[AI-PHI] 20th SESSION

Causerie on Reasoning

AI-Phi Community

20/03/2025

Causerie

- A Topic
- Framing the topic
- Questions
- Ideally… some output

Output

About AI-Phi

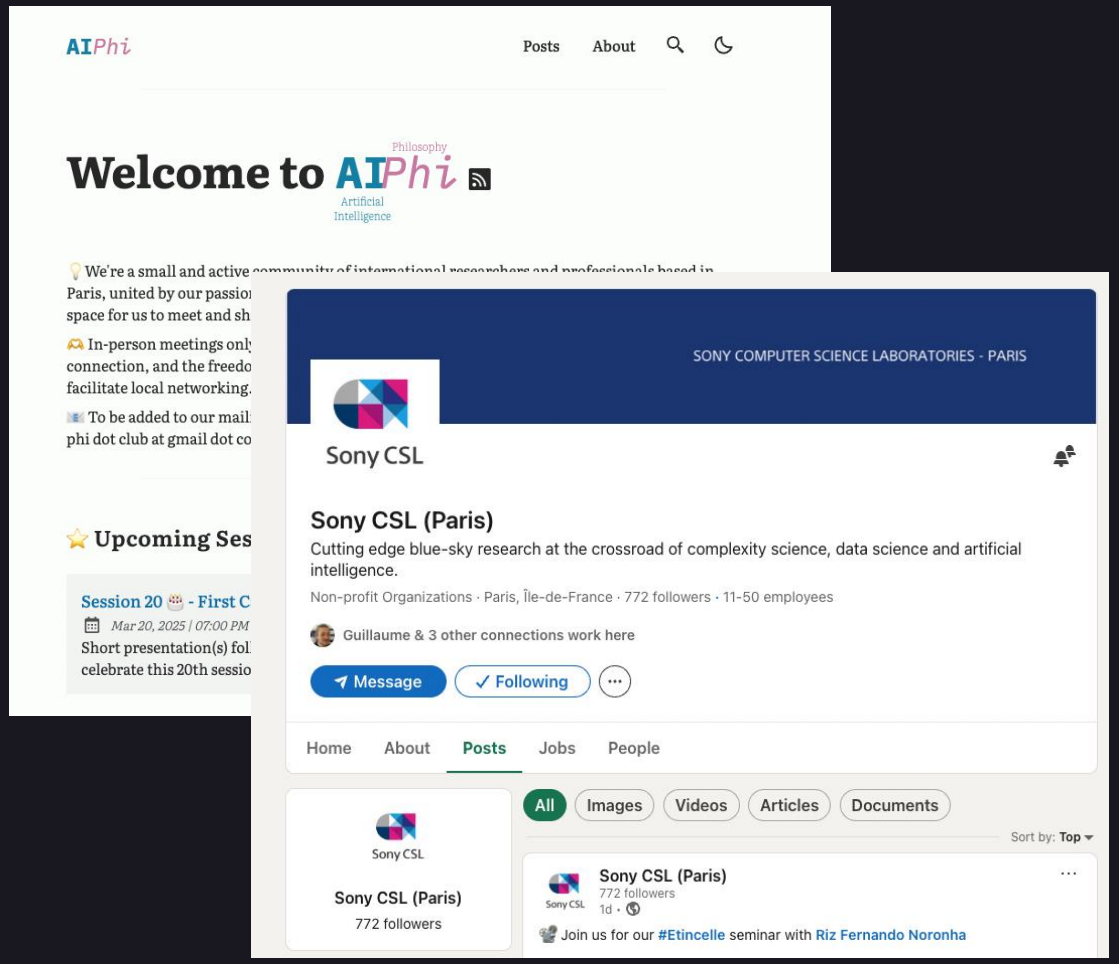
What we discussed in general

Opinions:

- “I think” – Person A
- “therefore” – Person B
- “wait…” – Person C
- “am I reasoning?” – Person D

