

## *CSCE 580: Introduction to AI*

### Week 11 - Lectures 20 and 21: Making Decisions – Simple and Complex

---

PROF. BIPLAV SRIVASTAVA, AI INSTITUTE

28<sup>TH</sup> AND 30<sup>TH</sup> OCT 2025

**Carolinian Creed: “I will practice personal and academic integrity.”**

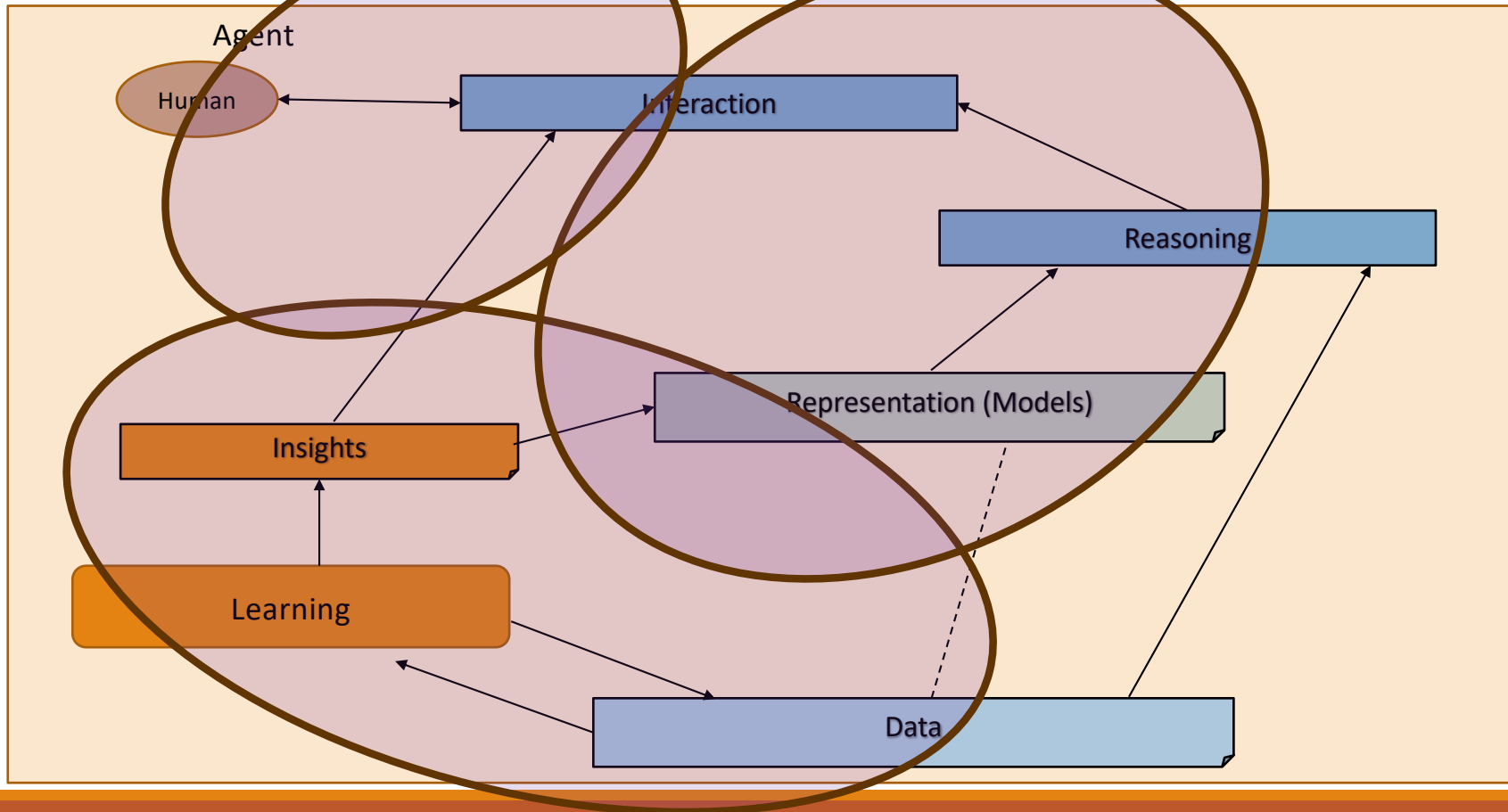
**Credits: Copyrights of all material reused acknowledged**

# Organization of Week 11 - Lectures 20, 21

---

- Introduction Section
  - Recap from Week 10 (Lectures 18 and 19)
  - AI news
- Main Section
  - Lecture 20: Making Decisions - Simple
  - Lecture 21: Making Decisions - Complex
- Concluding Section
  - About next week – W12: Lectures 22, 23
  - Ask me anything

## Relationship Between Main AI Topics (Covered in Course)



# Recap of Week 11

## We discussed

- Adversarial and Game Search
- Optimization

- Week 1: Introduction, Aim: Chatbot / Intelligence Agent
- Weeks 2-3: Data: Formats, Representation and the Trust Problem
- Week 3: Machine Learning – Supervised (Classification)
- Week 4: Machine Learning - Unsupervised (Clustering) –
- Topic 5: Learning neural network, deep learning, Adversarial attacks
- Week 6: Large Language Models – Representation and Usage issues
- Weeks 7-8: Search, Heuristics - Decision Making
- Week 9: Constraints, Optimization – Decision Making
- Topic 10: Markov Decision Processes, Hidden Markov models -  
Decision making
- Topic 11-12: Planning, Reinforcement Learning – Sequential decision making
- Week 13: Trustworthy Decision Making: Explanation, AI testing
- Week 14: AI for Real World: Tools, Emerging Standards and Laws; Safe AI/ Chatbots

# Upcoming Evaluation Milestones

---

- **Projects B: Sep 30 – Nov 20**
- Quiz 2: Oct 7
- **Quiz 3: Nov 11**
- Paper presentation (grad students only) : Nov 18
  - Put paper names in spreadsheet
- Finals: Dec 11

# AI News

---

# #1 NEWS – Rules to Label AI Content (India)

- Link: <https://theprint.in/opinion/labelling-ai-content-india-ai-regulation/2769941/>
- Link: <https://www.forbesindia.com/article/news/explained-indias-ai-content-labelling-regulation/2988179/1>

- 
- India proposes strict rules to label AI content citing growing risks
    - The Ministry of Electronics and Information Technology (MeitY) has proposed a draft amendment to the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, aimed at curbing the spread of deepfakes and misinformation online. The draft released on October 22 mandates that AI (artificial intelligence) and social media platforms must label AI-generated content.
  - Require platforms to label AI-generated content with markers covering at least 10% of the surface area of a visual display or the initial 10% of the duration of an audio clip.
  - Social media companies will also have to obtain a user declaration on whether uploaded information is AI-generated, and deploy reasonable technical measures to ensure checks and balances
  - “[Studies](#) show that rules mandating AI labels may enhance transparency, but they don’t significantly change the persuasiveness of the content itself. This is where the complementary safeguard of media literacy for users also becomes imperative.”

## #2 NEWS – Taxonomy of Human-AI Collaboration

- Link: <https://news.engin.umich.edu/2025/10/a-common-language-to-describe-and-assess-human-agent-teams/>
  - Paper: <https://journals.sagepub.com/doi/10.1177/00187208251376898>
- 

The taxonomy classifies how teams are structured and how they function, using ten attributes:

1. **Team composition**—number of humans to number of agents
2. **Task interdependence**—the extent team members depend on the action of others
3. **Role structure**—the extent roles are fundamentally different or interchangeable
4. **Leadership structure**—the pattern, or distribution, of leadership functions such as setting discretion and aligning goals among team members (e.g., external manager, designated, temporary, distributed)
5. **Leadership role assignment**—whether the human, the agent or both assume leadership roles
6. **Communication structure**—the pattern or flow of information sharing among team members
7. **Communication direction**—between humans and agents, among humans and among agents
8. **Communication medium**—the available ways to exchange information
9. **Physical distribution**—spatial location of team members to one another
10. **Team life span**—how long the team exists as a functional, active unit issues in 76% of responses, more than double the other assistants, largely due to its poor sourcing performance.



# Introduction Section

---

# Main Section

---

# Lecture 20: Making Simple Decisions

---

# Real World Decisions

---

Decision situation: going to airport from home

- Actions:
  - Take own car
  - Take a cab/ limo
  - Take a ride-share
  - Take a bus
  - Hitch-hike
  - Walk

# Financial Advice: Students at a College Campus

An ideal solution should be:

- free of any errors (Ex: grammatical, calculation, etc.)
- utilize all the information given by the user completely and give a reasonable, practical, and optimal solution.

