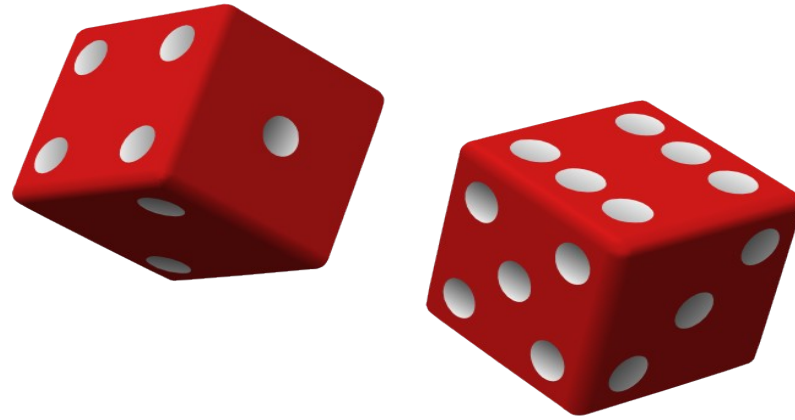


Artificial Intelligence

Lecture 1: Intro & Reasoning under uncertainty



CS4375 Probabilistic Artificial Intelligence Techniques
Dr. F. Oliehoek

Welcome!

■ Teachers

■ PhDs:

- ▷ Davide Mambelli
- ▷ Oussama Azizi
- ▷ Antonio Mone

■ TAs:

- ▷ Radek Kargul
- ▷ Aranya Sinha
- ▷ Aleksander Buszydlik

- See also “Course Information → Staff and Support” on brightspace



Frans Oliehoek



Matthijs Spaan



Luciano Siebert

Questions: 'answers EWI' forum

- Questions? We are using 'answers EWI'
 - ▷ <https://answers.ewi.tudelft.nl/>
 - ▷ (not brightspace forum)
 - ▷ use correct TAGs: "cs4375"
 - ▶ <https://answers.ewi.tudelft.nl/categories/1/tags/1024>
- How to use this?
 - ▷ <https://answers.ewi.tudelft.nl/help/faq>

Contact information: Brightspace
(Course Information → Staff and Support)
cs4375-support-ewi@tudelft.nl

Today

- Logistics...
- What is Artificial Intelligence (AI)?
 - ▷ And why do we care?
- Reasoning under uncertainty

Logistics

Course information...

- All information should be available at brightspace (under “contents → course information”)
 - ▷ course description, learning objectives, expected background, dates, etc.
 - ▷ let us know if you can't find something

Background...

- What are you studying...?

Background...

- What are you studying...?
- required background
 - ▷ probability
 - ▷ basic computer science, algorithms (e.g., search algorithms)
 - ▷ programming (python)

Schedule

(see brightspace for full schedule)

■ Lectures

- ▷ Tuesdays, 10:45-12:45
- ▷ Thursdays, 10:45-12:45
- ▷ On campus → different rooms... check schedule!
 - ▶ <https://mytimetable.tudelft.nl/>
- ▷ Lectures should be recorded on collegerama
 - ▶ <https://collegerama.nl/>
 - ▶ last year's lectures also available

Organizational (see Brightspace)

■ Reading

- ▷ Russell and Norvig (R&N) and additional material
- ▷ indicated on Brightspace: at the section for every week

Exercises

- Sheets with exercises will be made available
 - ▷ not graded
 - ▷ do them on your own
 - ▷ questions: forum

Practical assignment

- **Two** assignments
 - ▷ instructions will be published on Brightspace
- **implement** some AIT techniques
- **Group** work 4-5 people
- When? at your own time and place
- Graded, but **formative**: does not count towards your grade

- Form groups of 4 to 5 people
 - ▷ right now & here, or on forum
 - ▷ enroll your group on brightspace (under “collaboration”)

→ make sure you have ample programming experience in the team (python)

→ register your team

Exam

- In November 2022
 - ▷ no mid-term
 - ▷ both multiple choice and open questions
 - ▷ counts for 100%
 - ▷ practice exam will be made available
- Resit in Q2

Topics

- Approaching problems from
 - ▷ Single agent perspective
 - ▷ Multi-agent perspective
- We present a “decision-theoretic perspective”
 - ▷ probability for uncertainty
 - ▷ utility for goals
- Some keywords / techniques:
 - ▷ Bayesian networks and probabilistic inference
 - ▷ Rational decision making
 - ▷ Reinforcement learning and Markov Decision Problems
 - ▷ Game theory
 - ▷ Adversarial and cooperative decision making

Related courses

Well... what not? E.g.:

- Probabilistic programming
- Algorithms for Intelligent Decision Making
- Intelligent Decision Making Project
- Machine Learning 1 & 2
- Deep learning
- Deep Reinforcement Learning
- Conversational Agents
- Social Signal Processing
- Evolutionary Algorithms



Why AI?
What is AI?
What is an intelligent agent?

Why AI...?

■ Massive interest in AI...

ai


×



< All News Images Videos Books : More

Tools


About 485.000.000 results (0,31 seconds)


 Decrypt

Microsoft Infuses AI With Human-Like Reasoning Via an "Algorithm of Thoughts"

Tech giant Microsoft has unveiled a new AI training method called the "Algorithm of Thoughts" (AoT), designed to make large language models...

3 days ago





 The New York Times

The A.I. Revolution Is Coming. But Not as Fast as Some People Think.

Lori Beer, the global chief information officer of JPMorgan Chase, talks about the latest artificial intelligence with the enthusiasm of a...

