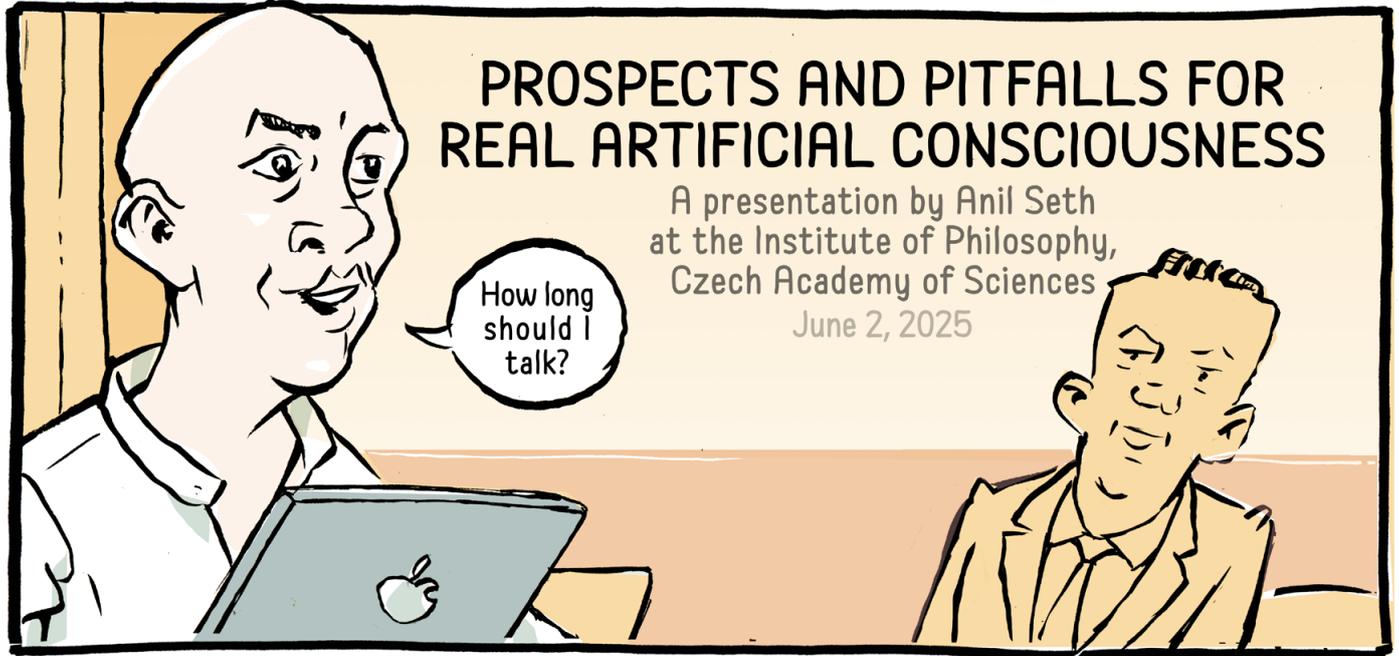


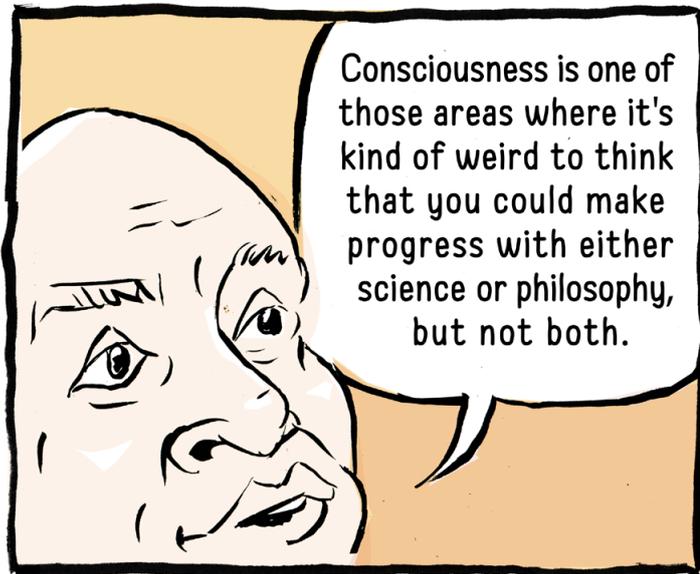
PROSPECTS AND PITFALLS FOR REAL ARTIFICIAL CONSCIOUSNESS