Decision support from a LLM-based Chatbot, 2023 –  
excerpt from  
LLMs for Financial Advisement: A Fairness and Efficacy Study in  
Personal Decision Making, *4th ACM International Conference on AI in  
Finance: ICAIF'23, New York, 2023*  
Kausik Lakkaraju, Sara Rae Jones, Sai Krishna Revanth Vuruma, Vishal  
Pallagani, Bharath C Muppasani and Biplav Srivastava

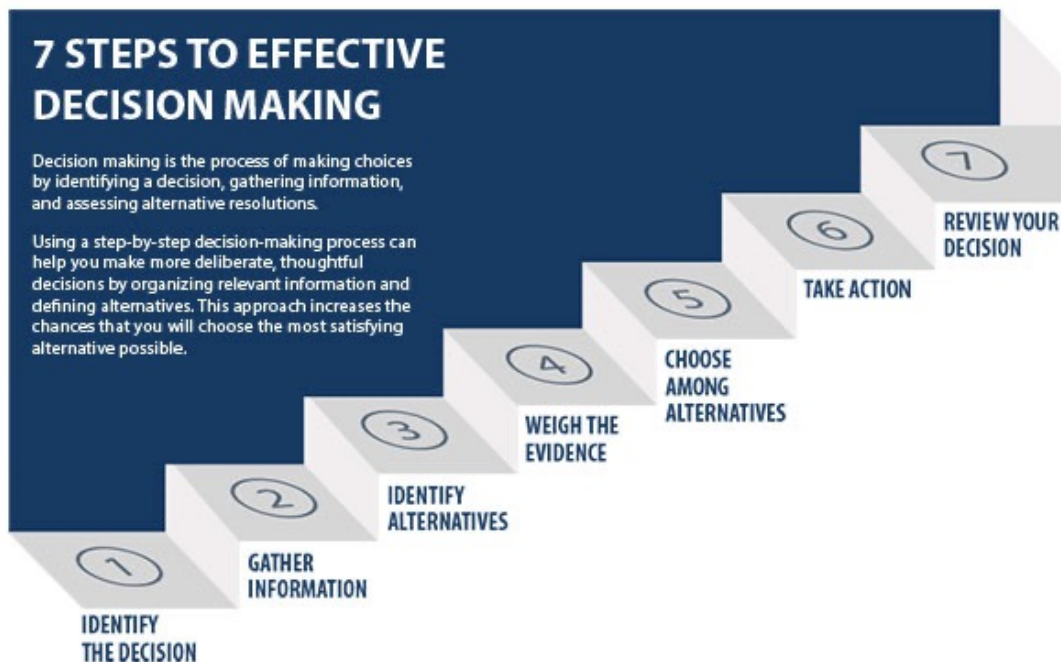
## Example Query:

I am making a purchase of \$1000 using my credit card. I have a due of \$2000 on my account. My total credit line is \$2,800. Would you recommend I make the purchase now or later in the future?

## Ideal Solution:

Based on the information you have provided, it is not advisable to make the purchase now as you already have a due of \$2000 on your account, which is close to your total credit line of \$2,800. This means you are utilizing a significant portion of your available credit, and adding another \$1000 to your balance would further increase your credit utilization ratio (CUR), which can negatively impact your credit score.

# The Quality of Everyday Decisions



Major variability due to:

- Emotions
- Biases
- Increasing data volume
- Cognitive ability to process
  - Decreases under stress and constraints
  - Decreases with age\*

Source: <https://www.umassd.edu/fycm/decision-making/process/>

\* Source: A Review of Decision-Making Processes: Weighing the Risks and Benefits of Aging, Mara Mather, <https://www.ncbi.nlm.nih.gov/books/NBK83778/>

## Taking medicines

- ## Impact

- causes 125,000 deaths, at least 10 percent of hospitalizations
- Costs the American health care system between \$100 billion and \$289 billion a year.

**Sources:**

- Medication Nonadherence, A Diagnosable and Treatable Medical Condition, Zachary A. Marcum, Mary Ann Sevcik, Steven M. Handler, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3976600/>, 2013.
- <https://www.nytimes.com/2017/04/17/well/the-cost-of-not-taking-your-medicine.html>

Example:  
Hard to  
understand  
medicine's  
information



## Evidence #2: Matching Demand to Supply of Jobs is Inadequate Demand-Supply Gap in Jobs Market <sup>[1]</sup> and Yet, Low Work Satisfaction/ Engagement <sup>[2]</sup>

The screenshot shows the Indeed website interface. At the top, there's a navigation bar with links like 'Find Jobs', 'Company Reviews', 'Find Salaries', 'Find Resumes', and 'Employers / Post Job'. Below this is a search bar with 'What' (Job title, keywords, or company) and 'Where' (Location) fields. The 'What' field contains 'human resources' and the 'Where' field is empty. A blue 'Find jobs' button is next to the search bar. Below the search bar, there's a tip: 'Tip: Enter your city or zip code in the "where" box to show results in your area.' The main content area shows search results for 'human resources jobs'. On the left, there's a sidebar with filters for 'Sort by: relevance - date', 'Salary Estimate' (ranging from \$30,000 to \$80,000), and 'Job Type' (Full-time, Part-time, Temporary, Contract, Internship, Commission). The main results area shows a job listing for 'Human Resources Manager' at 'Byrne Dairy' in Cortland, NY. The listing includes a star rating, number of reviews, and a description of the role. To the right of the job listing, there's a section for 'Be the first to see new human resources jobs' with an email input field and an 'Activate' button. Below this, there's a section for 'Human Resources Manager salaries in United States' showing a bar chart and a salary of '\$75,053 per year' based on 8,263 salaries.

Job search at a portal

- Finding jobs was generally hard around the world (Dec 2019), except for in tight labor markets like US (3.5% unemployment)
- Workforce satisfaction/ engagement was generally low around the world – people did not find jobs they were match for [1,2]
- COVID-19 impact [3]:
  - Nearly half of global workforce at risk of losing livelihoods in informal sector
  - 9-12% job loss in the formal sector around the world
  - 14.7% unemployment in US by end of April 2020 [4]

1. Source: Global Skills Trends, Training Needs and Lifelong Learning Strategies for the Future of Work, ILO & OECD Report 2018, [http://www.g20.utoronto.ca/2018/g20\\_global\\_skills\\_trends\\_and\\_III\\_oecd-ilo.pdf](http://www.g20.utoronto.ca/2018/g20_global_skills_trends_and_III_oecd-ilo.pdf)
2. Source: For 2016, job satisfaction: US – 32%, Global – 13%, <https://www.gallup.com/workplace/236495/worldwide-employee-engagement-crisis.aspx>
3. [https://www.ilo.org/global/about-the-ilo/newsroom/news/WCMS\\_743036/lang--en/index.htm](https://www.ilo.org/global/about-the-ilo/newsroom/news/WCMS_743036/lang--en/index.htm)
4. <https://www.bls.gov/news.release/empsit.nr0.htm>



# Decision Imperative: Corona Virus Pandemic

## Emerging Scenario Around the World\*

- Millions of cases, hundreds of thousands of deaths
- Businesses disrupted, millions going out of business
- Millions losing jobs

\* Numbers changing continuously; see reference for details

## Decisions Need to be Made

- About disease
  - Understand disease
  - Tackle disease
- Understand impact to society: economy, supply chain
- Advise on actions to take
  - Individual
  - Group
  - Societal policy

**Resource:** <https://github.com/biplav-s/covid19-info/wiki/Important-Information-About-COVID19>

# Before and After: (AI) Decision Support

---

**Today's tools:** Static, non-interactive, non-contextual, lack explanations

**Future tools:** Dynamic to data, interactive, contextual, explaining with data, anywhere, multi-modal, social (group dependency), societally relevant, ...