Closing remarks

Format



Publishing

Reasoning

Being unreasonable

Acting based on emotions

Making statements without justification

Being incoherent

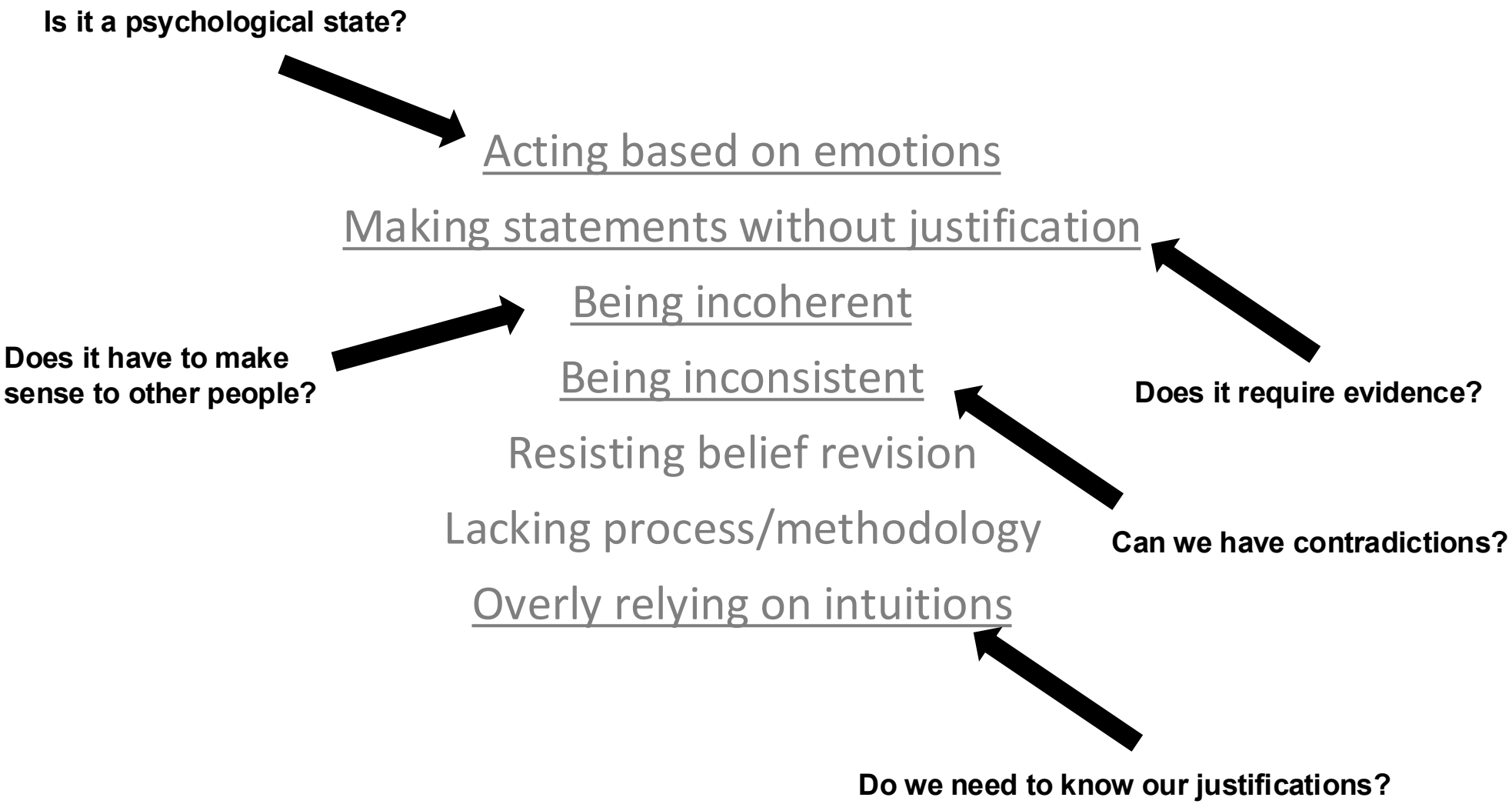
Being inconsistent

Resisting belief revision

Lacking process/methodology

Overly relying on intuitions

Being unreasonable



Types of Reasoning

Deductive

Premises > Conclusions

Inductive

Examples > General Statements

Abductive

Outcomes > Explanations

The grass is wet. Maybe it rained?