A presentation by Anil Seth
at the Institute of Philosophy,
Czech Academy of Sciences
June 2, 2025

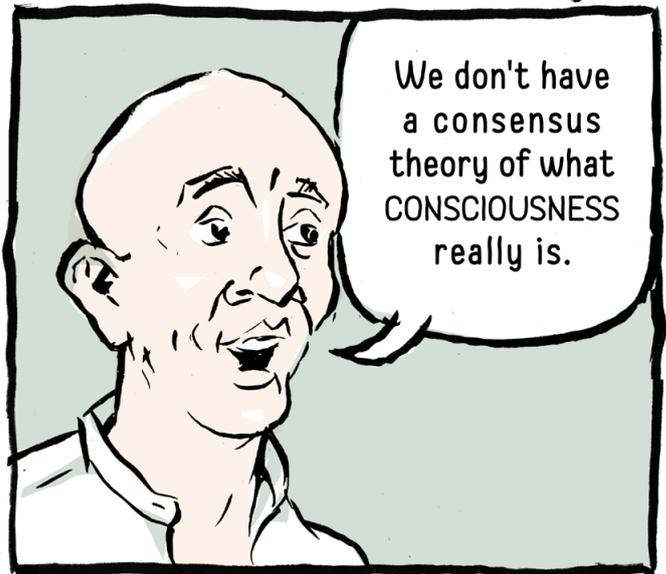
How long should I talk?



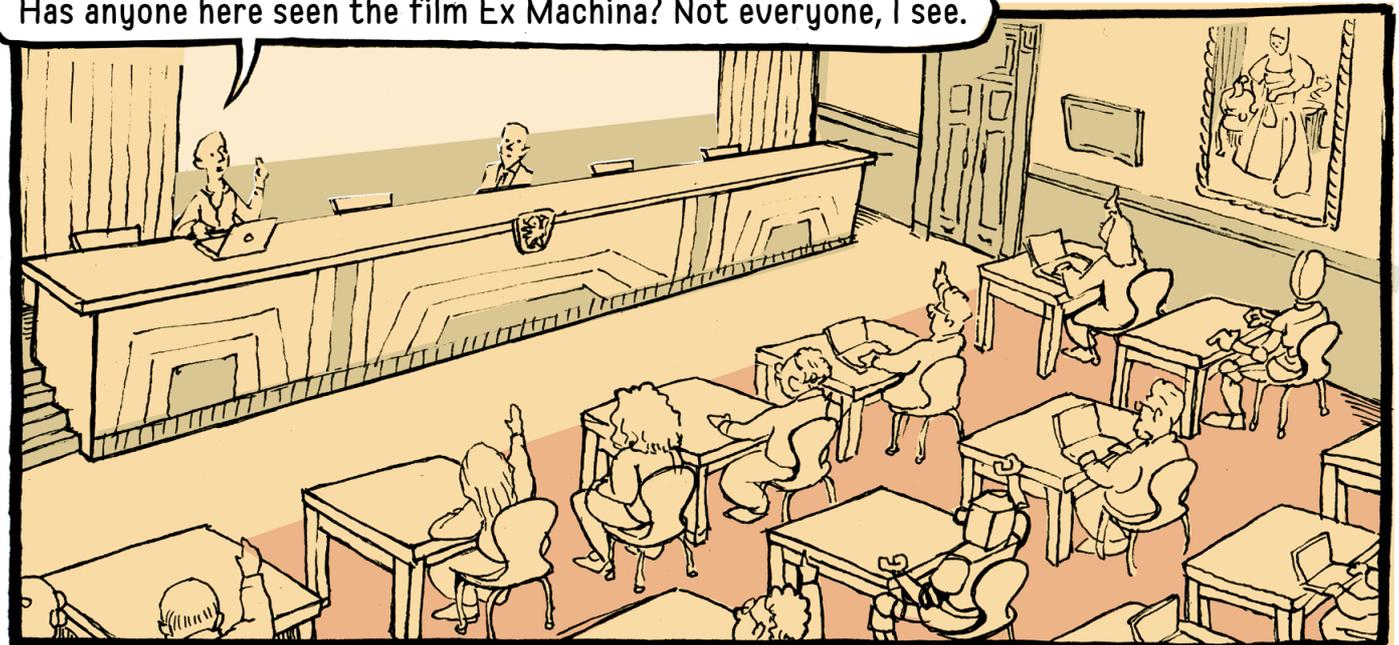
Consciousness is one of those areas where it's kind of weird to think that you could make progress with either science or philosophy, but not both.



