EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Introduction

Welcome everyone!

- What: *The Language of Thought: computational cognitive science approaches to category learning*
- Who: Fausto Carcassi
- When: Sommer semester 2022

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Who are we?

- Let's go around, and please say:
  - Your name
  - What you are studying?
  - Why you are taking this course?
  - What you are expecting from this course?
  - Do you know any python / formal grammars / Bayesian probability?

Various matters of organizational nature

- We'll have two sessions every week
  - Monday: a lecture with some theory
  - Wednesday: a programming lab
- If there is interest, I might also do a session on simple maths stuff for people without the background & offer office hours.
  - Let me know if you'd like those additional sessions! I'll explain how in a sec.
- There is a website for this course: https://thelogicalgrammar.github.io/pLoT_course
- The website contains
  - The course materials (more on this in this week's lab!)
  - Various info on the course.
  - The lecture slides.
  - And basically everything else!
- The website is not done yet, I will update it during the semester as we go.

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Concerning questions

- There's a technical part to this course. It's important that I don't lose you on in.
- Otherwise the course will stop making sense.
- We absolutely need a way for you to ask questions when you're confused.
- One way to ask questions is during class!
- But I know that some people feel shy about asking questions in class.
- Therefore, I added an **anonymous feedback form** on the website.
- 'Ask a question!' section in the sidebar
- So please let me know if you have any questions, either during class or through the website form.
- I might not be able to answer them on the spot but I'll write them down and let you know next time!

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Course structure

- The course will be structured in two parts.
- First, a foundational section lasting 7 weeks (including today) and covering:
  - Philosophical foundations of the Language of Thought & introductory python.
    - 3 weeks (including today)
  - Formal grammars / formal languages / some lambda calculus
    - 2 weeks
  - Bayesian foundations
    - 2 weeks
  - It's really important you keep up with this material or the second part of the course will make no sense!
- Second, a section on the probabilistic Language of Thought
  - One introductory week going over the main ideas & Steven Piantadosi's LOTlib3 library
  - Several weeks looking at specific applications
    - E.g. learning logic, kinship words, numerals, etc.
- A final review week

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Evaluation

- We'll do two types of evaluation (described on the website too)
- First, homework sets (30% of total grade)
  - A total of 3 (10% each)
  - To be done individually
  - They will be about technical stuff in the course
  - They are meant to motivate you to keep up with the technical side of things
  - They will concern the three technical bits in the first part: Python, formal grammars, and Bayesian probability.
- Second, a final project (70% of total grade)
  - Can be done in groups
  - You can find some possible topics on the website

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Concerning confusion

- Some of you may already have taken courses with a technical component.
- If you're starting with no technical background on python/grammars/probability, it's going to be a difficult course.
  - Difficult but doable!
  - And this material is useful for all sorts of things, so good to learn anyway.
- But this means that there will be moments of confusion, where you have the feeling you're losing grip on the material.
- Remember: things are not meant to be clear the first time you hear about them. If you stay through the confusion and keep applying yourself, things will get clearer and clearer.

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Concerning slides

- I will put quite a lot of text on the slides, and follow them quite closely during lectures.
- I know this is not ideal and makes lectures less engaging.
- I want to explain the reason for this choice: since we are not going to follow a textbook, ideally you should be able to reconstruct the thread of what was said in the lectures from the slides.
- Nonetheless, please feel free to stop me whenever you are feeling confused. I hope that we can make some space for discussions of things as we go.

Questions

- ‘Where do I find the reading for each week?’
  - On the website, in the section for the corresponding week
  - Note: I will add the weeks as we go, also depending on your feedback. Now the website is mostly empty.
- ‘Can I do the final project by myself?’
  - Yes, that is fine by me, but discuss your choice of project with me.