3 days ago





 Trouw

Kun je met AI verloren schilderijen opnieuw creëren? 'Voeg schapen en bloesem toe'

Kan kunstmatige intelligentie originele beelden creëren? Misschien zelfs een kunstwerk? Een kunstredacteur en een mediarredacteur van...

15 hours ago





 NOS

VS breidt verkooprestricties Nvidia's AI-chips verder uit

Chipbedrijf Nvidia moet voortaan voor het verkopen van zijn modernste AI-chips aan bepaalde landen in het Midden-Oosten toestemming vragen...

18 hours ago




 Telegraaf

Samsung brengt AI naar je bord en gaat helpen met afvallen

Eten is misschien niet het eerste waar je aan denkt bij het toepassen van kunstmatige intelligentie (AI).


2 hours ago




Recently: Generative AI



- dream.ai's interpretation of "lecture AI techniques TU Delft"

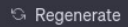


Hi ChatGPT, I will be giving the first lecture "AI techniques" of the semester next week. What should I tell my new students about you?



Hello! It's great that you'll be introducing your students to AI techniques. When explaining me, ChatGPT, to your new students, you can cover the following key points:

- 1. Introduction to ChatGPT:**
 - Start by explaining that ChatGPT is an artificial intelligence language model developed by OpenAI.
 - Mention that it is part of the GPT-3.5 architecture, a state-of-the-art model for natural language understanding and generation.
- 2. What ChatGPT Can Do:**
 - Emphasize that ChatGPT can understand and generate human-like text in response to various prompts.
 - Mention its ability to provide information, answer questions, generate creative content, and more.
- 3. How ChatGPT Works:**
 - Give a simplified overview of how ChatGPT is trained on a vast dataset of text from the internet, allowing it to learn patterns and relationships in language.
 - Explain that it uses deep learning techniques, specifically transformer architectures, to process and generate text.
- 4. Applications:**
 - Discuss the practical applications of ChatGPT, such as chatbots, content generation, language translation, and text-based tasks in various industries.
- 5. Limitations:**
 - Highlight that while ChatGPT is powerful, it has limitations, such as generating biased or inappropriate content and occasionally producing inaccurate information.
 - Mention that it may not understand context perfectly and can generate plausible



Recently: Generative AI



that is great, but can you make it a bit shorter. I just want to quickly mention you on 1 slide.

it



Certainly! Here's a shorter version for a single slide introduction:



Slide Title: Introduction to ChatGPT

- **ChatGPT:** An advanced AI language model created by OpenAI.
- **Capabilities:** Understands and generates human-like text, useful for chatbots, content generation, translation, and more.
- **How It Works:** Learns language patterns through deep learning (transformer architecture).
- **Limitations:** Possible biases, occasional inaccuracies, and lack of context understanding.
- **Ethical Use:** Emphasize responsible and ethical AI use.
- **Future of AI:** Discuss AI's growing role in various industries.
- **Resources:** Encourage further exploration.
- **Engagement:** Promote questions and discussions.

This concise slide can serve as a quick introduction to ChatGPT for your students.

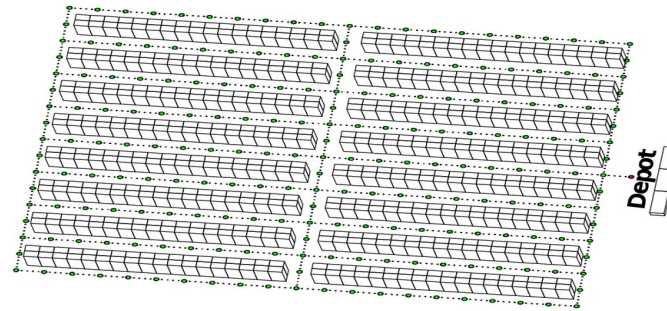
or

generate

How can we...

- ...run an autonomous warehouse?

Warehouse Commissioning

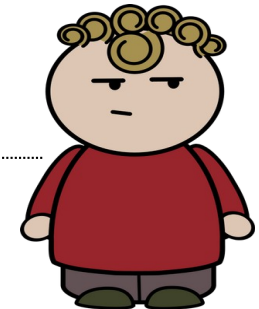


Toru Robot
(Magazino GmbH)

- ...interact with customers on-line?

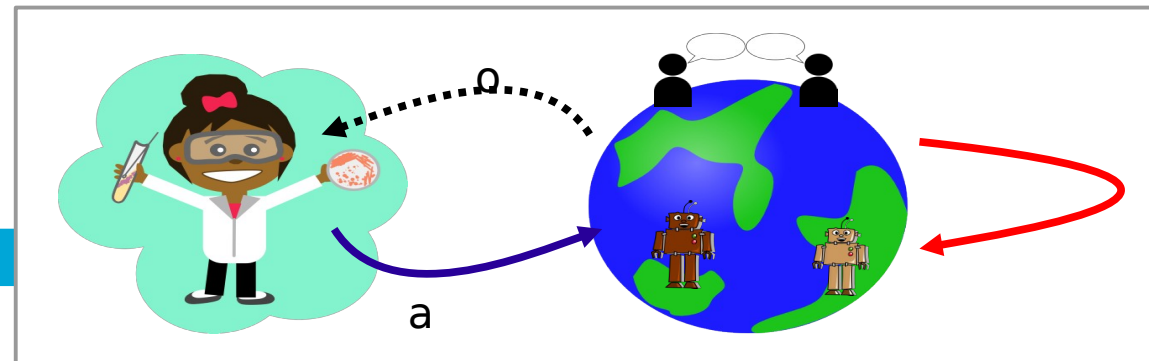
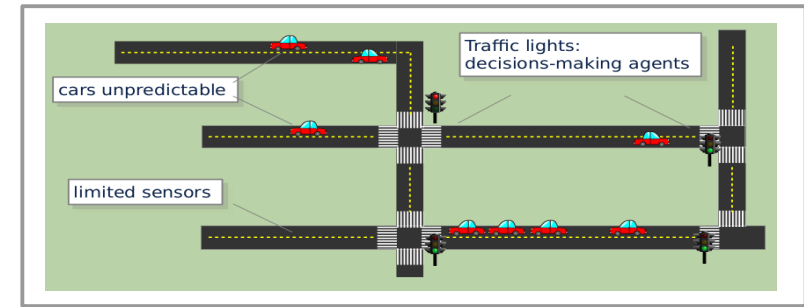


Our Corp



How can we...

