsf{NP},\mathsf{UM}}(\text{Bob} \mid \mathsf{F} = sun) =& P^{\mathsf{NP}}(\mathsf{W} = sun \mid \mathsf{F} = sun) \cdot U^{\mathsf{UM}}(\text{Bob} \mid \mathsf{W} = sun) + \\ & P^{\mathsf{NP}}(\mathsf{W} = rain \mid \mathsf{F} = sun) \cdot U^{\mathsf{UM}}(\text{Bob} \mid \mathsf{W} = rain) \\ =& 0.90 \cdot (-20) + 0.10 \cdot 90 = (-9) \end{aligned}$$

Decision Trees



Solving Decision Trees

(0.62*79)+(0.38*50.4)=
68.132



(0.54*74.7)+(0.46*56)=
65.55

Acknowledgements

- Slide 3: https://www.nasa.gov/sites/default/files/thumbnails/image/pia19808-main_tight_crop-monday.jpg
- Slide 4: https://www.nasa.gov/sites/default/files/thumbnails/image/pia19399_msl_mastcammosaiclocations.jpg
- Slide 5: https://ichef.bbci.co.uk/news/624/media/images/55165000/jpg/_55165401_exomarssimulation.jpg
- Core examples are from F.S. Hillier, G.J. Lieberman, Operations Research, 1994. (esp. Ch 6 and 12)

Acknowledgements

- Much of the decision theory discussion is a paraphrased and condensed version of chapters in Luce and Raiffa's excellent book on Games and Decisions
- The Bayes choice example in the last section is taken from a tutorial by Gal and Pfeffer at AAAI 2008.