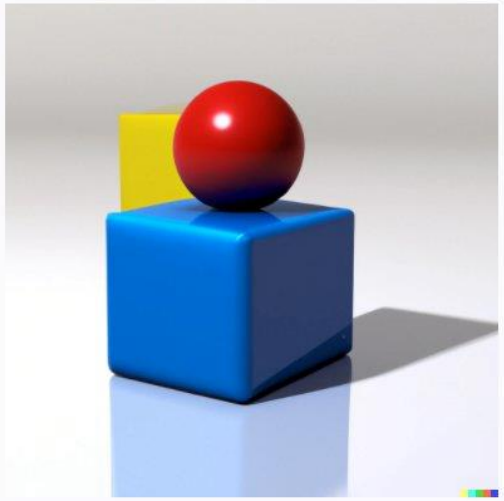
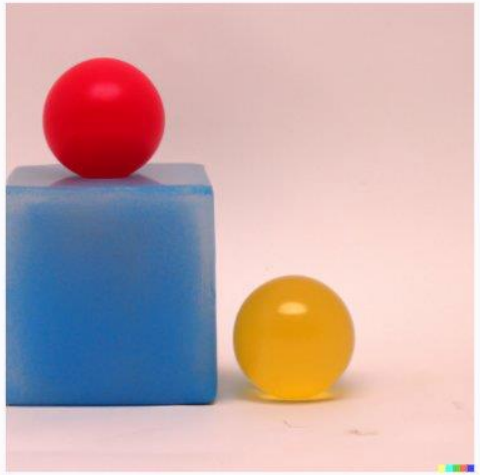
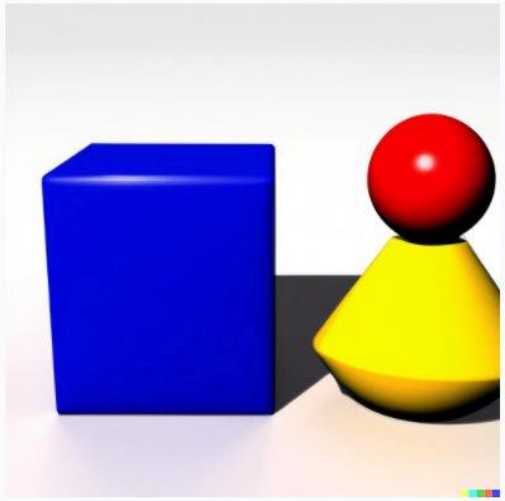
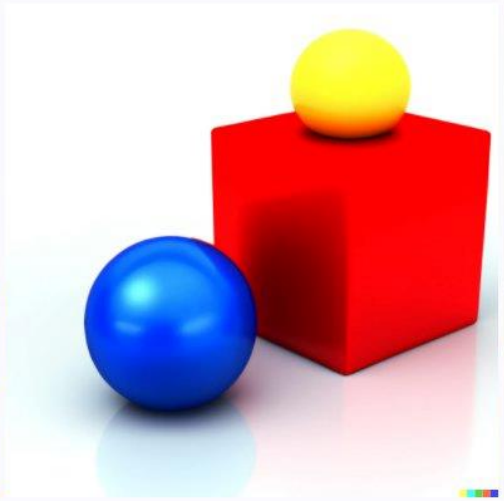
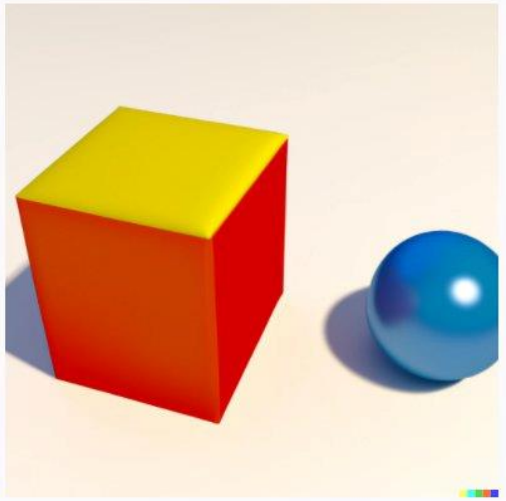
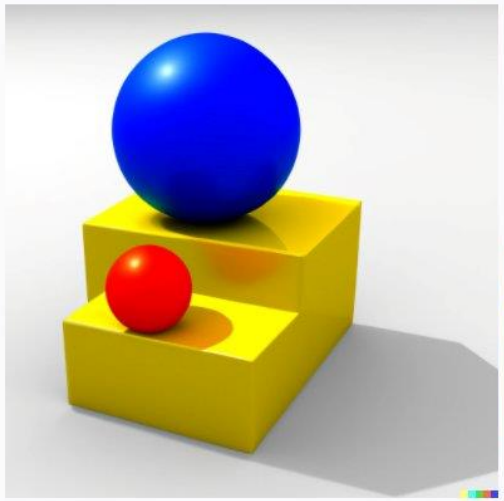
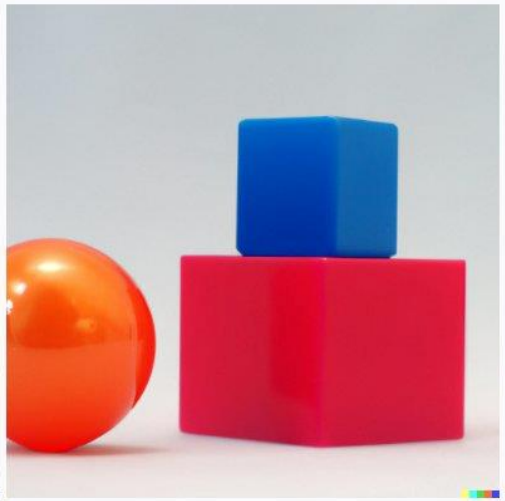
- Other questions?