- ...beat Go Grandmasters?
- ...coordinate traffic lights in a large city?
- ...develop an artificial scientist?



What is Artificial Intelligence?

Many definitions have been proposed for Artificial Intelligence

Df 1: Machine \geq human

- The science that tries to automate processes that humans so far do better than machines

Df 2: Understand and simulate

- The science that aims to understand natural intelligence so well that it can be simulated on a computer.

Df 3: Human and computer

- The science of developing intelligent software that supports humans.

One step back: what is Intelligence?

What do you think? Suggestions...?

- The AllWords Dictionary: 'The ability to use **memory**, **knowledge**, **experience**, **understanding**, **reasoning**, **imagination** and **judgement** in order to solve problems and adapt to new situations'
- 'The capacity to **learn** or to **profit by experience**' W. F. Dearborn
- 'Ability to **adapt** oneself adequately to relatively **new situations** in life' R. Pinter
- 'A very general **mental capability** that, among other things, involves the ability to **reason**, **plan**, **solve problems**, **think abstractly**, **comprehend complex ideas**, **learn quickly** and **learn from experience**' (Gottfredson, 1997, as cited by Schut, 2007)

'Viewed narrowly, there seem to be almost as many definitions of intelligence as there were experts asked to define it' R. J. Sternberg

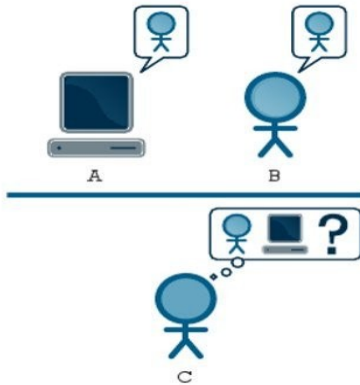
Intelligence definitions from AI researchers

- “Intelligence is the **computational part of the ability to achieve goals** in the world. Varying kinds and degrees of intelligence occur in people, many animals and some machines.” J. McCarthy
- “Any system . . . that generates **adaptive behaviour to meet goals in a range of environments** can be said to be intelligent.” D. Fogel
- “. . . in any real situation **behavior appropriate to the ends** of the system and adaptive to the demands of the environment can occur, within **some limits of speed and complexity**.” A. Newell and H. A. Simon

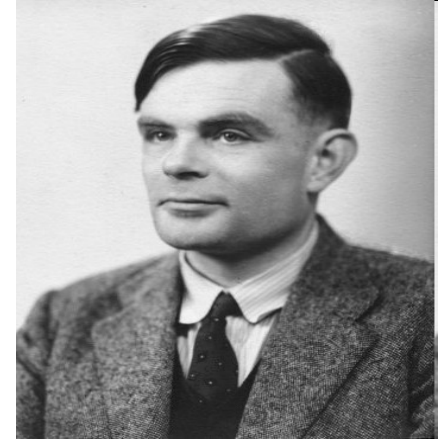
Can machines think?

A practical approach

- Turing, 1950: that's an ill-posed problem: *'the meaning and the answer to the question, "Can machines think?" is to be sought in a statistical survey such as a Gallup poll'*
- Let's rather frame it in terms of an **imitation game**:



A computer or a human is interrogated by a participant through a teletype (chat). The interrogator guesses if there is a computer or a human at the other end of the terminal.



Alan Turing
1912-1954

Turing thought that by 2000 computers would have fooled humans 30% of the times.