Counterfactual

If this were the case what would happen?

Causal

X happened because of Y

Spatial

If I rotated this object by 90 degrees, what would it look like?

But what is reasoning?

Reasoning

?

?

Directed thought

Focused on generating new knowledge

Based on existing knowledge

With some justification

?

?

Reasoning

When are we reasoning?

Not Reasoning

Reasoning

Random

Rabbits on mars!

Baaaa

Poorly Justified

Repeat what I heard in a similar situation

Say *anything* just so that you said something

I heard it from x

I feel that it is true

"I heard it from x and they are an expert"

I feel that it is true because…

Well justified

This is how I think it works and why I think it is true

I used this formal system and argued for my application of it

Reasoning

Does any of this have to be true/consistent/useful to be reasoning?

Not Reasoning

Reasoning

Purely Random

Poorly Justified

Well justified

Rabbits on mars!

Repeat what I heard in a similar situation

"I heard it from x"

"I heard it from x and they are an expert"

This is how I think it works and why I think it is true

I used this formal system and argued for my application of it

Say anything so that you said something

I feel that it is true

I feel that it is true because…

Baaaa

Logic

Reasoning & Logic

Logic is an instrument (organon) for for discovering knowledge in science, ethics and metaphysics.

Famous for Syllogistic Logic:

Premise 1: All humans are mortal.

Premise 2: Socrates is a human.

Conclusion: Socrates is mortal.

$A = B$

And $B = C$

Therefore $C = A$

Laws of Thought:

Law of Identity: A thing is what it is.

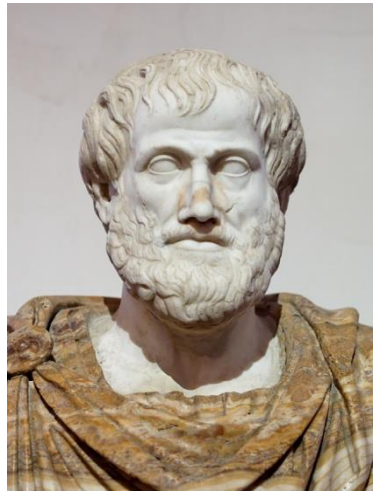
(A is A)

Law of Non-Contradiction: A statement cannot be both true and false at the same time.

(A cannot be both A and not-A)

Law of the Excluded Middle: A statement must be either true or false, with no middle ground.

(Either A or not-A)



Aristotle: 384 - 322 BC

Reasoning & Logic

"… an art is needed to direct the act of reasoning, so that by it a man when performing the act of reasoning might proceed in an orderly and easy manner and without error. And this art is logic, that is, the science of reason."

– Thomas Aquinas, Commentary on the Posterior Analytics of Aristotle.

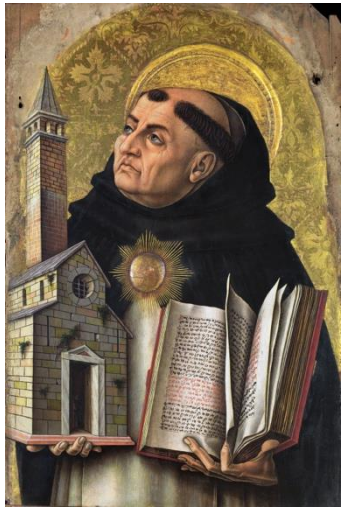
"Art" (Latin *ars*, Greek *techne*) = ability to do something

The logic of Aquinas:

1. Understand what is indivisible. i.e. what 'things' are there?
   (dog, animal, sky, blue)

2. How you can combine or divide things by making judgements (truths).
   (dogs are animals, the sky is blue)

3. Advancing in such a way that what is known arrives at what is unknown.

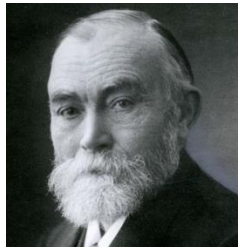


Thomas Aquinas: 1225-1274 AD

Reasoning & Logic

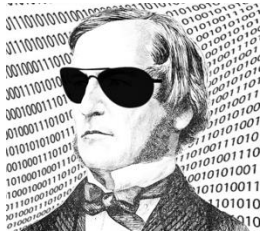
Formal/Mathematical Logic

From the 19th Century



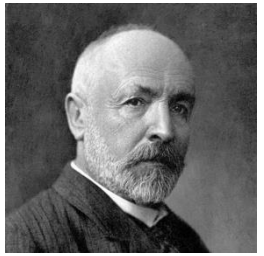
Gottlob Frege

Predicate Logic



George Boole

Boolean Algebra



Georg Cantor

Set Theory

• • •

Concrete Expressions > Abstract Symbols

Deductive:

Premises > Conclusion

Correctness

True premises result in true conclusions

…when following valid inference rules

Formal/Mathematical Logic

Where is the reasoning happening?

The mathematician or the logical system?

Can the logical system reason?

Reasoning in AI

Rule-based Systems (Deductive)

A list of rules (if-then)

A knowledge base of facts

An inference engine

- Applies rules to the known facts
- Infers new facts

Used for making conclusions/decisions

Machine Learning (Inductive)

Learn patterns and relationships from data

No requirement for explicit definitions of rules

Typically requires the definition of ‘features’ and ‘feature engineering’

Experts decide what features are relevant to model

Deep Learning (Inductive)

A subset of Machine Learning

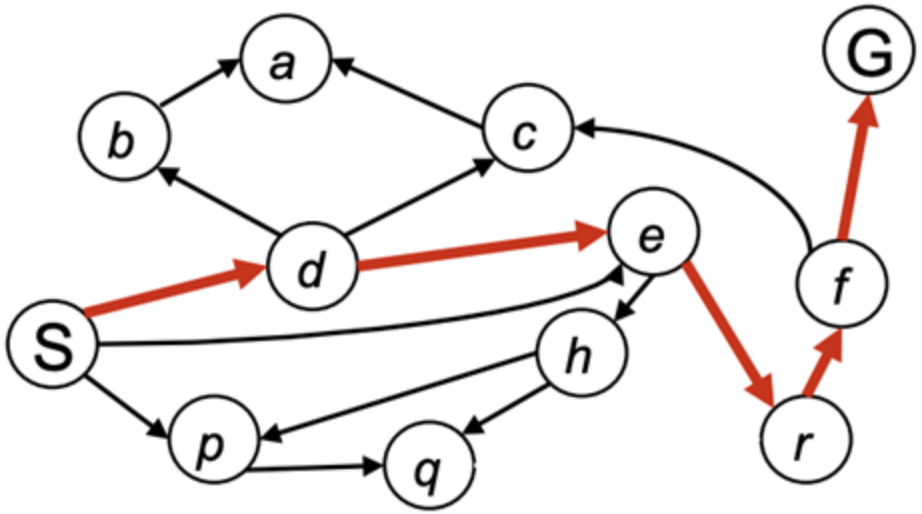
Representations learned from data

- Less reliance on domain experts
- No feature engineering

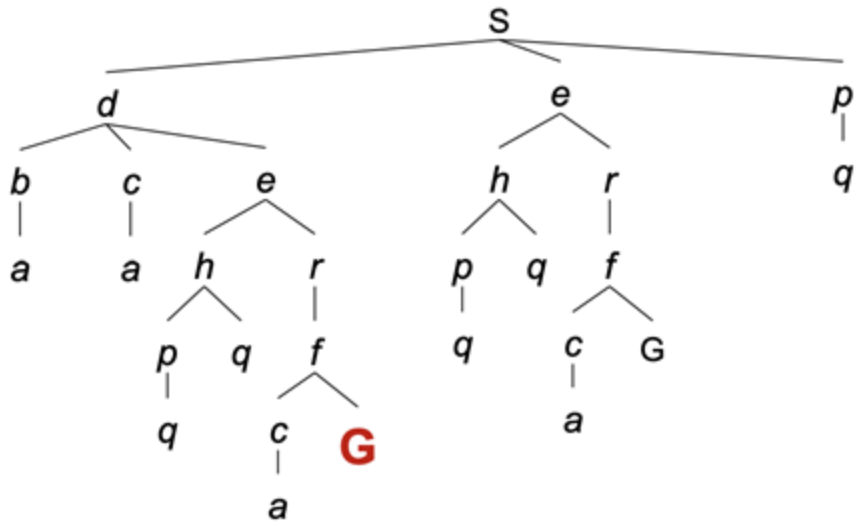
Latent spaces - An abstract representation of data