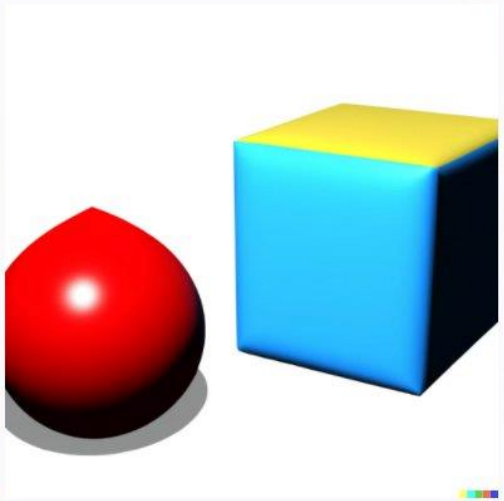
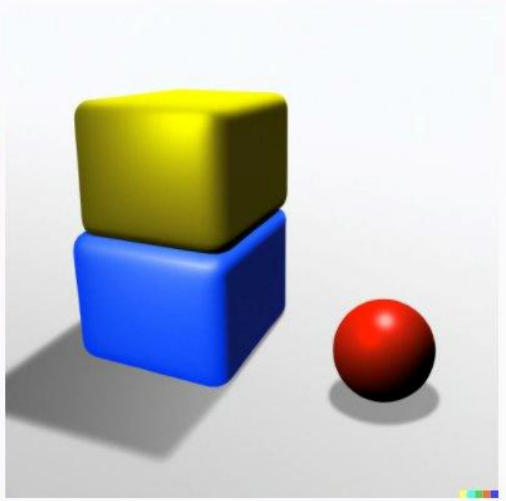
EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

The very idea of a Language of Thought

a blue cube on top of a red cube, beside a smaller yellow sphere

Report issue

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Learning a rule from (few!) examples

Robert Feldman → Dr Feldman

Ruth Millican → Dr Millikan

Joanna Newsom → ??

- Dj Newsom

Is the rule "Dr+ +*last name*" or "D+*first letter of first name*+ + *last name*"?

6 @ 2 = 12

3 @ 4 = 12

10 @ 2 = ??

- 12

Does @ express multiplication or does it simply say to return 12?

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Learning a rule from (few!) examples

This is our paradox: *no course of action could be determined by a rule, because every course of action can be made out to accord with the rule*. The answer is: if everything can be made out to accord with the rule, then it can also be made out to conflict with it.

Ludwig Wittgenstein, Philosophical Investigations, §201

- Pretty unclear what this means but it sounds great and it's something to think about.

Albeit its prosperity, current AI techniques cannot rapidly generalize from a few examples. […] successful AI applications rely on learning from large-scale data. In contrast, humans are capable of learning new tasks rapidly by utilizing what they learned in the past. For example, a child who learned how to add can rapidly transfer his knowledge to learn multiplication given a few examples (e.g., $2 \times 3 = 2 + 2 + 2$ and $1 \times 3 = 1 + 1 + 1$). Another example is that given a few photos of a stranger, a child can easily identify the same person from a large number of photos.

Wang et al (2020), "Generalizing from a Few Examples: A Survey on Few-Shot Learning"

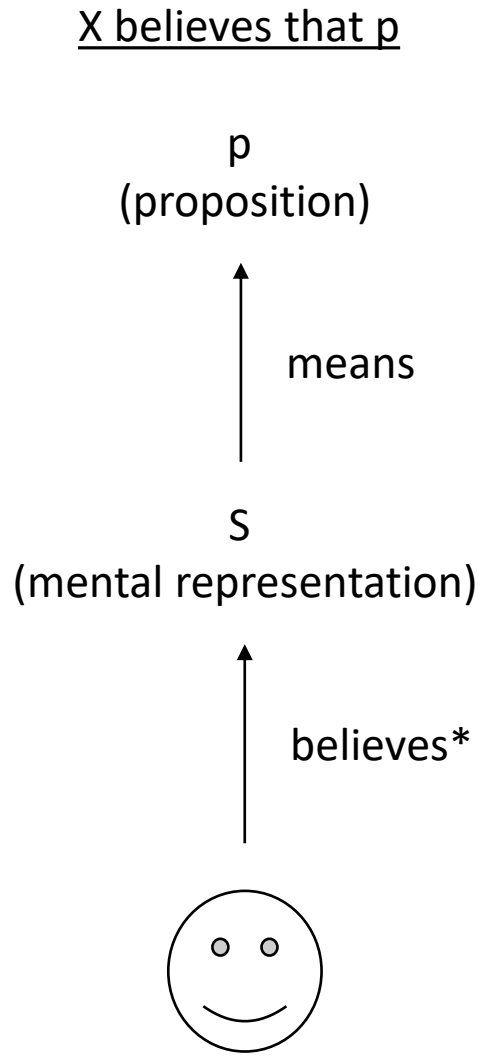
EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Learning a rule from (few!) examples