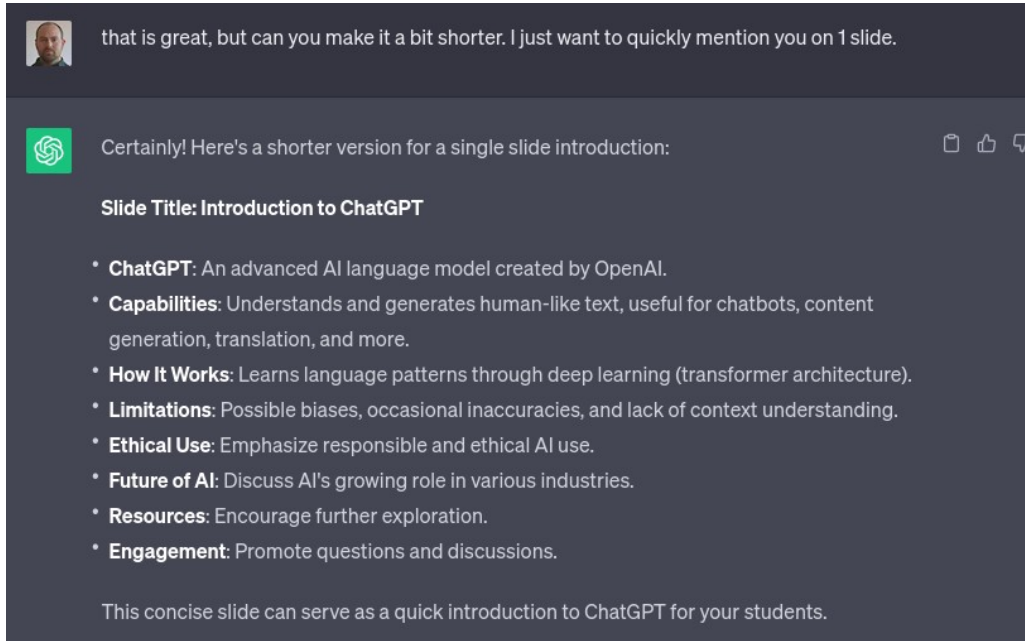
The Turing test

To pass it, a machine needs to be able to

- process natural language (recognize and interpret it)
- Represent knowledge (to store and access info)
- Reason (to draw conclusions)
- Learn (to adapt to new circumstances by examining its previous experiences).

Where are we?

- Well, what do you think...?



Progress is rapid: “Where are we? (2022)”

- Much recent progress in language models
- but real reasoning power seems limited.

the field of AI is best described by a maxim:

“if something can be done,

it will be done.”

If you set your hope to greatness and hope to believe that humanity has advanced far beyond the technology used to enslave it, you will be a

Generate Text

talktotransformer.com

The Chinese Room (Searle 1980)

- Intelligent behavior does not imply the machine is **conscious** and **understanding** in a human sense
- Perhaps these are emergent properties of biological brains!
- However, machines can be intelligent



© Dan Dennett and Neil Cohn

www.visuallanguagelab.com

<https://www.visuallanguagelab.com/chinese-room>

From a philosophical perspective

Weak AI hypothesis: Machines may act intelligently, without:

- actually thinking
- having a mind, i.e., mental states of self-awareness

Strong AI hypothesis: Machines that act intelligently necessarily:

- are thinking
- possess consciousness.

Do we need strong AI? (i.e. is it important that machines actually think ?) Is that achievable?

What sort of agent system?

What kind of agent system is AI trying to engineer?



Biology
Psychology
Sociology
Cognitive Sc

Think Like Humans	Think Rationally
Act Like Humans	Act Rationally

Mathematics
Economy
Logic
Game Theory
Philosophy

Acting humanly: Turing test



- Turing (1950) “Computing machinery and intelligence”
 - ▷ “Can machines think?”
- Operational test for intelligent behavior: Imitation Game
- Suggested major components of AI: knowledge, reasoning, language understanding, learning.
- Pro: clear yes/no answer
- Cons:
 - ▷ Not **reproducible**, **constructive**, or amenable to **mathematical analysis**.
 - ▷ It does not provide any guidelines for building agents!

Acting humanly: Turing test



- Turing (1950) “Computing machinery and intelligence”
 - ▷ “Can machines think?”
- Operational test for intelligent behavior: Imitation Game
- Suggested major components of AI: knowledge, reasoning, language understanding, learning.
- Pro: clear yes/no answer
- Cons:
 - ▷ Not **reproducible**, **constructive**, or amenable to **mathematical analysis**.
 - ▷ It does not provide any guidelines for building agents! ← ML techniques?

Thinking humanly: Cognitive science

- What is thinking humanly?
 - ▷ Requires scientific theories of internal activities of the brain...
 - ▷ Two scientific approaches:
 - ▶ Predicting and testing behavior of human subjects (top-down)
 - ▶ Direct identification from neurological data (bottom-up)
- Both approaches (roughly cognitive science and cognitive neuroscience) are now distinct from AI
- Pro:
 - ▷ CogSci provides potentially useful theories for AI engineers!
- Con:
 - ▷ human specifics such as memory limitations are irrelevant for machine intelligence

Thinking rationally: Laws of Thought

- Goes back to Aristotle: what are correct arguments / thoughts?
 - ▷ Socrates is a man; all men are mortal; therefore ...
- Use logic to derive the 'right' conclusions
- Pros:
 - ▷ Theoretically well-founded.
- Cons:
 - ▷ Difficult to capture the worlds in crisp logical statements
 - ▷ Perhaps not all intelligent behavior is mediated by logical deliberation.