Search

State Space



Search Tree



Moving through a state space to maximise or minimise a goal

Search

Allows for systematic exploration

Reasoning often involves exploring many potential paths

Particularly in some domains:

Constrain satisfaction, planning, adversarial games

Search

Is search a tool used to arrive at good answers?

Or is it a core part of reasoning itself?

Large Language Models (?)

'Reasoning' models are a chimera…

Model of… language… *and* thought (indirectly)… *and* the world (indirectly)

Some kind of instruction following.

Search + reinforcement learning.

An inductively trained system that…

You can ask to do many

types of reasoning.

Large Language Models (?)

'Reasoning' models are a chimera…

Model of… language… *and* thought (indirectly)… *and* the world (indirectly)

Some kind of instruction following.

Search + reinforcement learning.

An inductively trained system that…

You can ask to do many types of reasoning.

But is it really?

A Handful of Useful Concepts

Intension & Extension

Intension

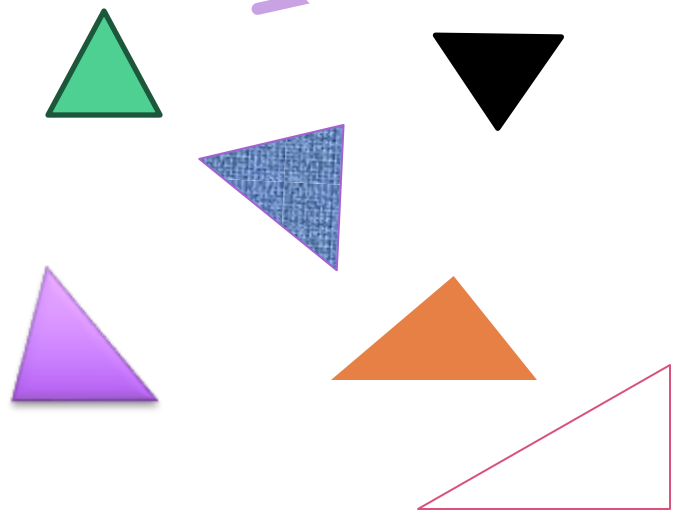
What does the concept mean?

Extension

What does the concept refer to?

Triangle

"A triangle is a polygon with three corners and three sides, one of the basic shapes in geometry. The corners, also called vertices, are zero-dimensional points while the sides connecting them, also called edges, are one-dimensional line segments"



Intension & Extension

Are statistical models fundamentally extensional?

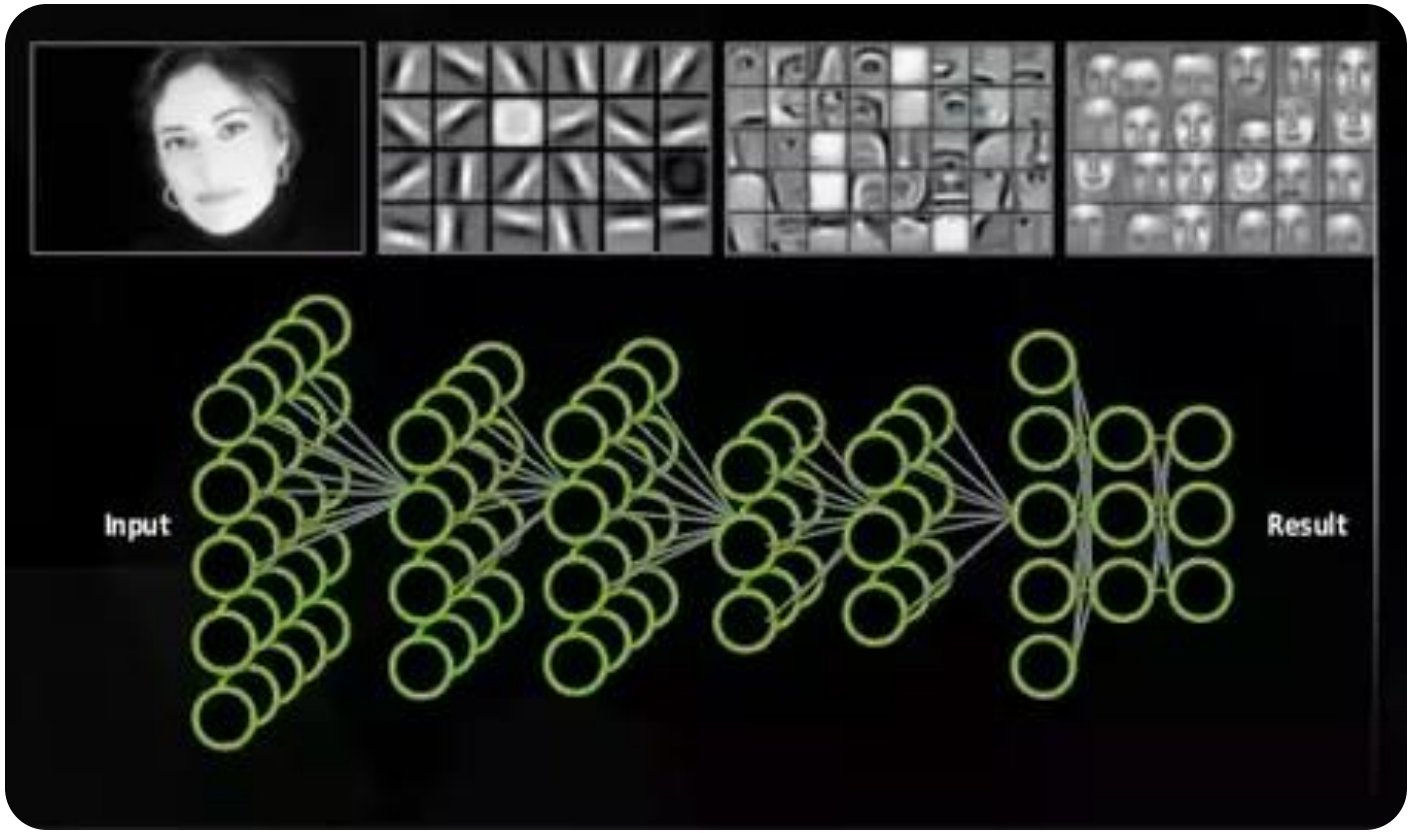
They just predict extensions of a concept via pattern matching.

Or… have they learned predictive intensions?

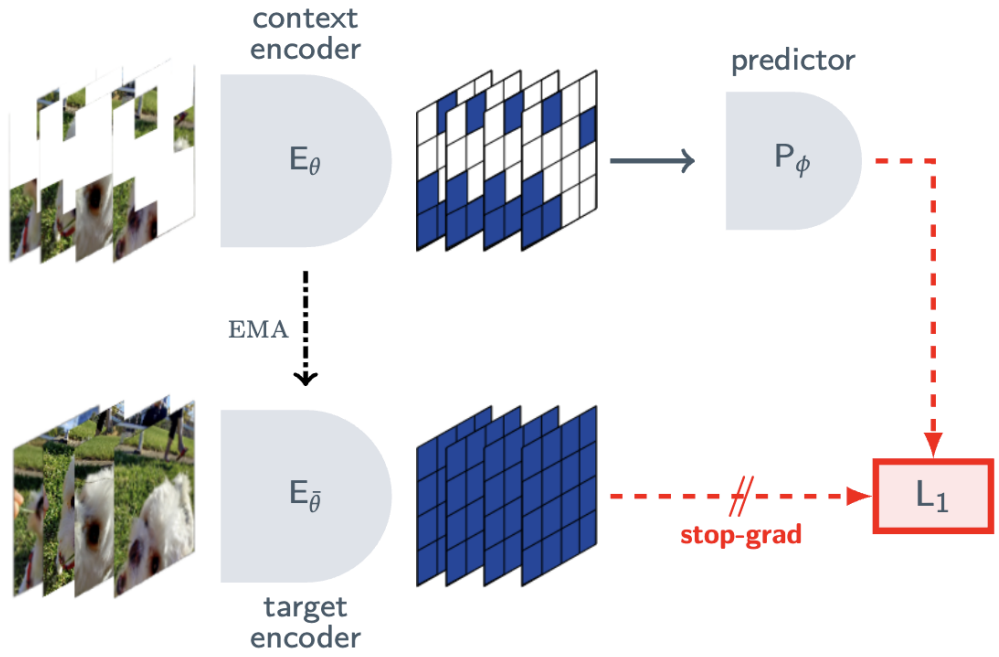
In the magic of their latent spaces.

Intension & Extension

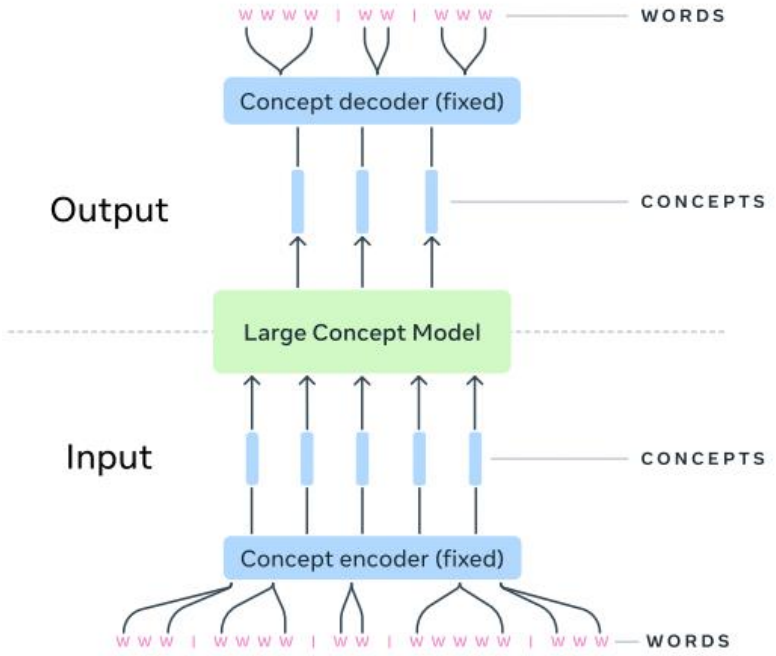
Intension ? Extension



Intension & Extension



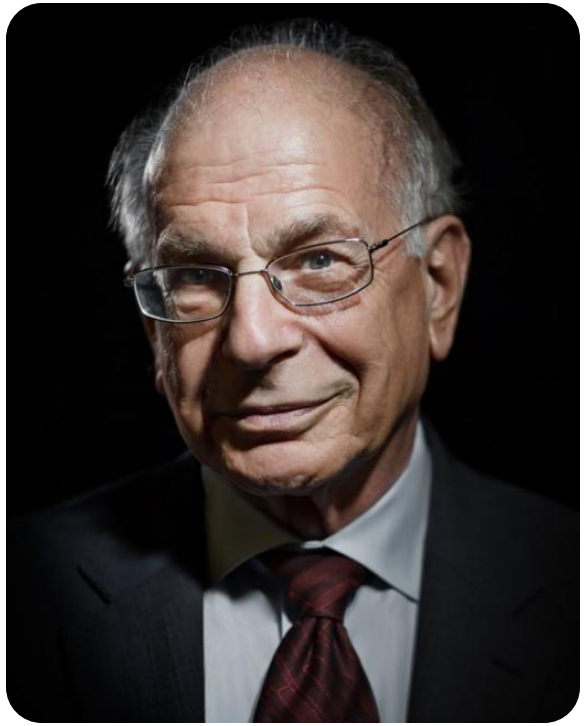
Yann Lecun's Joint Embedding Predictive Architecture. FAIR at Meta.



Large Concept Models: Language Modelling in a Sentence Representation Space. FAIR at Meta.

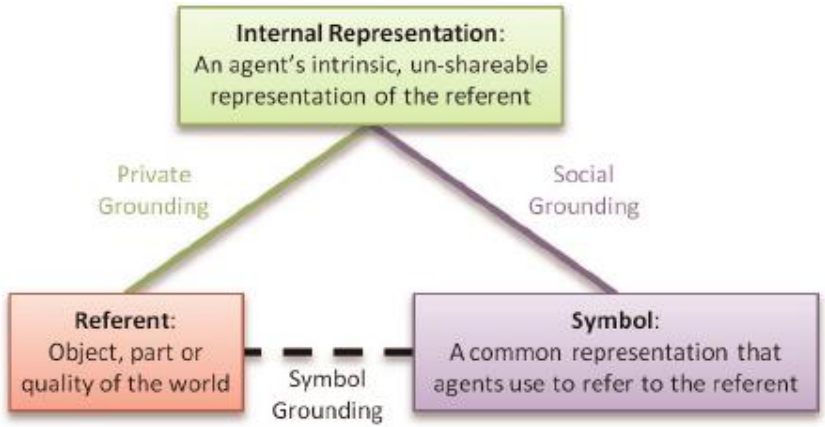
Thinking Fast and Slow

System 1	System 2
Fast	Slow
Subconscious	Conscious
Automatic	Effortful
Everyday Decisions	Complex Decisions
Error Prone	Reliable



Daniel Kahneman

Symbol Grounding

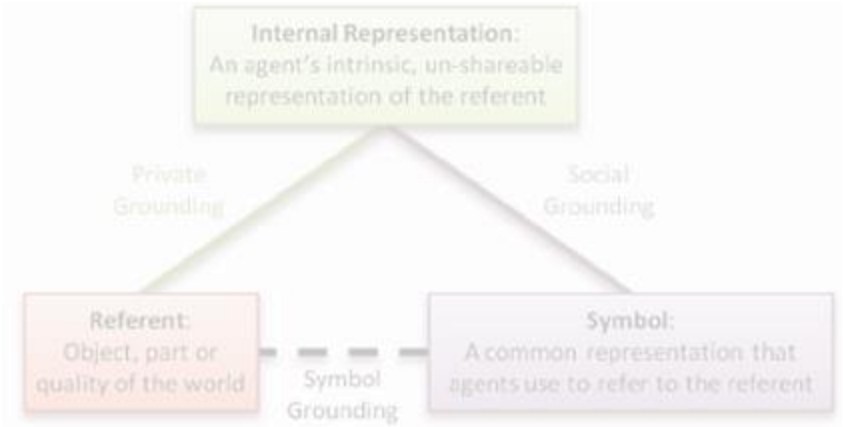


The Semiotic Triangle

"How can the semantic interpretation of a formal symbol system be made *intrinsic* to the system, rather than just parasitic on the meanings in our heads?"

The symbol grounding problem. 1990. Stevan Harnad.

Symbol Grounding



The Semiotic Triangle

"How can the semantic interpretation of a formal symbol system be made *intrinsic* to the system, rather than just parasitic on the meanings in our heads?"

Are you ever really reasoning if you don't know what you are reasoning about?

The symbol grounding problem. 1990. Stevan Harnad.

Memetic Reasoning / Functional Mimicry

To what extent does anybody really ‘reason’ formally?

Can we just learn simple heuristics/rules to mimic this reliably?

How much of our knowledge have we (as individuals) learned and internalised from first principles?

How much do we just accept knowledge from society as facts?