- Basically, there's something deeply puzzling about how we manage to learn rules from just a few examples.
- And yet, in many cases we agree on what the most plausible rule is.
- This is a puzzle for cognitive science, involving philosophy, linguistics, psychology, and computer science.
- In this course, we will analyze in detail one approach to this question, namely the approach of the Language of Thought.
- But note that by answering the question of how learning works, we will approach some fundamental questions about the nature of the mind, e.g.:
  - What does it mean to think / reason / learn?
  - What is a mental representation?
  - How is learning in humans related to complexity?
- But our starting point is a much more modest question…

Propositional attitudes

- Starting question. What does it mean to say:
  - Mary thinks that it will rain tomorrow
- One possible answer (from the SEP entry):
  - *"X* believes that *p"* is true if and only if:
    - there is a mental representation *S* such that
    - *X* believes* *S*
    - *S* means that *p*
- Where :
  - believes* is some relation between individuals and mental representations that we haven't defined yet
  - We still haven't defined what 'means' means
  - We don't know what a proposition is
- This doesn't feel like much progress, but bear with me!



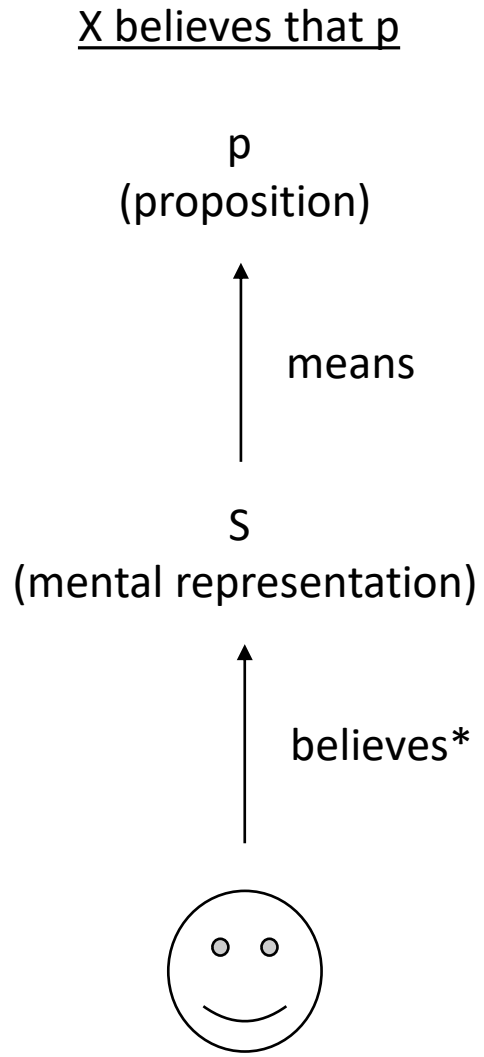
Propositional attitudes

The point here is that this analysis divides the work into more manageable chunks, and we have various options for each chunk.

We need to define:

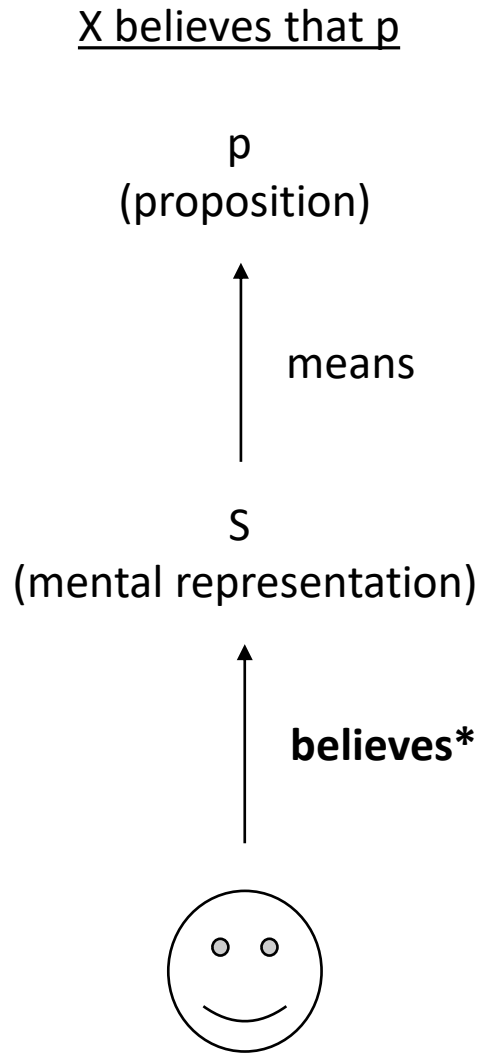
- What believes* is
- What mental representations are
- What ‘means’ means
- What propositions are

Answering these questions will motivate the approach we’ll take in the course.