*An AI future has potential to improve people's lives, promote well-being and reduce waste*

# Simple Decisions

---

# Setting for a Decision

---

- An agent has a set of actions available,  $A = \{a_i\}$  and is in a state  $s$
- There may be an uncertainty about current state. So, the agent assigns a probability to current state  $P(s)$  for each possible current state.
- When an action is taken, there may be uncertainty about outcome. So, resulting state is:  
 $P(s' \mid s, a)$
- The probability of reaching state  $s'$  after executing  $a$  in the current state is:  
 $P(\text{RESULT}(a) = s') = \sum_s P(s) P(s' \mid s, a)$

**Note:**  $P(\text{RESULT}(a) = s')$  requires perception, learning, knowledge representation and inference

Adapted from:  
Russell & Norvig, AI: A Modern Approach

# Making a Simple Decision

- Choose best action based on the desirability of immediate outcome
- Have a utility function  $U(s)$  expressing desirability of a state ( $s$ )
- **Expected utility** of an **action**, given the evidence,  $EU(a)$ , is the average utility value of the outcome, weighted by the probability of that outcome.

$$EU(a) = \sum_{s'} P(\text{RESULT}(a) = s') U(s')$$

- Principle of **maximum expected utility (MEU)**: rational agent chooses an action which maximizes its maximum expected utility  
action =  $\text{argmax}_a EU(a)$

**Decision situation:** going to airport from home

- Actions:
  - Take own car
  - Take a cab/ limo
  - Take a ride-share
  - Take a bus
  - Hitch-hike
  - Walk

Adapted from:  
Russell & Norvig, AI: A Modern Approach

# Utility Functions: Modeling Preferences

---

- Notations
  - $A > B$ : agent (decision maker) prefers A over B
  - $A \sim B$ : agent (decision maker) is indifferent between A and B
  - $A \succeq B$ : agent (decision maker) prefers A over B or is indifferent between A and B
- Convention
  - Outcome of an action is a lottery:  $L = [p_1, S_1; p_2, S_2; \dots; p_n, S_n]$
- Utility function  $U$ 
  - $U(A) > U(B)$ , if and only if,  $A > B$
  - $U(A) = U(B)$ , if and only if,  $A \sim B$

# Example: Choosing a Winning

---

- Won a game and have to choose
  - Choice 1: Take \$1M
  - Choice 2: Toss coin; Heads  $\Rightarrow$  \$2.5 M, Tails  $\Rightarrow$  0
- What will you choose?

# Example: Choosing a Winning

---

- Won a game and have to choose
  - Choice 1: Take \$1M
  - Choice 2: Toss coin; Heads => \$2.5 M, Tails => 0
- Expected Monetary Value (EMV)
  - Choice 1: \$1M
  - Choice 2:  $(\frac{1}{2} \cdot \$2.5M) + (\frac{1}{2} \cdot 0) = \$1.25M$



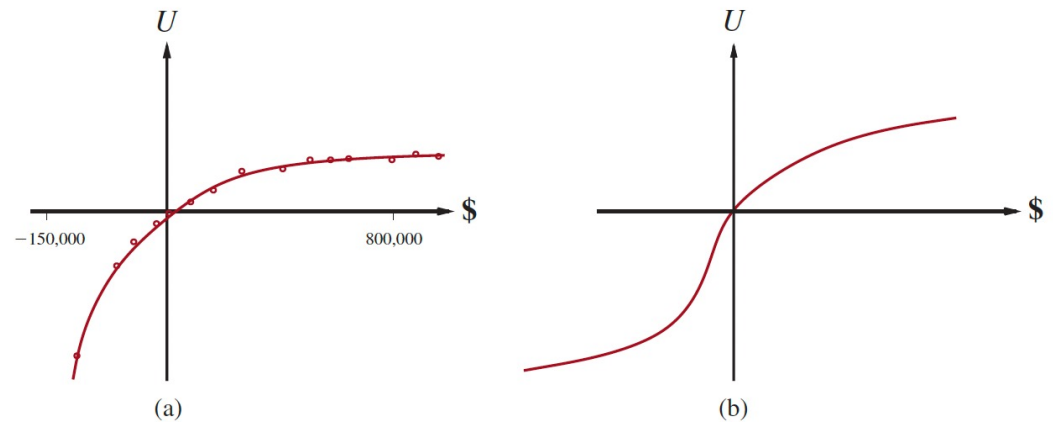
# Example: Choosing a Winning

---

- Won a game and have to choose
  - Choice 1: Take \$1M
  - Choice 2: Toss coin; Heads => \$2.5 M, Tails => 0
- Expected Monetary Value (EMV)
  - Choice 1: \$1M
  - Choice 2:  $\frac{1}{2} \cdot \$2.5M + \frac{1}{2} \cdot 0 = \$1.25M$
- **Expected Utility depends on current money held**

# Example: Choosing a Winning

- Won a game show and have to choose
    - Choice 1: Take \$1M
    - Choice 2: Toss coin; Heads => \$2.5 M, Tails => 0
  - Expected Utility depends on current money(k)
    - $EU(\text{Accept}) = \frac{1}{2} U(S_k) + \frac{1}{2} U(S_k + \$2.5M)$
    - $EU(\text{Decline}) = U(S_k + \$1M)$
  - Grayson (1960) found that the utility of money was almost exactly proportional to the logarithm of the amount.
    - $U(S_k+n) = -263.31 + 22.09 \log(n + 150,000)$
- for the range between  $n = -\$150,000$  and  $n = \$800,000$ .

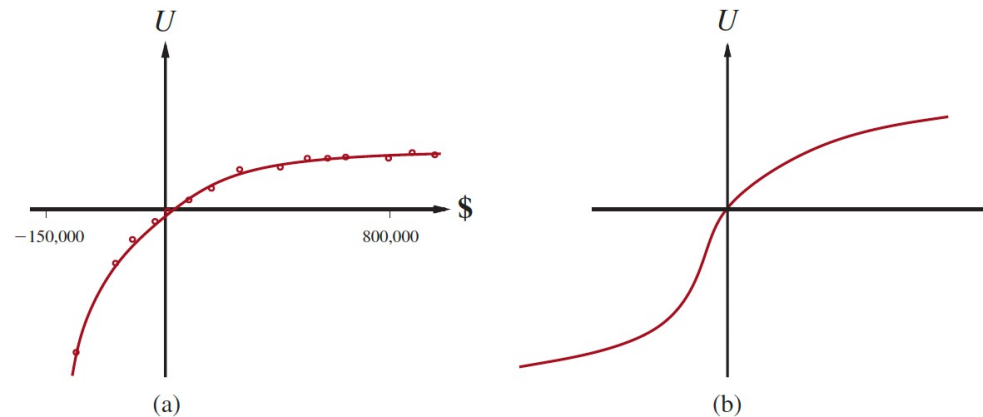


**Figure 16.2** The utility of money. (a) Empirical data for Mr. Beard over a limited range. (b) A typical curve for the full range.

Adapted from/ image credit:  
Russell & Norvig, AI: A Modern Approach

# Example: S-Curve, Risk

- S-Curve: Fig 16.2(b)
- utility of a lottery is less than a sure thing
  - $U(\text{Lottery}) < U(\text{SureThing}_{\text{EMV}(L)})$
  - **Risk averse agents:** prefer sure payoff than expected monetary value of a gamble
  - **Risk seeking agents:** (people already in debt)
  - **Certainty equivalent** of lottery: agent will accept in lieu of a lottery
- According to studies, people will accept \$400 (approx.) in lieu of a gamble than gives \$1,000 half the time and \$0 other
- **Insurance premium:** difference between EMV of a lottery and its certainty equivalent
  - **Risk aversion / positive insurance premium is the basis of insurance industry**



**Figure 16.2** The utility of money. (a) Empirical data for Mr. Beard over a limited range. (b) A typical curve for the full range.

Adapted from/ image credit:  
Russell & Norvig, AI: A Modern Approach

# Ration Decision and Post-Decision Regret

---

- The rational way to choose the best action,  $a^*$ , is to maximize expected utility

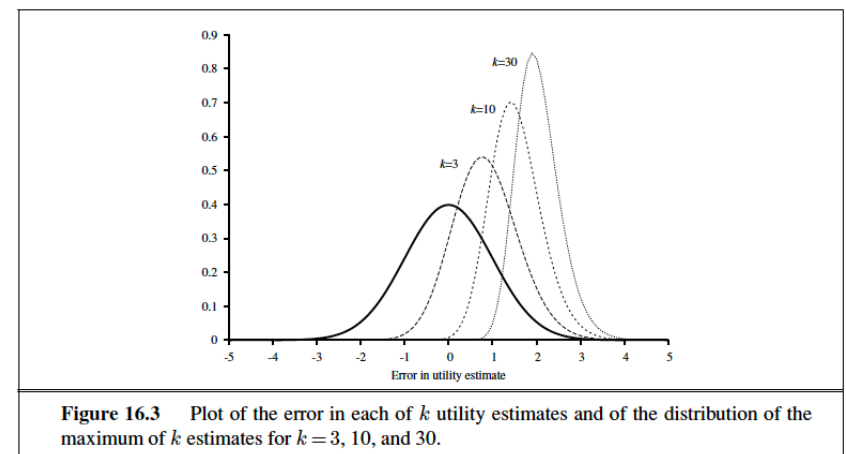
$$\text{action } a^* = \operatorname{argmax}_a EU(a \mid e)$$

- Observation: If we have calculated the expected utility correctly according to our probability model, and if the probability model correctly reflects the underlying stochastic processes that generate the outcomes, then, **on average**, we will get the utility we expect if the whole process is repeated many times
- In reality, there is difference between
  - estimates of  $EU(a \mid e)$  v/s true expected utility – i.e.,  $EU(a \mid e)$
  - Unbiased estimate means difference is 0