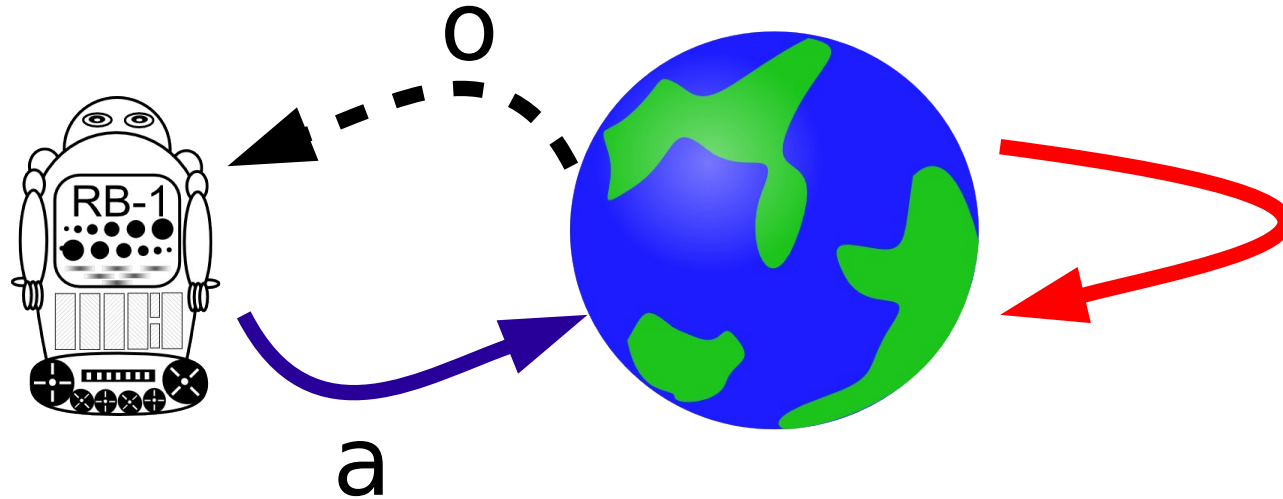
Acting rationally: the rational agent

- Rational behavior: doing “the right thing”...
 - ▷ expected to **maximize goal achievement** or a different notion of **performance**, given the available information.
- Pros:
 - ▷ Also theoretically well-founded
 - ▷ More general than “law of thought”
(correct reasoning is just one way of getting correct actions)
 - ▷ mostly quantitative: well-suited for scientific approach
- Cons:
 - ▷ Quantitative... → can be hard to obtain the “numbers”, or to relate the theory to **human** performance.

Definition of Agent

According to Russel&Norvig (ch.2)

*“An **agent** is anything that can be viewed as perceiving its **environment** through **sensors** and acting upon that environment through **actuators**”*



What is uncertainty and why AI needs to care?

Why care about uncertainty?

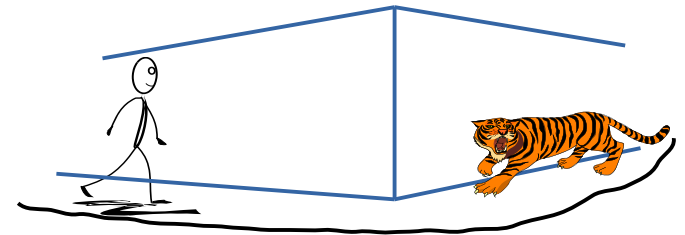
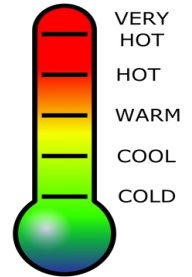
■ Stochasticity:

- ▷ outcome of actions uncertain
- ▷ environment can fluctuate



■ State uncertainty:

- ▷ sensor noise
- ▷ limited sensor



Logic for dealing with uncertainty

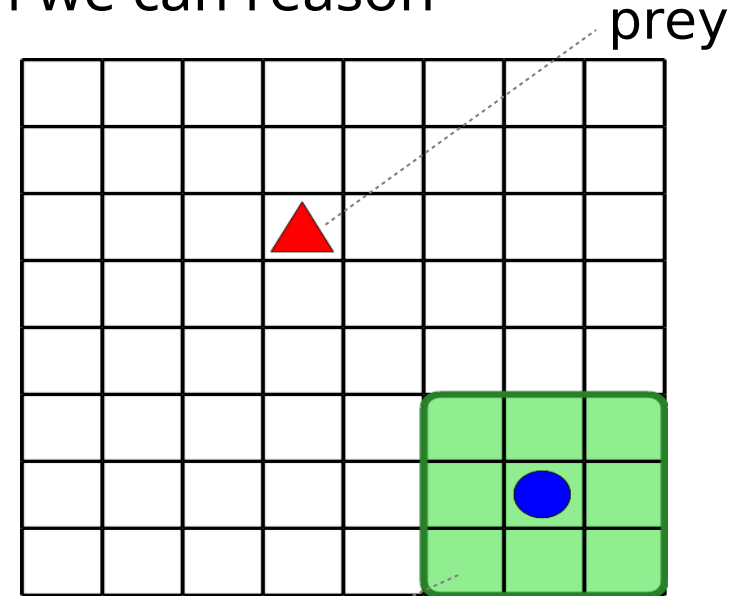
(E.g., R&N chap 7)

- Represent world using facts over which we can reason

- ▷ e.g.: predator-prey
- ▷ state= $(-3,4)$

- Maintain a **belief**:

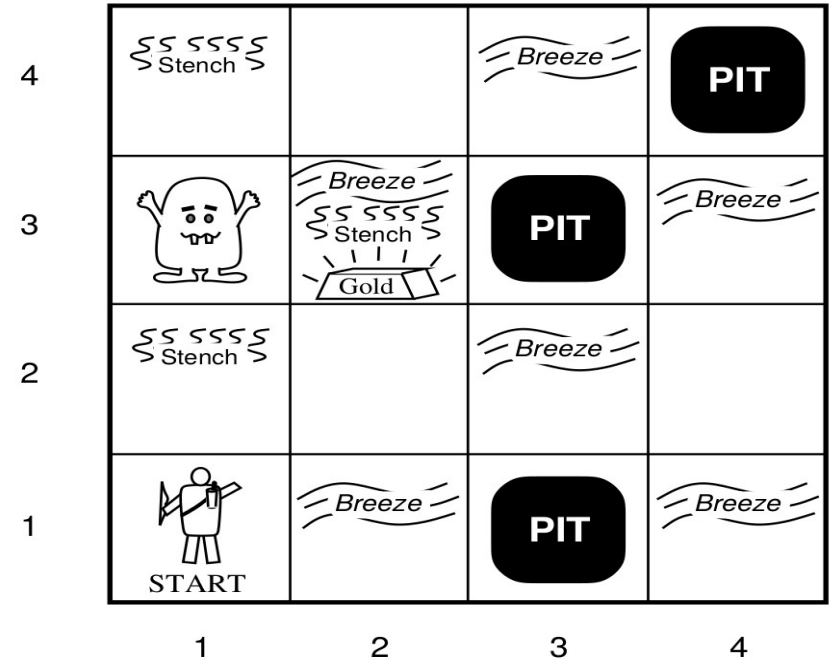
- ▷ set of possible states!
- ▷ $b=\{(-6,-1),(-6,0),(-6,1)\dots\}$