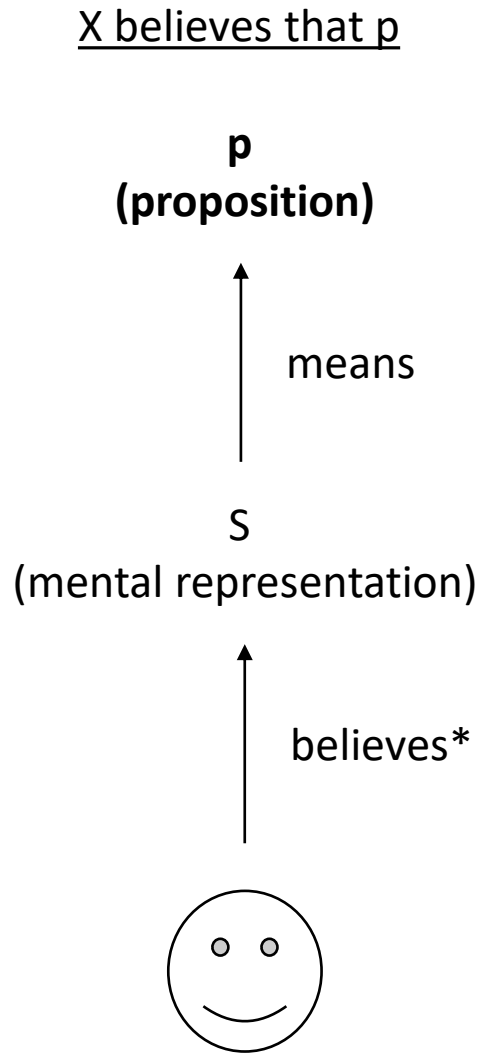
believes*

- The most popular account of believes* is the *functionalist* one (e.g. Fodor, 1987).
- Functionalism is currently the most popular approach among philosophers.
- SEP: "Functionalism in the philosophy of mind is the doctrine that what makes something a mental state of a particular type does not depend on its internal constitution, but rather on the way it functions, or the role it plays, in the system of which it is a part"
- In this account, believes* is characterized by a certain *function* it plays in mental activity.
- For instance:
  - believes* + certain desires produce certain actions
  - believes* play some role in inferences
- Questions for you:
  - What are specific examples of this role in action? E.g., in 'Mary believes that it will rain tomorrow'.
  - What else could we include in the functional role of believes*?
- The same analysis can be applied to desires*, fears*, and any other relation we hold to mental representations.



Proposition

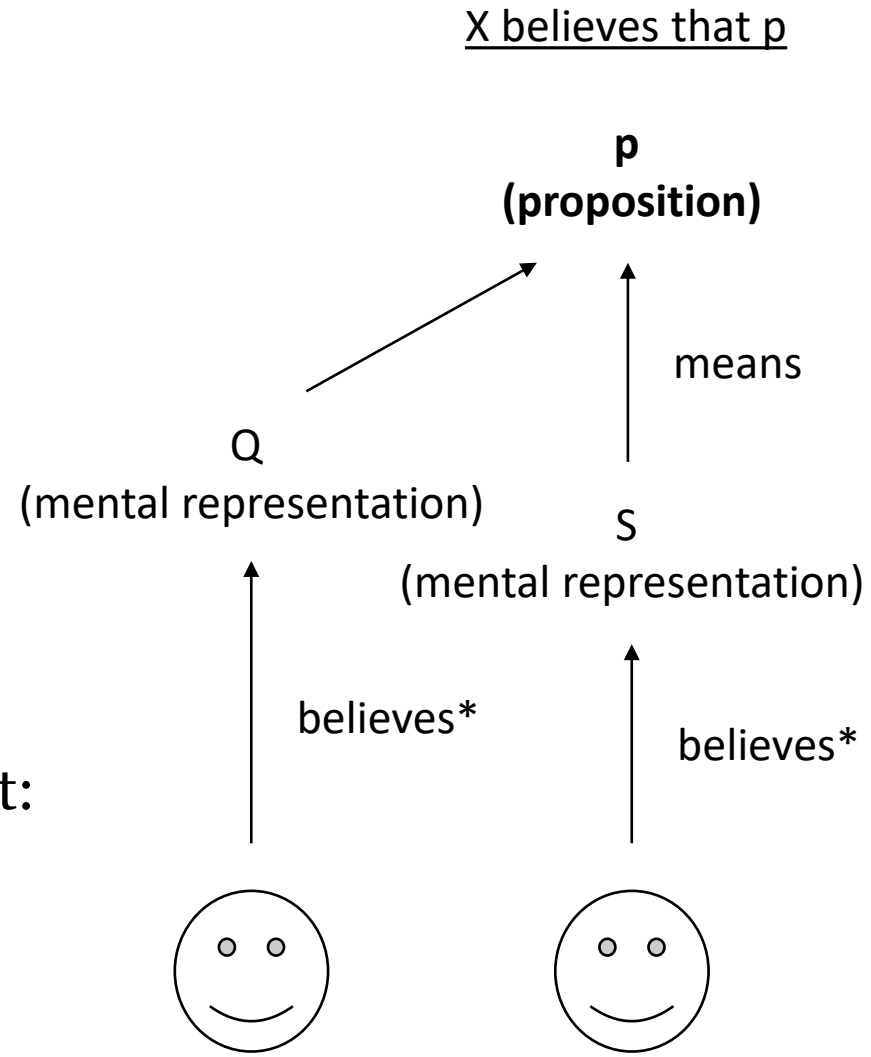
- Various ways of cashing out what a proposition is:
  1. the primary bearer of truth-value
  2. the object of belief and other “propositional attitudes”
  3. the referent of that-clauses
  4. the meaning of sentences
- Definition 2 would make our analysis of ‘believes’ circular.
- But otherwise you can just pick the one you prefer and it won’t be a problem for the rest of the course.