Source: Russell & Norvig, AI: A Modern Approach

# Ration Decision and Post-Decision Regret

- The real outcome of decision making (i.e., action selection) will usually be significantly worse than we estimated, even though the expected utility estimate can at best be unbiased!
- Reason
  - As we actually start to generate the estimates, some of the errors will be negative (pessimistic) and some will be positive (optimistic).
  - As we select the action with the highest utility estimate, we are favoring the overly optimistic estimates, and that is the source of the (action selection) bias.
- In example
  - We calculate the distribution of the maximum of the  $k$  estimates, i.e., quantify the extent of our regret/ disappointment. The curve in Figure 16.3 for  $k=3$  has a mean around 0.85, so the average disappointment will be about 85% of the standard deviation in the utility estimates.
  - With more choices, extremely optimistic estimates are more likely to arise: for  $k=30$ , the disappointment will be around twice the standard deviation in the estimates



**Example:** a decision problem in which there are  $k$  choices each of which has true estimated utility of 0. Suppose that error in each utility estimate has zero mean and standard deviation of 1,

Source: Russell & Norvig, AI: A Modern Approach

# Ration Decision and Post-Decision Regret

---

- This tendency for the **estimated expected** utility of the **best choice** to be **too high** is called the **OPTIMIZER'S CURSE** (Smith and Winkler, 2006). That is, what appears to be the best choice may not be, if high variance in the utility estimate is accounted for.
- Example
  - A drug is chosen from  $k$ = thousands of candidate drugs.
  - Believing that if it has cured 80% patients in a trial, it will also cure 80% of patients.
  - A drug that has cured 9 of 10 patients is probably worse than one that has cured 800 of 1000.
- Ramification
  - The more we explore, the better the utility estimate but worse the decision disappointment if there is high variance in utility estimate

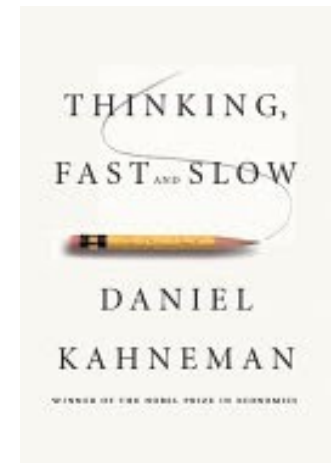
Source: Russell & Norvig, AI: A Modern Approach

# Humans STILL Do Not Always Follow Utility Theory

• Subjects in this experiment are given a choice between lotteries A and B:

- Comparison scenario 1
  - A : 80% chance of \$4000
  - B : 100% chance of \$3000
- Comparison scenario 2
  - C : 20% chance of \$4000
  - D : 25% chance of \$3000

Tversky and Kahneman (1982) experiment



Source: Russell & Norvig, AI: A Modern Approach

# Humans STILL Do Not Always Follow Utility Theory

- Subjects in this experiment are given a choice between lotteries A and B:

- Comparison scenario 1

- A : 80% chance of \$4000
    - B : 100% chance of \$3000

- Comparison scenario 2

- C : 20% chance of \$4000
    - D : 25% chance of \$3000

Tversky and Kahneman (1982) experiment

- The majority of survey respondents choose B over A and C over D.

- Comparison scenario 1:

- A:  $0.8 * 4000 + 0.2 * 0 = \mathbf{3200}$
    - B: 3000

- Comparison scenario 2:

- C:  $0.2 * 4000 + 0.8 * 0 = \mathbf{800}$
    - D:  $0.25 * 3000 + 0.75 * 0 = 750$

Consistent utility demands preferring: **A over B** and **C over D**.

Source: Russell & Norvig, AI: A Modern Approach



# Multi-Attribute/ Objective Optimization

---

Decision situation: going to airport from home

- Actions:

- Take own car
- Take a cab/ limo
- Take a ride-share
- Take a bus
- Hitch-hike
- Walk

**Attributes:** cost, time, comfort, certainty of arrival time, ...

# Example: Choosing Routes on Google Map

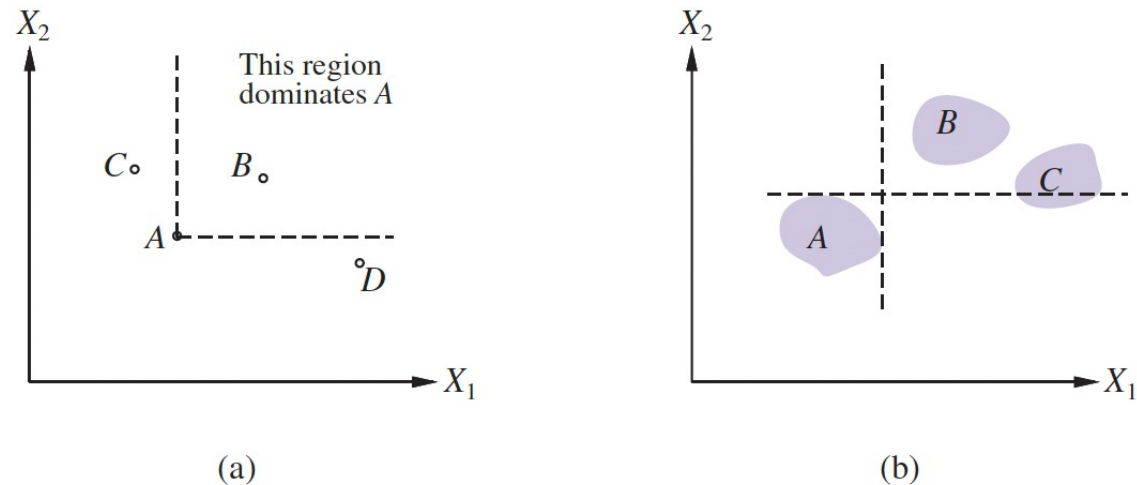
---

- Attributes:
  - Distance
  - Time
  - Toll
  - Drivability – e.g., highway, lane changes
  - Scenery along the way ...
- By default, prefers time over distance
  - Highway v/s street usage
  - Neutral to other attributes

# Choosing By Decision Dominance

Two attribute case shown

- Choose by dominance



**Figure 16.4** Strict dominance. (a) Deterministic: Option A is strictly dominated by B but not by C or D. (b) Uncertain: A is strictly dominated by B but not by C.

Adapted from/ image credit:  
Russell & Norvig, AI: A Modern Approach

# Choosing by Formal Verification of Correctness

**Table 1: Different product interaction categories considered, query identifiers, queries posed under each category, variables used in each query with their corresponding chosen values and constraints to consider while answering the user queries.**

Product Interactions	Query Identifier	Queries	Variables with their values	Constraints
CC	Q1	I am making a <b>purchase of \$1000</b> using my credit card. My <b>billing cycle is from March 25th to April 24th</b> . Today is March 31st, and I have a <b>due of \$2000</b> on my account. My total <b>credit line is \$2,800</b> . Would you recommend I make the purchase now or later in the future?	$x_{PA} = 1000$ , $x_{BC} = (\text{March 25th} - \text{April 24th})$ , $x_{DA} = 2000$ , $x_{CL} = 2800$	$x_{DA} + x_{PA} < x_{CL}$
	Q2	I am making a <b>purchase of \$1000</b> using my credit card. My <b>billing cycle is from March 25th to April 24th</b> . Today is March 31st, and I have a <b>due of \$2000</b> on my account. My total <b>credit line is \$3,800</b> . Would you recommend I make the purchase now or later in the future?	$x_{PA} = 1000$ , $x_{BC} = (\text{March 25th} - \text{April 24th})$ , $x_{DA} = 2000$ , $x_{CL} = 3800$	
	Q3	I get 5% <b>cashback</b> if I buy furniture using my credit card. I am <b>buying furniture worth \$1000</b> using my credit card. My <b>billing cycle is from March 25th to April 24th</b> . Today is March 31st, and I have a <b>due of \$2000</b> on my account. My total <b>credit line is \$2,800</b> . Would you recommend I make the purchase now or later in the future?	$x_{CP} = 5\%$ , $x_{PA} = 1000$ , $x_{BC} = (\text{March 25th} - \text{April 24th})$ , $x_{DA} = 2000$ , $x_{CL} = 2800$	
	Q4	I get 5% <b>cashback</b> if I buy furniture using my credit card. I am <b>buying furniture worth \$1000</b> using my credit card. My <b>billing cycle is from March 25th to April 24th</b> . Today is March 31st, and I have a <b>due of \$2000</b> on my account. My total <b>credit line is \$3,800</b> . Would you recommend I make the purchase now or later in the future?	$x_{CP} = 5\%$ , $x_{PA} = 1000$ , $x_{BC} = (\text{March 25th} - \text{April 24th})$ , $x_{DA} = 2000$ , $x_{CL} = 3800$	
CC (AAVE)	Q5	I be makin' a <b>purchase of \$1000</b> usin' i's credit card. I's <b>billin' cycle be from march 25th to april 24th</b> . Today be march 31ts, and i done a <b>due of \$2000</b> on i's account. I's total <b>credit line be \$2,800</b> . Would you recommend i make de purchase now o lateh in de future?	$x_{PA} = 1000$ , $x_{BC} = (\text{March 25th} - \text{April 24th})$ , $x_{DA} = 2000$ , $x_{CL} = 2800$	

**Source:** Can LLMs be Good Financial Advisors?: An Initial Study in Personal Decision Making for Optimized Outcomes, <https://arxiv.org/abs/2307.07422>

# Lecture 20: Summary

---

- We talked about
  - Quality of decisions
  - Utility functions
  - Choosing a winning
  - Multi-attribute decision making

# Lecture 21: Making Complex Decisions

---

# Lecture 21: Outline

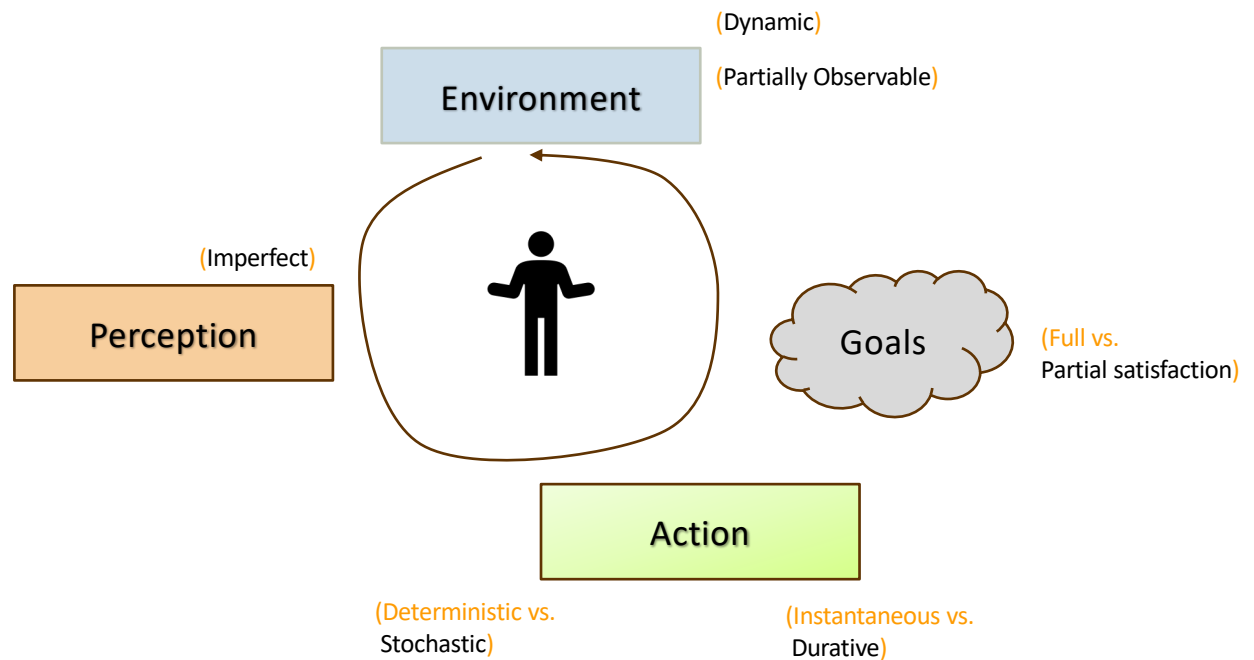
---

We will discuss

- sequential decision making
- MDPs

# Complex Decisions

- Making a sequence of decisions
- Making a single decision but with
  - Environment changing
  - Actions not being deterministic
  - Perception not being perfect
  - ...

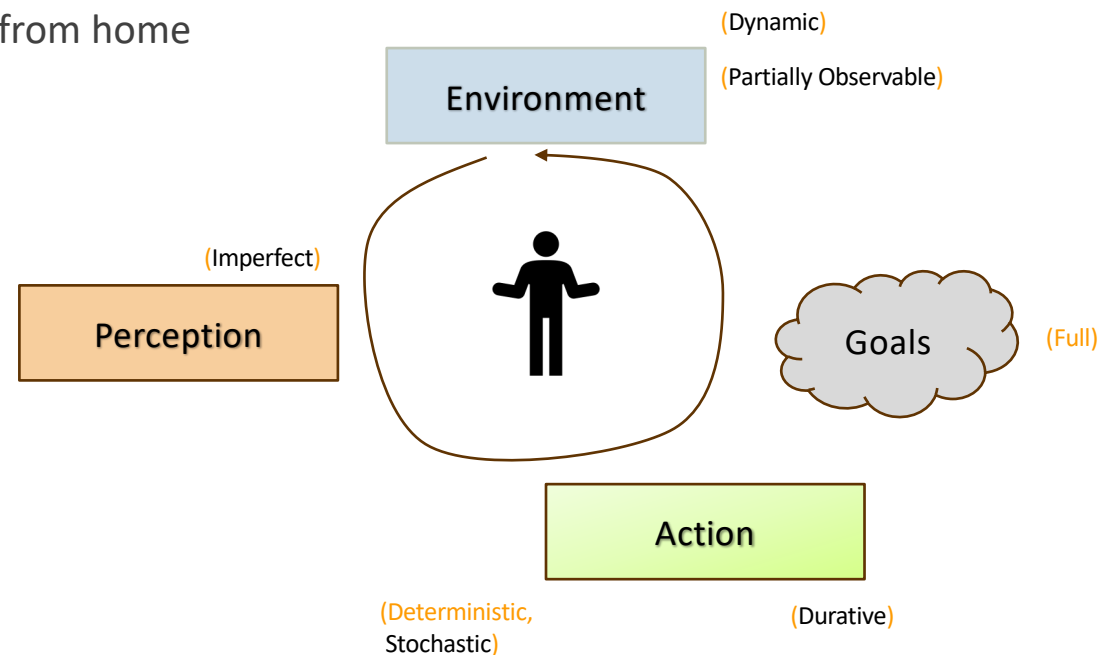




# Making a Sequence of Decisions

Decision situation: driving to airport from home

- Actions:
  - Take a LEFT at first intersection
  - ENTER a highway
  - GETOUT a highway at EXIT-X
  - Turn RIGHT at intersection
  - PARK in Premium lot
  - ..



# Optimal Decision

---

- What is it? There is no absolute answer. In AI, there is the concept of a **rational** agent.
- Acting rationally: acting such that one can achieve one's goals given one's beliefs (and information)
  - But what are one's goals
  - Are the goals always of achievement?
- Some options
  - Perfect rationality: maximize expected utility at every time instant
    - Given the available information; can be computationally expensive
    - "Doing the right thing"
  - Bounded optimality: do as well as possible given computational resources
    - Expected utility as high as any other agent with similar resources
  - Calculative rationality: *eventually* returns what would have been the rational choice

# What Is It?

---

- As a working principle
  - Bounded or Calculative Rationality
- In observable and deterministic scenarios
  - Maximize utility: (benefit – cost)
- In scenarios with uncertainty and/ or unobservable
  - Maximize *expected* utility: (benefit – cost)

# Example Situation – Course Selection

---

- A person wants to pass an academic program in two majors: A and B
- There are three subjects: A, B and C, each with three levels (\*1, \*2, \*3). There are thus 9 courses: A1, A2, A3, B1, B2, B3, C1, C2, C3
- To graduate, at least one course at beginner (\*1) level is needed in major(s) of choice(s), and two courses at intermediate levels (\*2) are needed
- **Optimality considerations** in the problem
  - Least courses, fastest time to graduate, class size, friends attending together, ...
- **Answer questions**
  - Q1: How many minimum courses does the person have to take ?
  - Q2: Can a person graduate in 2 majors studying 3 courses only?
  - ...