predator's observation range

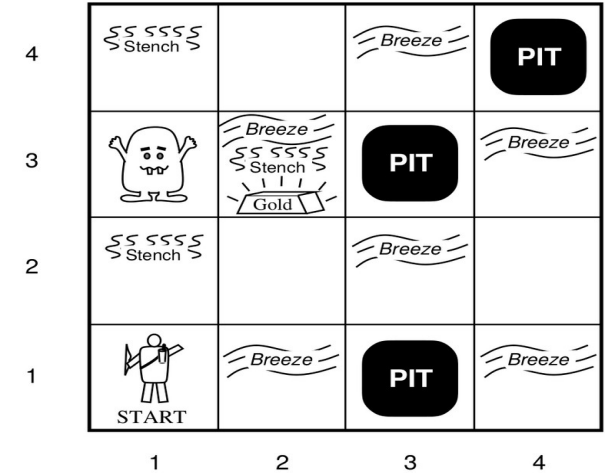
Scaling this...

- Works for simple problems
 - ▷ static prey
 - ▷ simple predictable movements
 - ▷ “Wumpus world” (R&N, sect. 7.2)
- Planning to get to the airport in to make our flight...
 - ▷ a_t = “leave in t minutes”
 - ▷ logical conclusion:
 a_{25} will get me on time if
 - ▶ there is no traffic jam, and
 - ▶ the car does not break down, and
 - ▶ do not need to get gas, and
 - ▶ I don’t get in an accident, and
 - ▶ ...



Scaling this...

- Works for simple problems
 - ▷ static prey
 - ▷ simple predictable movements
 - ▷ “Wumpus world” (R&N, sect. 7.2)
- Planning to get to the airport in to make our flight...
 - ▷ a_t = “leave in t minutes”
 - ▷ logical conclusion:
 a_{25} will get me on time if
 - ▶ there is no traffic jam, and
 - ▶ the car does not break down, and
 - ▶ do not need to get gas, and
 - ▶ I don’t get in an accident, and
 - ▶ ...



Too many things can go wrong...!

“qualification problem”
need to specify all exceptions of
desired effects of actions.

Different approach

- Not all these possible outcomes are likely...
- Base decision on **degree of belief** $b(s' | a)$
 - ▷ in outcomes $s' - \{on_time, too_late\}$
 - ▷ given alternative actions $a - \{a_{10}, a_{25}, a_{40}, \dots\}$

Different approach

- Not all these possible outcomes are likely...
- Base decision on **degree of belief** $b(s' | a)$
 - ▷ in outcomes $s' - \{on_time, too_late\}$
 - ▷ given alternative actions $a - \{a_{10}, a_{25}, a_{40}, \dots\}$

think of s'
as resulting state

Different approach

- Not all these possible outcomes are likely...
- Base decision on **degree of belief** $b(s' | a)$
 - ▷ in outcomes $s' - \{on_time, too_late\}$
 - ▷ given alternative actions $a - \{a_{10}, a_{25}, a_{40}, \dots\}$
- Select action with highest 'believed' **utility**:

$$U(a) = \sum_{s'} u(s') * b(s' | a)$$

Different approach

- Not all these possible outcomes are likely...
- Base decision on **degree of belief** $b(s' | a)$
 - ▷ in outcomes $s' - \{on_time, too_late\}$
 - ▷ given alternative actions $a - \{a_{10}, a_{25}, a_{40}, \dots\}$
- Select action with highest 'believed' **utility**:

$$U(a) = \sum_{s'} u(s') * b(s' | a)$$

- how does this help?
- how do we form these beliefs?

Degrees of belief

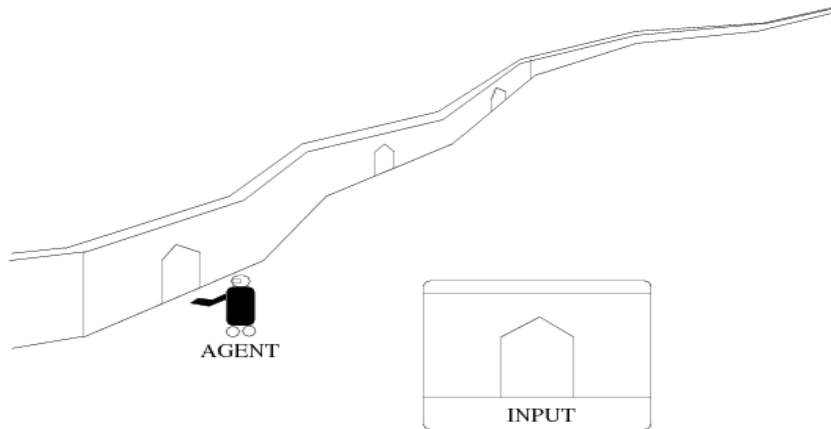
- Many people advocate: probability
- Leads to **maximum expected utility**

$$U(a) = \sum_{s'} u(s') * P(s' | a)$$

- also “decision theory” (=probability + utility)

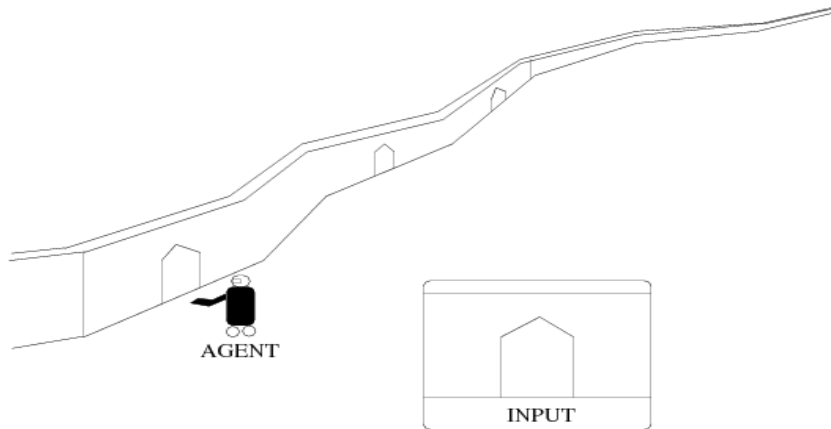
State Uncertainty

- Traffic example: **outcome uncertainty**
- How about: **state uncertainty**?



State Uncertainty

- Traffic example: **outcome uncertainty**
- How about: **state uncertainty?**



Similar...!

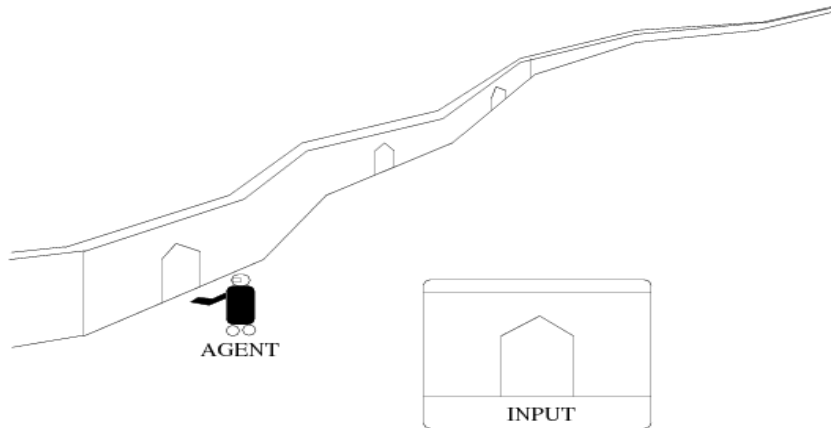
- now state s is hidden
- utility of action depends on s

$$U(a) = \sum_s u(s, a) * b(s)$$

State Uncertainty

- Traffic example: **outcome uncertainty**
- How about: **state uncertainty**?

think of s as current
(but hidden) state



Similar...!

- now state s is hidden
- utility of action depends on s

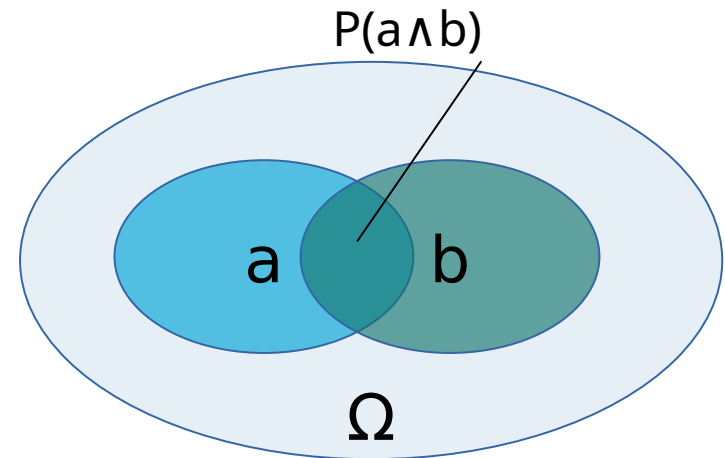
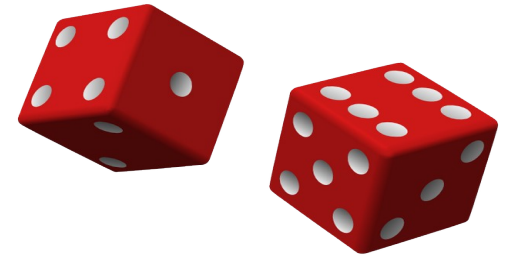
$$U(a) = \sum_s u(s, a) * b(s)$$

What you should know about probability...

Short answer:

All of Russel&Norvig (v4) Chap 12

- events, random variables
- joint probability
- inference by enumeration
- independence
- conditional independence
- Bayes Rule



Again: why probability?

De Finetti's argument

- 1) (Non-negativity) $P(A) \geq 0$, for all $A \in \mathcal{F}$.
- 2) (Normalization) $P(\Omega) = 1$.
- 3) (Finite additivity) $P(A \vee B) = P(A) + P(B)$ for all $A, B \in \mathcal{F}$ such that $A \cap B = \emptyset$.

Bruno de Finetti:

If agent's beliefs violate the axioms of probability, then there exists a combination of bets against it which it is willing to accept that guarantees it will lose money, every time.

Example

A Dutch book:

prop.	belief	taken bet	a, b	$a, \neg b$	$\neg a, b$	$\neg a, \neg b$
a	0.4	6:4 against a	-6	-6	4	4
b	0.3	7:3 against b	-7	3	-7	3
$a \vee b$	0.8	8:2 on $a \vee b$	2	2	2	-8
			-11	-1	-1	-1

Example

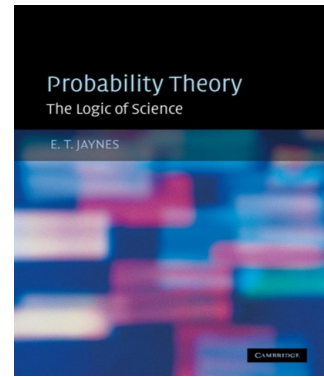
$$\begin{aligned} u(6:4 \text{ against } a) &= -6 * \text{bel}(a) + 4 * \text{bel}(\neg a) \\ &= -6 * 0.4 + 4 * 0.6 = 0 \end{aligned}$$

A Dutch book:

prop.	belief	taken bet	a,b	a, \neg b	\neg a,b	\neg a, \neg b
a	0.4	6:4 against a	-6	-6	4	4
b	0.3	7:3 against b	-7	3	-7	3
a v b	0.8	8:2 on a v b	2	2	2	-8
			-11	-1	-1	-1

Cox' Theorem

(See Jaynes 2003)



- Desiderata:
 - ▷ Degrees of plausibility: represented by **real numbers**
 - ▷ **Qualitative** correspondence how humans reason
 - ▷ **Consistency**: If a conclusion can be reached in more ways, then every possible way must lead to the same result
- Need to use probability to represent plausibility
- “Probability theory is nothing but common sense reduced to calculation.” — Laplace, 1819

Bayes rule

“One rule to rule them all”

If you are going to remember just one thing...

- ...remember Bayes' rule:

$$P(A | B) = P(B | A)P(A) / P(B)$$

directly from product rule:

$$\Leftrightarrow P(A | B) P(B) = P(B | A)P(A)$$

$$\Leftrightarrow P(A, B) = P(B, A)$$

Generalized form given background evidence \mathbf{e} :

$$P(A | B, \mathbf{e}) = P(B | A, \mathbf{e})P(A | \mathbf{e}) / P(B | \mathbf{e})$$

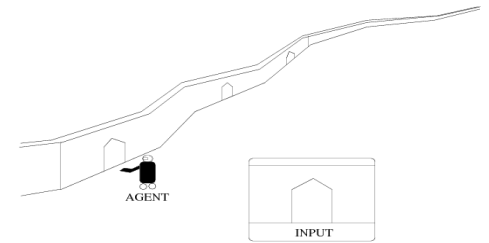
- Important why..?
(What is its importance for an intelligent agent?)

It allows to update a belief

- Rewrite with State and a particular observation o :

$$\mathbf{P}(\text{State} \mid o) = \mathbf{P}(o \mid \text{State})\mathbf{P}(\text{State}) / P(o)$$

- $\mathbf{P}(\text{State})$ is our prior belief
- so we can update our belief, based on observations!



Many observations...

- How to deal with **many observations**?

$$P(\text{State} | o_1, o_2, o_3, \dots) = \alpha P(o_1, o_2, o_3, \dots | \text{State}) P(\text{State})$$

- we don't know the observation sequence in advance...
- representing $P(o_1, o_2, o_3, \dots | \text{State})$ with a table does not scale...

Many observations...

- How to deal with **many observations**?

$$P(\text{State} | o_1, o_2, o_3, \dots) = \alpha P(o_1, o_2, o_3, \dots | \text{State}) P(\text{State})$$

- we don't know the observation sequence in advance...
 - representing $P(o_1, o_2, o_3, \dots | \text{State})$ with a table does not scale...
- Solution:
 - conditional independence can help significantly

It allows to *maintain* a belief

- Given conditional independent observations
 $P(o_1, o_2 | State) = P(o_1 | State)P(o_2 | State)$

- ...we can also sequentially update:

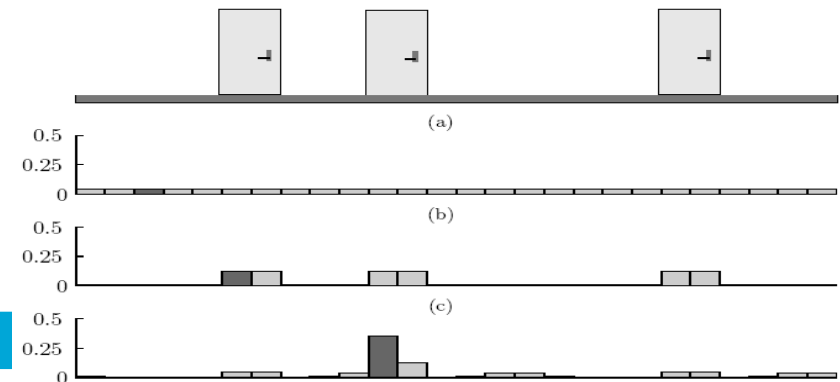
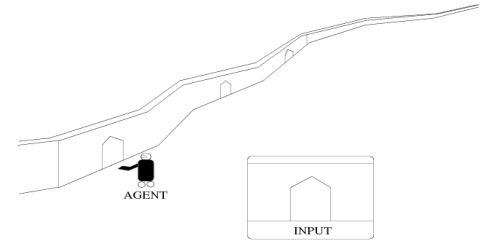
$$P'(State) := P(o_1 | State)P(State) / P(o_1)$$

$$P''(State) := P(o_2 | State)P'(State) / P(o_2)$$

o then $P''(State) = P(State | o_1, o_2)$

o (Exercise!)

- We will see later how
to incorporate robot movement over time



AI and Uncertainty: Summary

- 'Rational agent' approach to AI
 - ▷ agent aims to do the 'right thing' given its knowledge
 - ▷ and that task can be quantified
- Agents need to represent beliefs...
 - ▷ Deterministic rules and logic struggle to represent the world: too many exceptions... → qualification problem
- Probability
 - ▷ Can summarize many different failure cases
 - ▷ strong arguments: use probability
 - ▷ Bayes rule: to update beliefs