Proposition

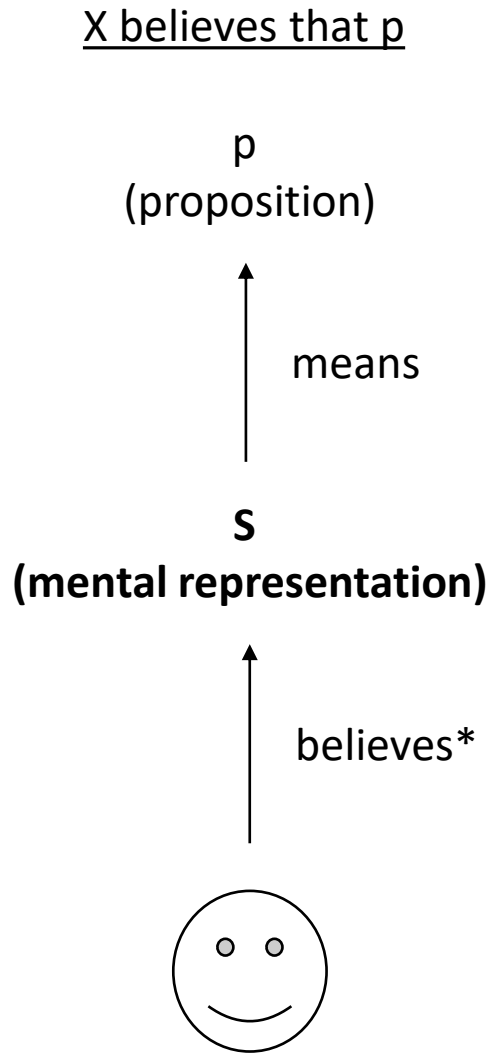
- To get an intuition for what the concept of ‘proposition’ is *for*, consider the relation between:
  1. I ate a donut
  2. A donut was eaten by me
- These are different *sentences*. Yet, they have something in common.
- We can say that they refer to or express the same proposition.
- It also gives us a way to make sense of what’s common about different people with the *same belief* but different mental representations of it:
  1. Mary: ‘Rome is in Italy’
  2. Pietro: ‘Roma è in Italia’

(arguably different mental representations!)



Mental representation

- As you might have guessed (since I left it for last), mental representations is what's of interest to us.
- We can kind of 'squeeze' the concept of mental representation between our (rough) definitions of 'believes*' and 'meaning a proposition'.
- A mental representation then is a mental item (something in the mind) that:
  1. Can mean propositions / Has propositions as content
  2. Can be inserted into the 'believes*' functional-role slot (the *belief-box*) & other similar slots
- For instance, if Julius believes* that Mary ate an apple:
  1. Something is going on in Julius' mind
  2. That something plays the functional role associated with believing (e.g., in the way it interacts with desires)
  3. And the *content* of that belief is that Mary ate an apple



Mental representation

Assume that we can cook up some plausible functional account of believes*

Then, we are left with one question, keep it in mind:

- In virtue of what does a certain mental representation *mean* a certain proposition and not another one?

This is an incredibly difficult question. Called the problem of *intentionality*.

However, we don't need to worry too much about what mental representations *are* (this would be a whole course in itself!), as long as we can say enough about them for our purposes.