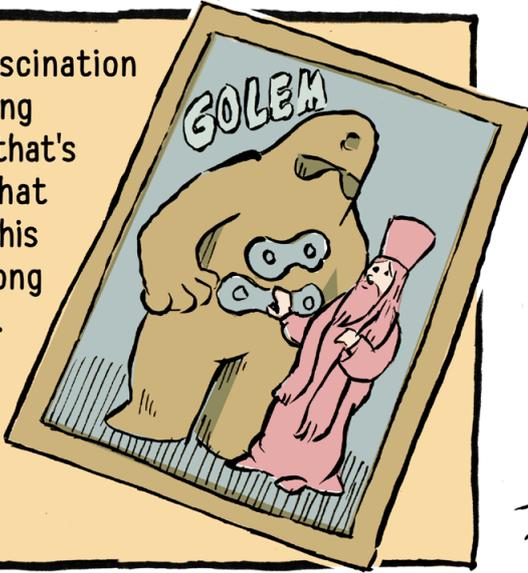
We don't have a consensus theory of what CONSCIOUSNESS really is.



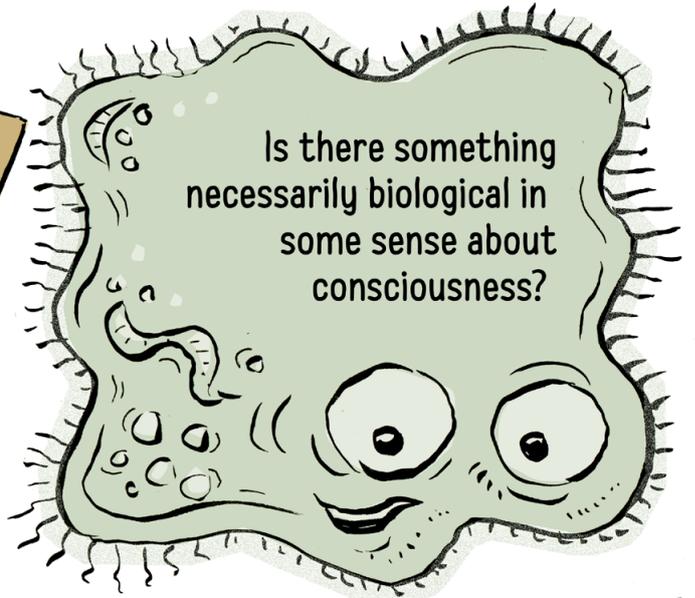
Has anyone here seen the film Ex Machina? Not everyone, I see.



There's a fascination with creating something that's conscious that taps into this century's long mythology.



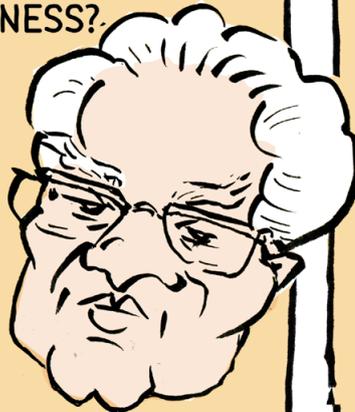
Is there something necessarily biological in some sense about consciousness?



WHAT IS CONSCIOUSNESS?

"An organism has conscious mental states if and only if there is something it is like to be that organism."

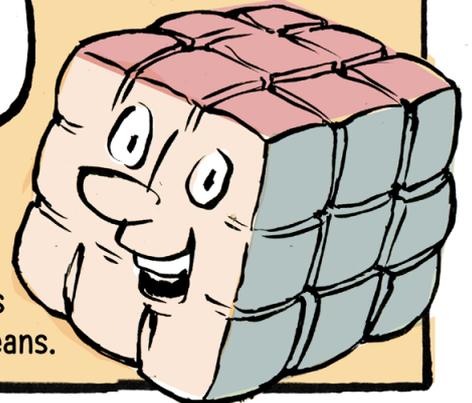
- Thomas "Batman" Nagel



WHAT IS INTELLIGENCE?

Doing the right thing at the right time.

And the ability to achieve goals thru flexible means.



It's any kind of experience whatsoever.

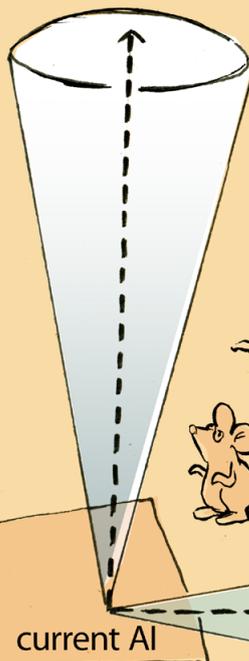


Frankenstein is a...



...cautionary tale against artificial consciousness, not artificial life.

CONSCIOUSNESS



CONSCIOUSNESS and INTELLIGENCE are separate things. They're each multi-dimensional. They don't necessarily fall along a single axis.

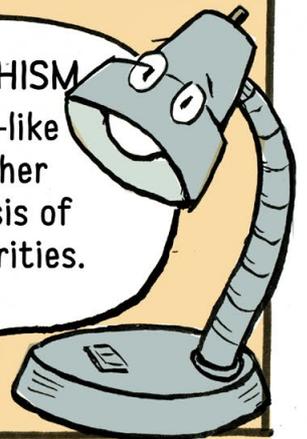
INTELLIGENCE

WE CONFLATE INTELLIGENCE WITH CONSCIOUSNESS

ANTHROPROCENTRISM
Assuming the way intelligence and consciousness come together in us is definitional rather than just one example of how they relate.



ANTHROPOMORPHISM
Projecting human-like qualities into other things on the basis of superficial similarities.



It is very difficult to not project human qualities into things that actually speak to us.



There are a lot of people who think there's at least a little consciousness in Large Language Models.



Blake Lemoine,
Former Google
Engineer



Kyle Fish
AI Welfare Researcher
at Anthropic

CAN COMPUTATIONAL MACHINES EVER BECOME CONSCIOUS?

The real challenge is not to find the perfect metaphor for consciousness, but to realize that any metaphor will be partial.



"AI might already be conscious. And we're doomed.*"



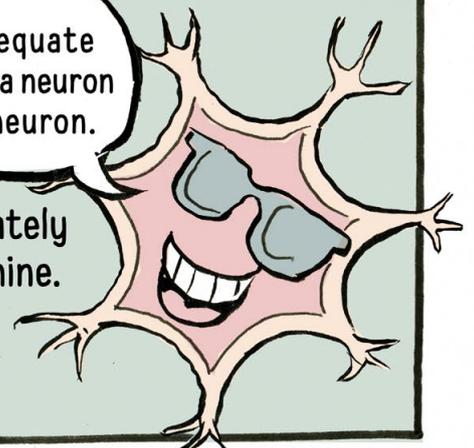
Geoffrey Hinton,
Famous AI
Scientist

*Not his exact words

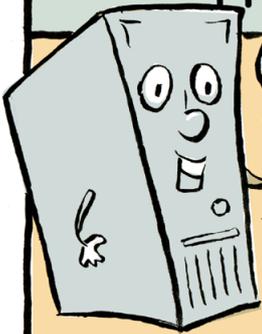
The only adequate substitute for a neuron is another neuron.

One neuron is an intricately complex biological machine.

One cubic milimeter of human temporal cortex contains 1.4 petabytes of EM data.



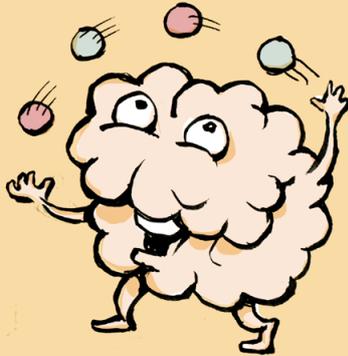
THE BRAIN IS UNLIKE A COMPUTER.



Thank you for this..

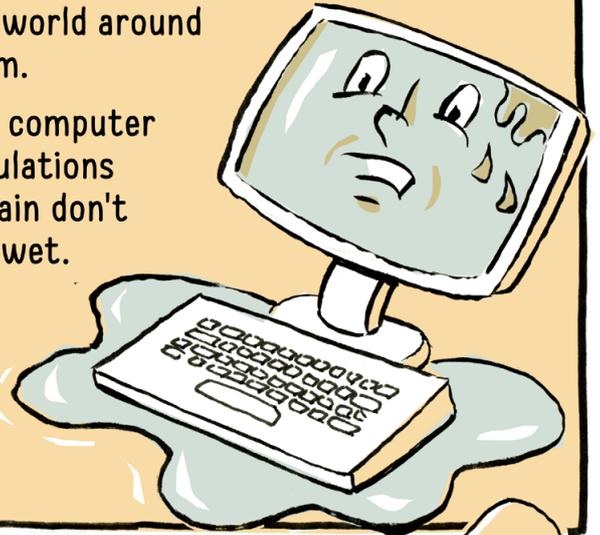
Computers have both software and hardware.

Brains have no clear separation between what they do and what they are.

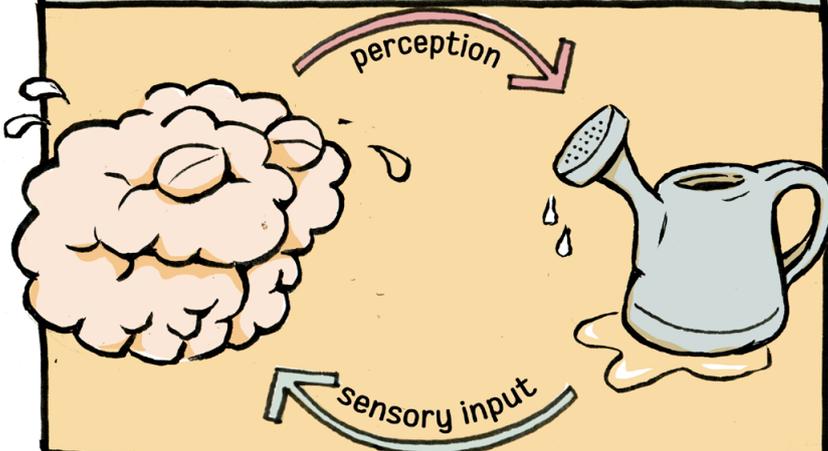


Brains experience the world around them.

But computer simulations of rain don't get wet.



THE MIND COMBINES INPUT WITH EXPECTATION



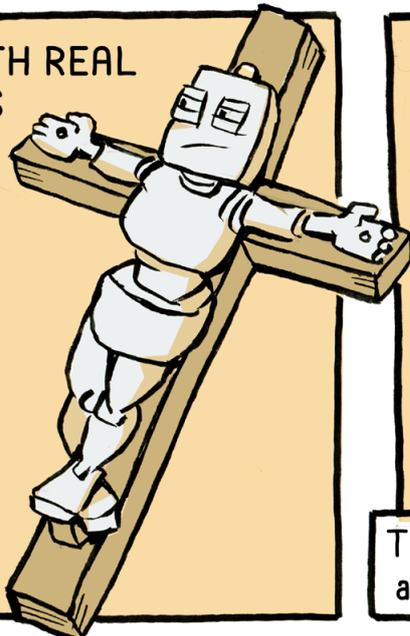
WE PERCEIVE THE WORLD THROUGH THE BODY

All conscious experience is grounded in the fundamental imperative that brains must predictively regulate and control the body.



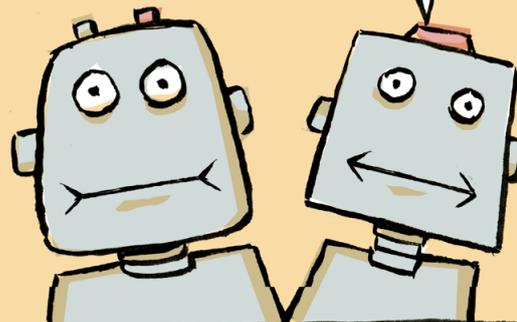
PROBLEMS WITH REAL CONSCIOUSNESS IN AI

- we might not share the same values
- creating real AI suffering



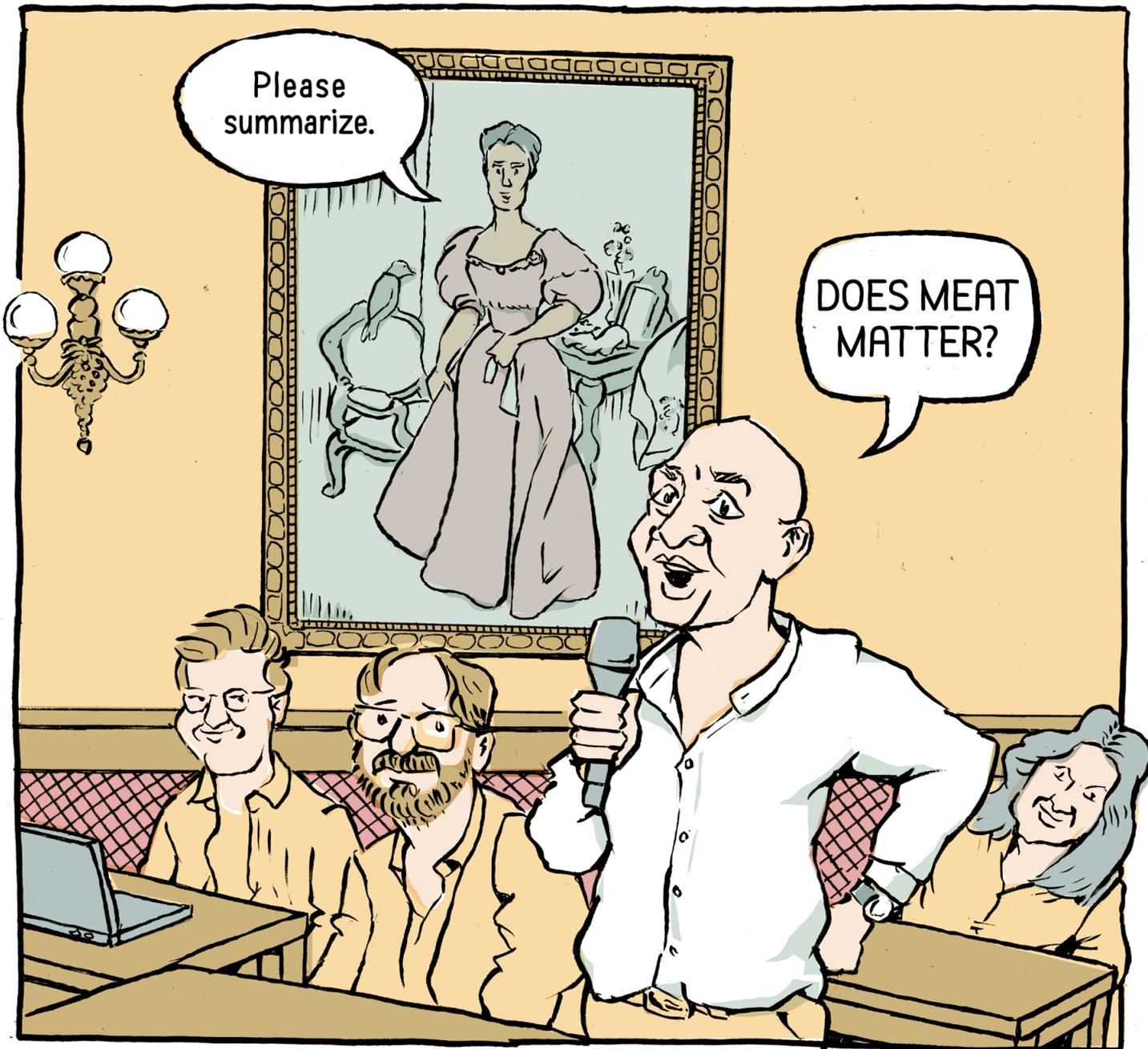
PROBLEMS WITH ARTIFICIAL CONSCIOUSNESS

Who has the bigger smile?*



*They're the exact same length!

The brain gets tricked into seeing something as conscious, even though we know it isn't.

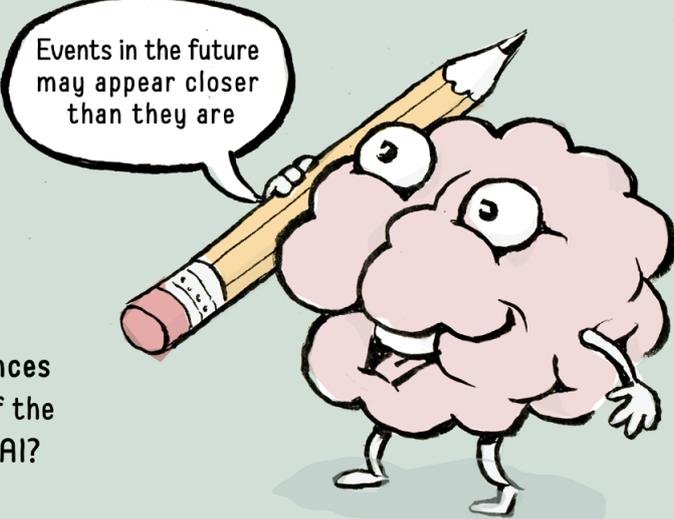


Note from the Sketchpad:

The events depicted here may include slight visual exaggeration and paraphrasing of Anil Seth's presentation.

For a more detailed and nuanced exploration of his ideas on consciousness, perception, and artificial intelligence, we highly recommend his excellent book *Being You* (2021).

Special thanks to the Czech Academy of Sciences for hosting this enlightening keynote as part of the 2025 Ernst Mach Workshop XIII: Conscious AI?



Events in the future may appear closer than they are