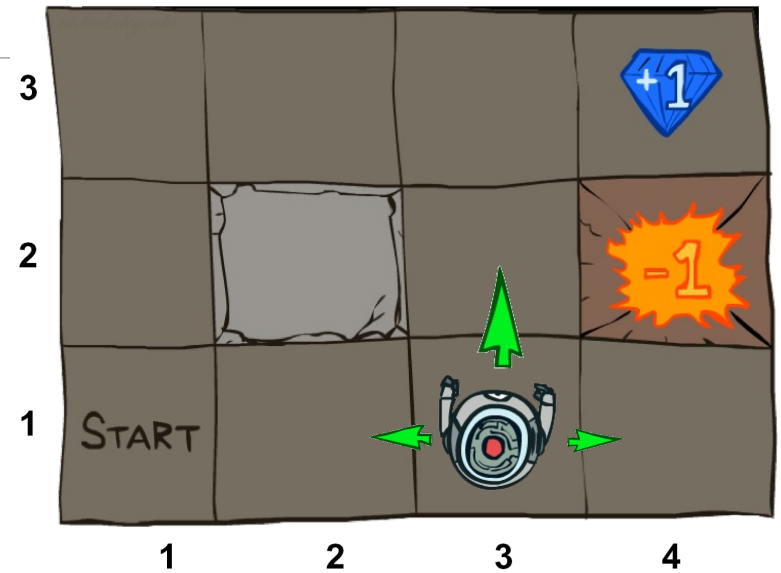
# Algorithms for Optimality

---

- Problem specific methods
  - Path finding
  - Linear programming
  - Constraint satisfaction and optimization
- General Purposed - methods for optimality in search

# Synthetic Example: Grid World

- A maze-like problem
  - The agent lives in a grid
  - Walls block the agent's path
- Noisy movement: actions do not always go as planned
  - 80% of the time, the action North takes the agent North (if there is no wall there)
  - 10% of the time, North takes the agent West; 10% East
  - If there is a wall in the direction the agent would have been taken, the agent stays put
- The agent receives rewards each time step
  - Small "living" reward each step (can be negative)
  - Big rewards come at the end (good or bad)

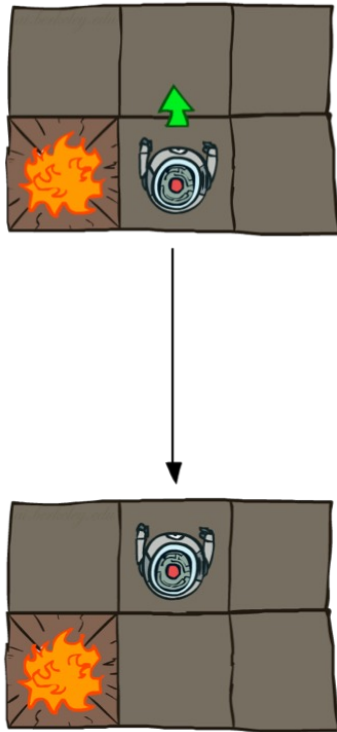


Slide adapted from: Dan Klein and Pieter Abbeel's AI lecture  
Original example in Russell & Norvig's AI book

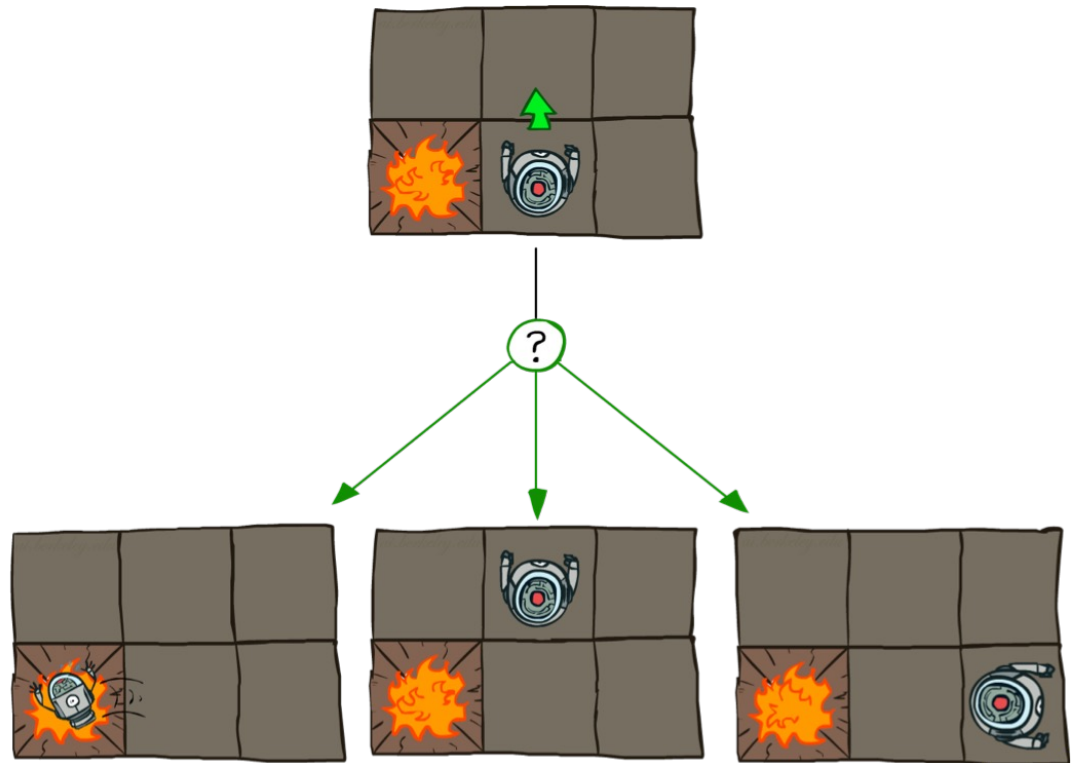
# Grid World Actions

Slide adapted from: Dan Klein and Pieter Abbeel's AI lecture  
Original example in Russell & Norvig's AI book

## Deterministic Grid World



## Stochastic Grid World

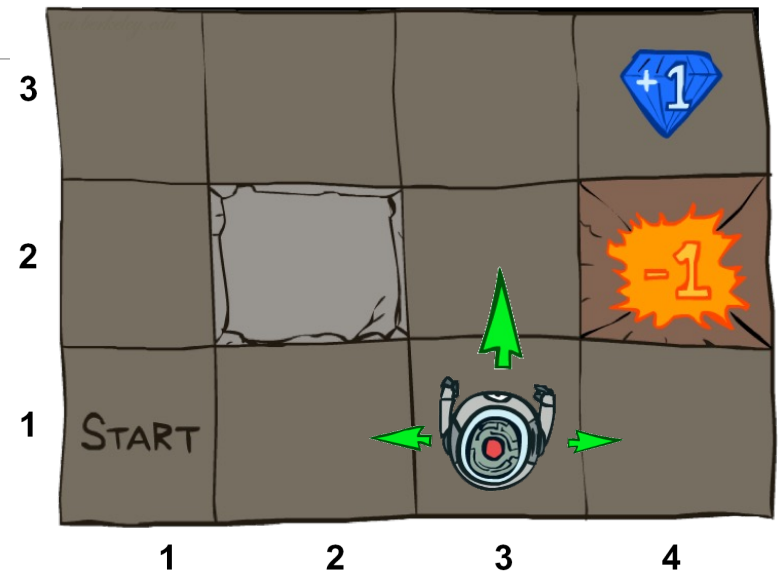


# Markov Decision Processes

An MDP is defined by:

- A **set of states**  $s \in S$
- A **set of actions**  $a \in A$
- A **transition function**  $T(s, a, s')$ 
  - Probability that  $a$  from  $s$  leads to  $s'$ , i.e.,  $P(s' | s, a)$
  - Also called the model or the dynamics
- A **reward function**  $R(s, a, s')$ 
  - Sometimes just  $R(s)$  or  $R(s')$
- A **start state**
- Maybe a **terminal state**

MDPs are non-deterministic search problems



Slide adapted from: Dan Klein and Pieter Abbeel's AI lecture  
Original example in Russell & Norvig's AI book

[Demo – gridworld manual intro (L8D1)]



# Markovian Assumption

---

“Markov” generally means that given the present state, the future and the past are independent

For Markov decision processes, “Markov” means action outcomes depend only on the current state

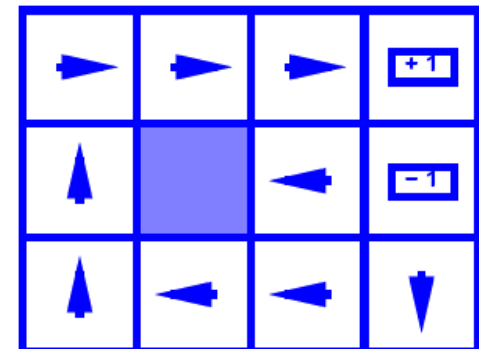
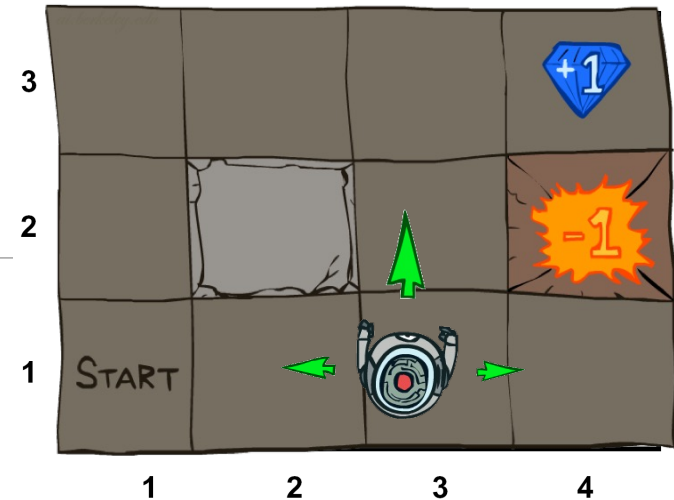
$$\begin{aligned} &P(S_{t+1} = s' | S_t = s_t, A_t = a_t, S_{t-1} = s_{t-1}, A_{t-1}, \dots, S_0 = s_0) \\ &= \\ &P(S_{t+1} = s' | S_t = s_t, A_t = a_t) \end{aligned}$$



Andrey Markov  
(1856-1922)

# Output: Policies

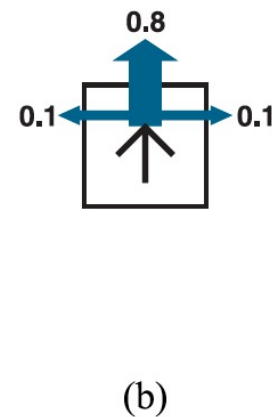
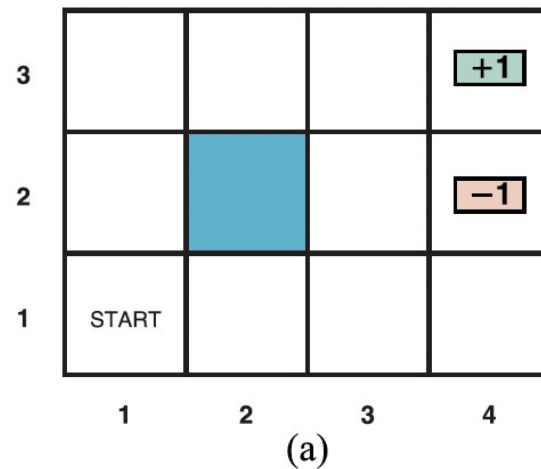
- In deterministic single-agent search problems, we have a **plan**, or sequence of actions, from start to a goal
- For MDPs, we want an optimal **policy**  $\pi^*: S \rightarrow A$ 
  - A policy  $\pi$  gives an action for each state
  - An optimal policy is one that maximizes expected utility if followed



$$R(s) = -0.01$$

Slide adapted from: Dan Klein and Pieter Abbeel's AI lecture

## Example 2:

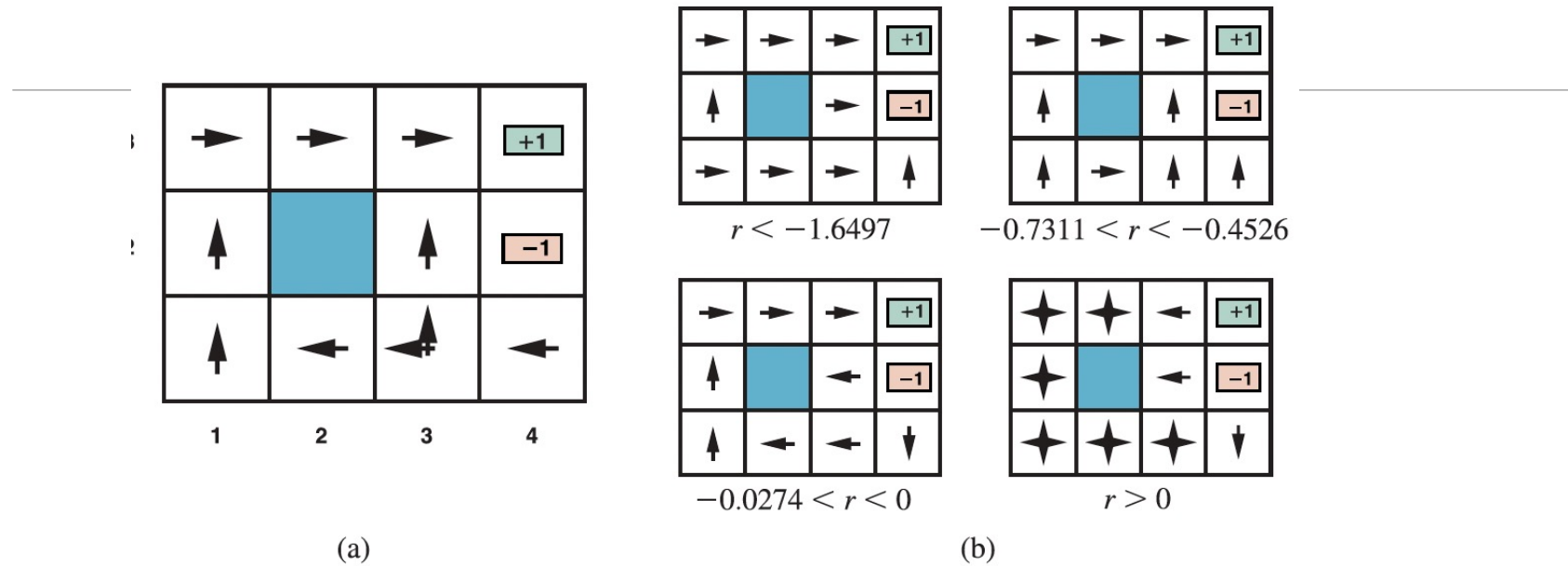


$r = -0.04$  for  
non-terminal states

**Figure 17.1** (a) A simple, stochastic  $4 \times 3$  environment that presents the agent with a sequential decision problem. (b) Illustration of the transition model of the environment: the “intended” outcome occurs with probability 0.8, but with probability 0.2 the agent moves at right angles to the intended direction. A collision with a wall results in no movement. Transitions into the two terminal states have reward +1 and -1, respectively, and all other transitions have a reward of -0.04.

Adapted from/ image credit:  
Russell & Norvig, AI: A Modern Approach

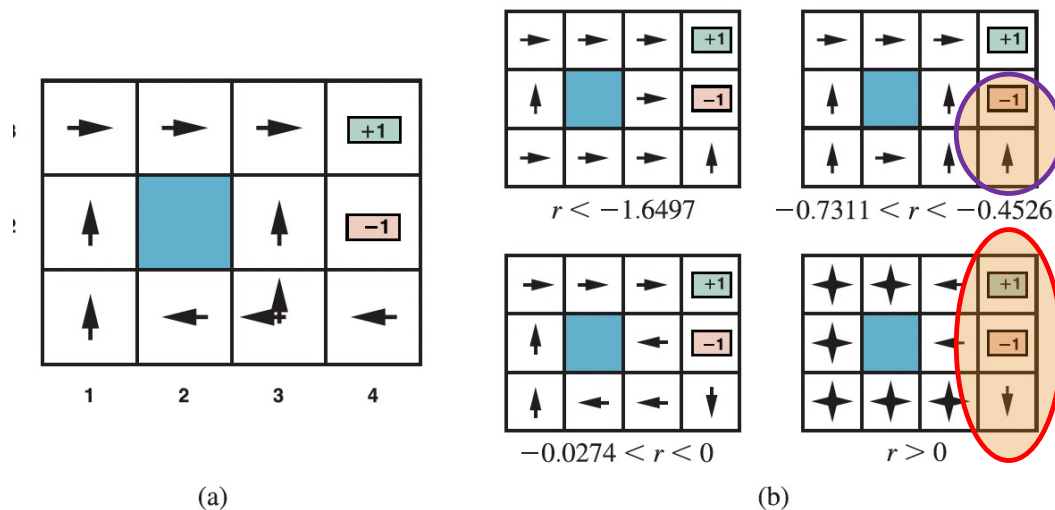
## Example 2: Optimal Policies Under Different Situations



**Figure 17.2** (a) The optimal policies for the stochastic environment with  $r = -0.04$  for transitions between nonterminal states. There are two policies because in state (3,1) both *Left* and *Up* are optimal. (b) Optimal policies for four different ranges of  $r$ .

Adapted from/ image credit:  
Russell & Norvig, AI: A Modern Approach

## Example 2:



Agent decides to directly go to -1 state rather than high cost of trying to go to +1

Agent decides NOT to go to any terminal state

**Figure 17.2** (a) The optimal policies for the stochastic environment with  $r = -0.04$  for transitions between nonterminal states. There are two policies because in state (3,1) both *Left* and *Up* are optimal. (b) Optimal policies for four different ranges of  $r$ .

Adapted from/ image credit:  
Russell & Norvig, AI: A Modern Approach

# On Finding Solution to MDP

---

- Dynamic programming: simplifying a problem by recursively breaking it into smaller pieces, solving it and assembling full solution from optimal solutions to sub-problems
- Optimal policy: a policy that yields the highest expected utility
- Setting: how much time we have – finite v/s infinite horizon
  - For **finite horizon**, solution may depend on time left. Policy is called **nonstationary**.
  - For **infinite horizon**, solution will not depend on time left. Policy is called **stationary**.
- Utility of a state sequence – by **additive discounted rewards**
  - $U_h ( [s_0, a_0; s_1, a_1; \dots] ) = R(s_0, a_0, s_1) + \gamma R(s_1, a_1, s_2) + \gamma^2 R(s_2, a_2, s_3) + \dots$

# On Finding Solution to MDP

---

- **Key Idea:** in an optimal policy, one would have chosen the action that maximizes the reward for the next step plus the expected discounted utility of the subsequent state
  - $\pi^*(s) = \operatorname{argmax}_a \sum_{s'} P(s' | s, a) [R(s, a, s) + \gamma U(s')]$
- **Key Idea:** The utility of a state is the expected reward for the next transition plus the discounted utility of the next state, assuming the agent chooses the optimal action
  - $U(s) = \max_a \sum_{s'} P(s' | s, a) [R(s, a, s) + \gamma U(s')]$
  - Bellman equation

# Finding Policy

---

- Value Iteration – iterate over value of states; offline; optimal
- Policy Iteration – iterate over policies ; offline; optimal
- Linear programming - offline; optimal
- Monte Carlo planning – online; approximate



# Exercise and Code

---

- MDP Solution Methods
  - From Book: AI – A Modern Approach,  
<https://github.com/aimacode/aima-python/blob/master/mdp.ipynb>
  - More applications  
[https://github.com/aimacode/aima-python/blob/master/mdp\\_apps.ipynb](https://github.com/aimacode/aima-python/blob/master/mdp_apps.ipynb)

Source: Russell & Norvig, AI: A Modern Approach

# Two Party Decisions - Games

---

- Games
  - Cooperative games
  - Non-cooperative games
    - Adversarial games
- What is value of cooperation ?
  - Prisoner's dilemma

# Two Party Decisions - Games

## Prisoner's dilemma

- Two prisoners are caught for a robbery. They can testify against each other (-5 years to other; 0 themselves), stay silent (-10 year if other testifies, but -1 if they do not).
- For A: testifying (defecting) is a better choice ( $-0 - 5 * \frac{1}{2} = -2.5$ ) over remaining silent (cooperating) ( $-1 - 10 * \frac{1}{2} = -6.5$ ) // Assuming B will decided with probability 0.5
- For B: similarly, testifying is better
- For both, cooperating is better: -1 each, but the authorities would try to prevent it

Prisoner A	Prisoner B	
	Prisoner B stays silent ( <i>cooperates</i> )	Prisoner B testifies ( <i>defects</i> )
Prisoner A stays silent ( <i>cooperates</i> )	Each serves 1 year	Prisoner A: 10 years Prisoner B: goes free
Prisoner A testifies ( <i>defects</i> )	Prisoner A: goes free Prisoner B: 10 years	Each serves 5 years

# Stable Marriage Problem

---

- The problem of finding a stable matching between two equally sized sets of elements given an ordering of preferences for each element. A matching is a bijection from the elements of one set to the elements of the other set. A matching is *not* stable if:
  1. There is an element  $A$  of the first matched set which prefers some given element  $B$  of the second matched set over the element to which  $A$  is already matched, and
  2.  $B$  also prefers  $A$  over the element to which  $B$  is already matched.
- Example Instances
  - Marriage: set 1 – men; set 2 – women
  - Jobs: Assignment of graduating medical students (set 1) to their first hospital appointments (set 2)
  - Servers: assigning users (set 1) to servers (set 2) in a large distributed Internet service

**Credit:** [https://en.wikipedia.org/wiki/Stable\\_marriage\\_problem](https://en.wikipedia.org/wiki/Stable_marriage_problem)

# Stable Marriage Problem - Solving

---

- Gale-Shapley Algorithm
  - for any equal number of men and women, it is always possible to solve the stable marriage problem and make all marriages stable.
  - Steps
    - In the first round, first a) each unengaged man proposes to the woman he prefers most, and then b) each woman replies "maybe" to her suitor she most prefers and "no" to all other suitors. She is then provisionally "engaged" to the suitor she most prefers so far, and that suitor is likewise provisionally engaged to her.
    - In each subsequent round, first a) each unengaged man proposes to the most-preferred woman to whom he has not yet proposed (regardless of whether the woman is already engaged), and then b) each woman replies "maybe" if she is currently not engaged or if she prefers this man over her current provisional partner (in this case, she rejects her current provisional partner who becomes unengaged). The provisional nature of engagements preserves the right of an already-engaged woman to "trade up" (and, in the process, to "jilt" her until-then partner).
    - This process is repeated until everyone is engaged.
  - Algorithm is guaranteed to produce a stable marriage for all participants in time  $O(n^2)$  where  $n$  is the number of men or women.
- Code example:
  - <https://github.com/biplav-s/course-tai/tree/573c1950381ed75eac1deaf93bf84de359f1f1b8/sample-code/future-material/stable-marriage-matching>

Credit: [https://en.wikipedia.org/wiki/Stable\\_marriage\\_problem](https://en.wikipedia.org/wiki/Stable_marriage_problem)

# Application of Decision Theory

---

- Help with individual decisions:
  - driving,
  - buying/ auctions, ...
- Help with group decisions:
  - hiring/ interviewing,
  - merger/ acquisition, ...
- Help with adversarial situations
  - Price discovery
  - Avoiding collusion
- Help with autonomous systems
  - Space crafts, drones, underwater navigation, ...

# Lecture 21: Summary

---

- We talked about
  - SDPs
  - MDPs
  - Prisoner's Dilemma
  - Stable Marriage Problem

# Week 11: Concluding Comments

## We talked about

- Lecture 18: Adversarial and Game Search
- Lecture 19: Optimization

- Week 1: Introduction, Aim: Chatbot / Intelligence Agent
- Weeks 2-3: Data: Formats, Representation and the Trust Problem
- Week 3: Machine Learning – Supervised (Classification)
- Week 4: Machine Learning - Unsupervised (Clustering) –
- Topic 5: Learning neural network, deep learning, Adversarial attacks
- Week 6: Large Language Models – Representation and Usage issues
- Weeks 7-8: Search, Heuristics - Decision Making
- Week 9: Constraints, Optimization – Decision Making
- Topic 10: Markov Decision Processes, Hidden Markov models - Decision making
- Topic 11-12: Planning, Reinforcement Learning – Sequential decision making
- Week 13: Trustworthy Decision Making: Explanation, AI testing
- Week 14: AI for Real World: Tools, Emerging Standards and Laws; Safe AI/ Chatbots



# Projects B: Sep 30 – Nov 20 (7 weeks; 400 points)

---

- End date: **Thursday, Nov 20**
  - Remember to update spreadsheet on data/ time when finished (**Column I**)
- Choices
  - Given by instructor
  - Defined by student using project-b teampate; reviewed and approved by instructor

# Upcoming Evaluation Milestones

---

- Projects B: Sep 30 – Nov 20
- Quiz 2: Oct 7
- Quiz 3: Nov 11
- Paper presentation (grad students only) : Nov 18
- Finals: Dec 11

# About Week 12 – Lectures 20, 21

---

# Week 12 – Lectures 22, 23

- Sequential Decision Making:
- Planning
- RL

- Week 1: Introduction, Aim: Chatbot / Intelligence Agent
- Weeks 2: Data: Formats, Representation, ML Basics
- Week 3: Machine Learning – Supervised (Classification)
- Week 4: Machine Learning - Unsupervised (Clustering) –
- Topic 5: Learning neural network, deep learning, Adversarial attacks
- Week 6: Large Language Models – Representation and Usage issues
- Weeks 7-8: Search, Heuristics - Decision Making
- Week 9: Constraints, Optimization – Decision Making
- Topic 10: Markov Decision Processes, Hidden Markov models -  
Decision making
- Topic 11-12: Planning, Reinforcement Learning – Sequential decision making
- Week 13: Trustworthy Decision Making: Explanation, AI testing
- Week 14: AI for Real World: Tools, Emerging Standards and Laws; Safe AI/ Chatbots

**Note:** exact schedule changes slightly to accommodate for exams and holidays.