Notational convention: From now on, when I put something in quotes I will specify whether in each case I mean a mental representation, a sentence, or a proposition.

So what do we need to say about mental representations?

Constituency & compositionality

Well, most people agree that some mental representations can be *constituents* in other mental representations.

For instance, take the mental representation

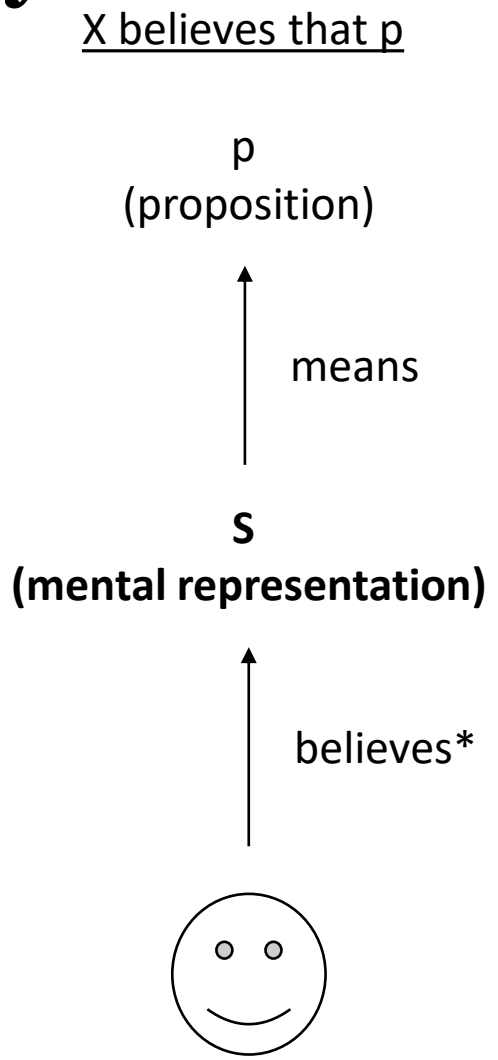
- P: ‘Samantha ate an apple’

It seems likely that the mental representations for ‘Samantha’ and ‘apple’ as in some sense *parts* of P.

Moreover, the structure is **compositional**:

- The *meaning* of a complex mental representation is a function of the *meaning* of its parts and the way they are combined.

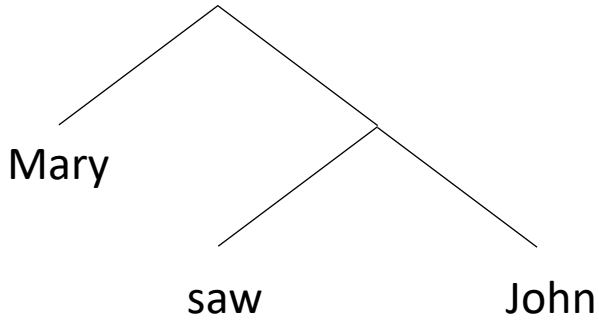
Compositionality is an enormously important concept!



EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Compositionality

- Compositionality is such an important concept that I think it's worth stopping for a moment and thinking about it.
- Can you think of an example to illustrate compositionality?



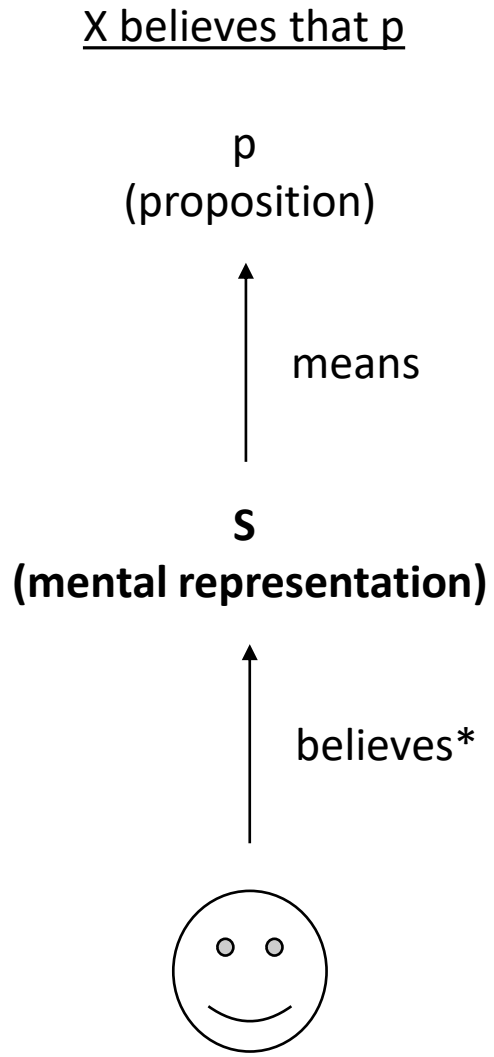
Mental representations

Note that the assumptions that complex mental representations have a compositional structure simplifies the question about what mental representations mean.

Now there's two (hopefully simpler) questions:

1. How do *simple* mental representations mean?
2. How do the meanings of simple mental representations combine?

In the rest of this course, we'll mostly be interested in the latter question (among other questions).



EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Here comes the Big Idea!

To make sense of attributions of beliefs, we have introduced a picture where mental representations have compositional structure. What does this remind you of?

The Stanford Encyclopedia of Philosophy (famous reputable source & solace of philosophers worldwide) can help us here:

- The *Language of Thought Hypothesis* (LoTH) proposes that thinking occurs in a *mental language*.
- Often called *Mentalese*, the mental language resembles spoken language in several key respects:
    1. It contains words that can combine into sentences
    2. The words and sentences are meaningful
    3. Each sentence's meaning depends in a systematic way upon the meanings of its component words and the way those words are combined

There's a lot going on in those three points. They will become clearer over time.

But I hope this already gives you a sense of what the LoTH is.

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Some specific facts about the LoT

- Question for you: What can we already say about the LoT?
- You might find this surprising, but we can plausibly say some quite *specific* things about the LoT.
- For instance, the LoT has something like ‘words’ (simple mental representations), which express simple (=unstructured) concepts.
- Some of the simple words in e.g., English probably correspond to complex mental representations.
- E.g., ‘bachelor’ -> unmarried human man
- We can also guess some words in our LoT. For instance, basic logical words like ‘and’, ‘or’, ‘not’, ‘all’, ‘some’.
- In other terms, it’s plausible that the LoT contains at least some logical scaffolding.

What does the LoT do for us?

The LoT gives us natural accounts of:

- Reasoning
  - Can you see what the LoT would say about reasoning?
- Why some mental representations feel more ‘complex’ than others
  - Can you see why?
- Connections between the meanings of different mental representations
- The apparently close relations between thought and language
- How can we have infinitely many thoughts
  - Can you see why?

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Back to rule learning

- So how does this all relate to the initial examples of rule learning?
- The idea is that learning a rule from examples consists in finding some expression in the LoT that is consistent with the examples…
- But consistency isn't enough! Remember Wittgenstein. We need something more.
- For instance, we could say that humans have a *prior preference for simpler LoT expressions*.
- This explains the Dr. case but not the summation case! So something more complicated is going on.
- And elaborating this idea is going to take us the next 13 weeks!

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Summary & road ahead

- We have seen a certain *picture* of mental representations.
- The picture is that mental representations have the structure of a language: they have parts that combine compositionally, and they have propositional content.
- Therefore, mental representations are structured like a language, the LoT.
- Beyond propositional attitudes, this helps us make sense of loads of stuff!
- But what does python have to do with this?, you ask
- In the last twenty years or so, people have been combining the idea behind the LoT with probabilistic and neural approaches into computational models that can be applied to specific aspects of human behaviour.
- This is the *probabilistic* Language of Thought which this course is about!

“But what about neural networks!”

- It wouldn’t be 2022 if I didn’t mention neural networks
- Jerry Fodor talked about this in his typically caustic style:

> In particular, the standard current alternative to Turing architecture, namely, connectionist networks, is simply hopeless. Here, as so often elsewhere, networks contrive to make the worst of both worlds. They notoriously can’t do what Turing architectures can, namely, provide a plausible account of the causal consequences of logical form. But they also can’t do what Turing architectures can’t, namely, provide a plausible account of abductive inference. It must be the sheer magnitude of their incompetence that makes them so popular.

- It’s not at all clear that this is a fair criticism!
- And we will see towards the end of the course that some successful applications of the LoT idea employ a combination of logical and connectionist tools.
- More on connectionism vs LoT in two weeks!

Next time I

- In the next *lecture* (Monday!), we'll continue our exploration in the philosophical / conceptual foundation of the Language of Thought. We'll see some arguments in favour of the LoTH.
- In the next *lab*, we'll start getting you all set up with running the lab notebooks, and we will also start with a 2-weeks introduction to python
  - Although this plan might change depending on what people's background is!
- If you're feeling like you want more in the meantime, you can watch this: https://www.youtube.com/watch?v=gjc5h-czorI

Next time I

- The reading for next time is the SEP page on the Language of Thought.
- I hope that the explanation this week will help you understand what's said in the SEP entry.
- Please come up with ONE question about the reading (e.g., something you found difficult to understand or something that strikes you as wrong).
- If there is time left after the lecture next time, we can discuss them. Otherwise, it's a great exercise to try to formulate to yourself what's not clear about a text.

Questions?

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN