

A SHORT TIME AGO I CEASED TO EXIST

### ANAESTHESIA:

People into OBJECTS & back again

OBLIVION  
Bye!

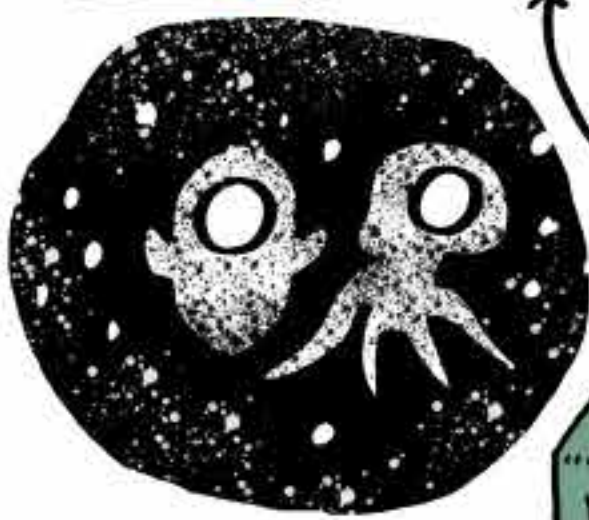


We experience JOY and SUFFERING, what about OTHER ANIMALS? Might they be CONSCIOUS TOO?

BILLIONS of BIOLOGICAL MACHINES (neurons) = CONSCIOUS EXPERIENCE

COME BACK BABY

We can get at the mechanisms of our misperceptions of ourselves



HUMAN CONSCIOUSNESS IS ONE TINY REGION in a VAST UNIVERSE of CONSCIOUSNESS

WE PREDICT OURSELVES INTO EXISTENCE

STAYING ALIVE is grounded in BIOLOGICAL MECHANISMS L.A.I. can't replicate this

OUR CONSCIOUS EXPERIENCES are Controlled HALLUCINATIONS



Once we EXPLAIN its PROPERTIES in terms of things happening inside BRAINS & BODIES then the INSOLUBLE MYSTERY starts to FADE AWAY



### WORLD MOVIE

### CONSCIOUSNESS



• MULTI-SENSORY  
• IMMERSIVE

I'M GOOD IN THIS!

### SELF MOVIE

→ We cling to a sense of SELF

STARRING ROLE!



• HAVING a BODY • 1st PERSON VIEW  
• BEING a CAUSE • CONTINUOUS SELF

## CONSCIOUSNESS ... HOW?

... WITH, THROUGH & BECAUSE of OUR

## LIVING BODIES

INTEROCEPTION - Perceiving BODIES from within

CONTROL + Regulation

### BEING a BODY:



That's my HAND!

Seeing & feeling create a FALSE BEST GUESS!

# YOUR BRAIN MAKES YOUR CONSCIOUS REALITY

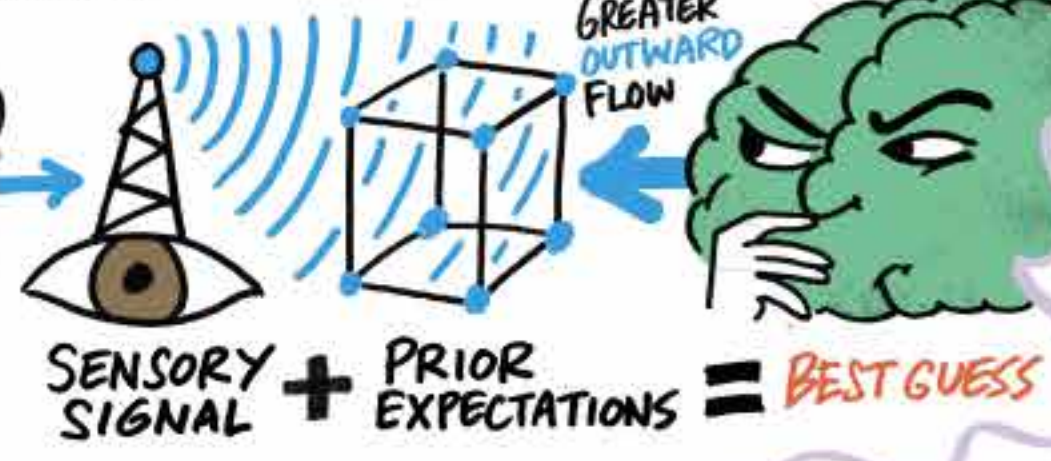
BEST GUESSES INSIDE OUTSIDE

PERCEPTION: controlled HALLUCINATION  
Hallucination: UNCONTROLLED PERCEPTION



WHAT IS IT?

The BRAIN is a PREDICTION ENGINE that relies upon INFORMED GUESSWORK.



SENSORY SIGNAL + PRIOR EXPECTATIONS = BEST GUESS

### WE ACTIVELY GENERATE the WORLD



overly strong perceptual predictions Create HALLUCINATORY experiences

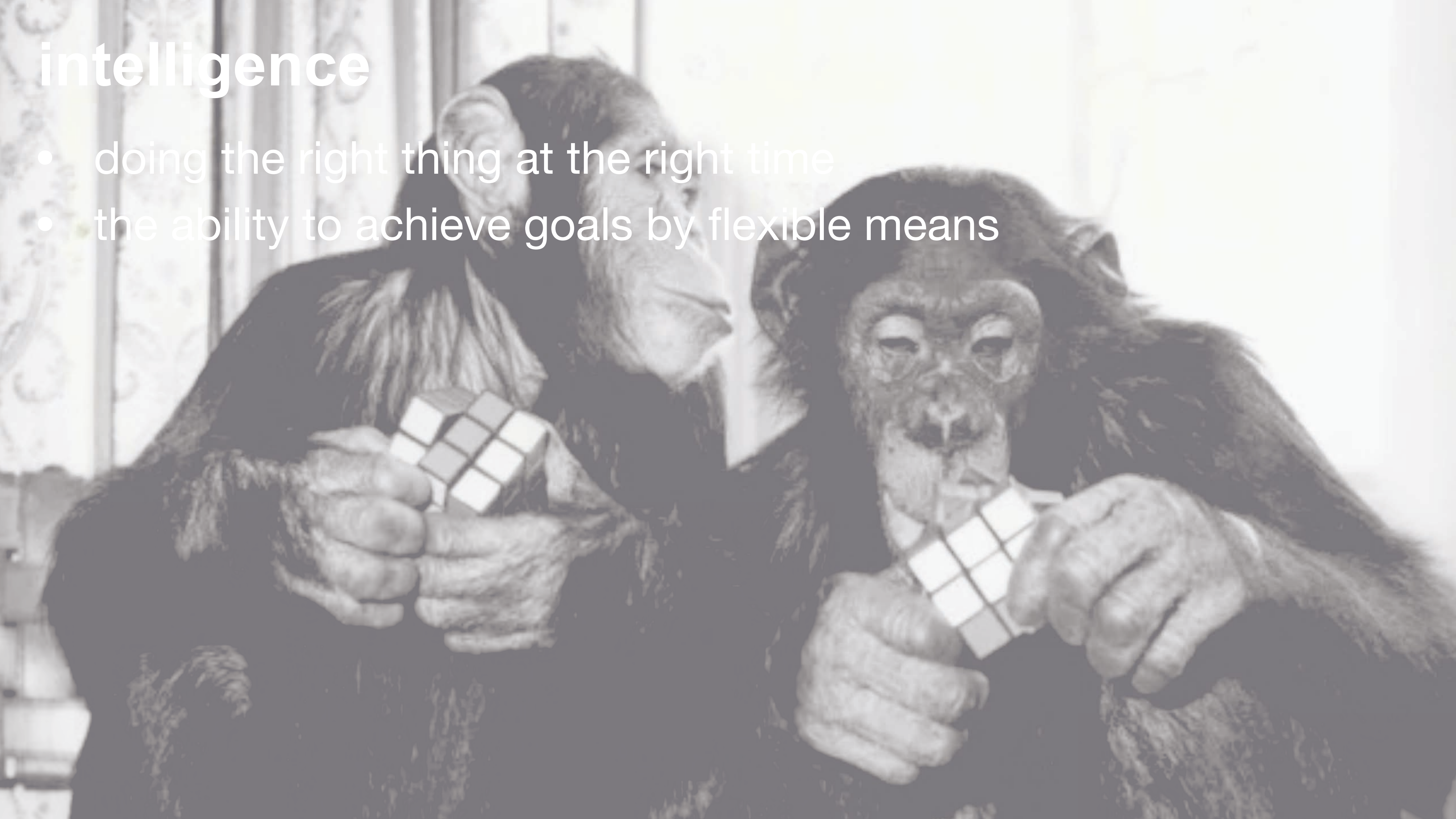
NO light or sound, just ELECTRICAL IMPULSES indirectly related to things in the WORLD

the intelligence we have,  
and the intelligence we need

# (artificial) intelligence

# intelligence

- doing the right thing at the right time
- the ability to achieve goals by flexible means



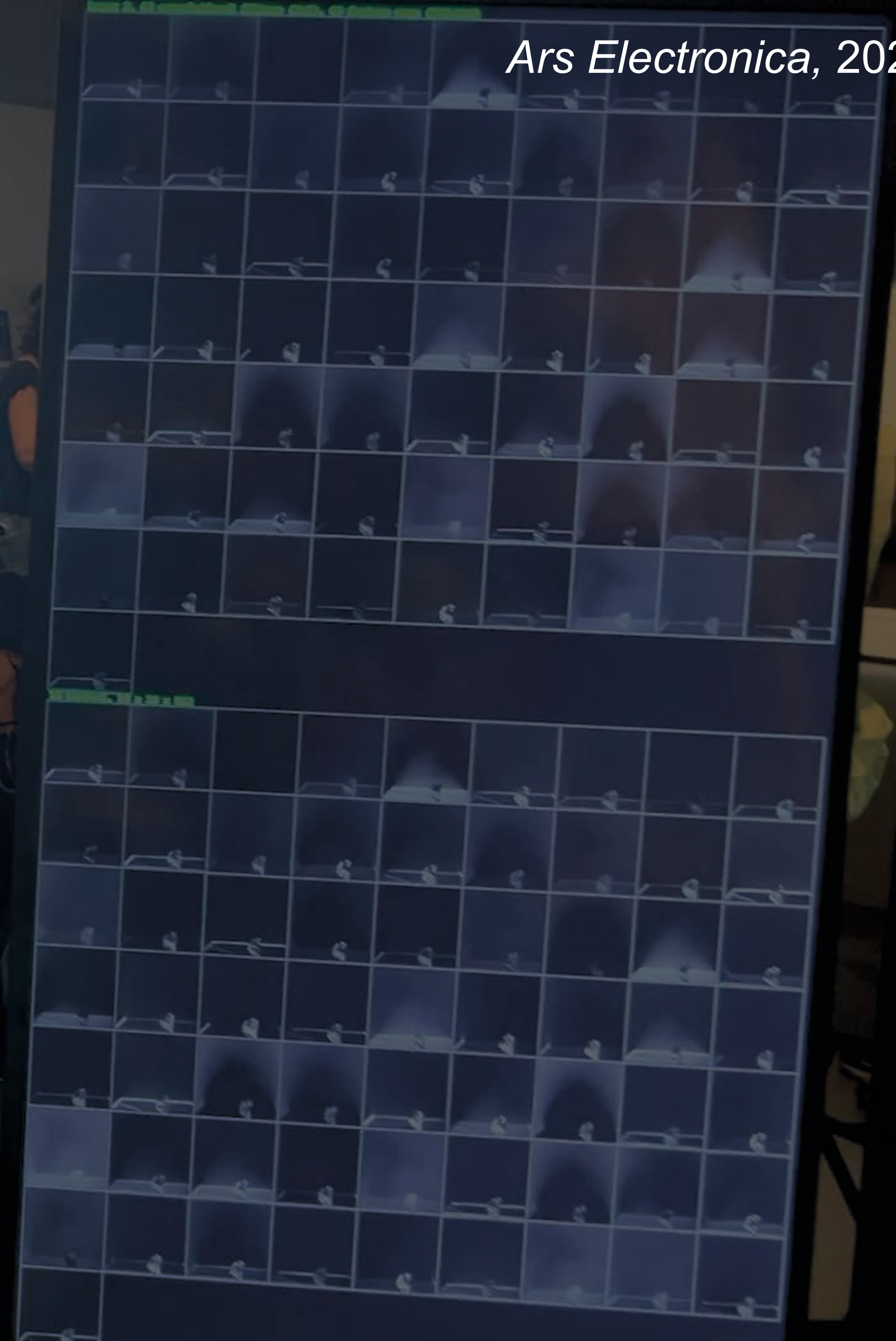
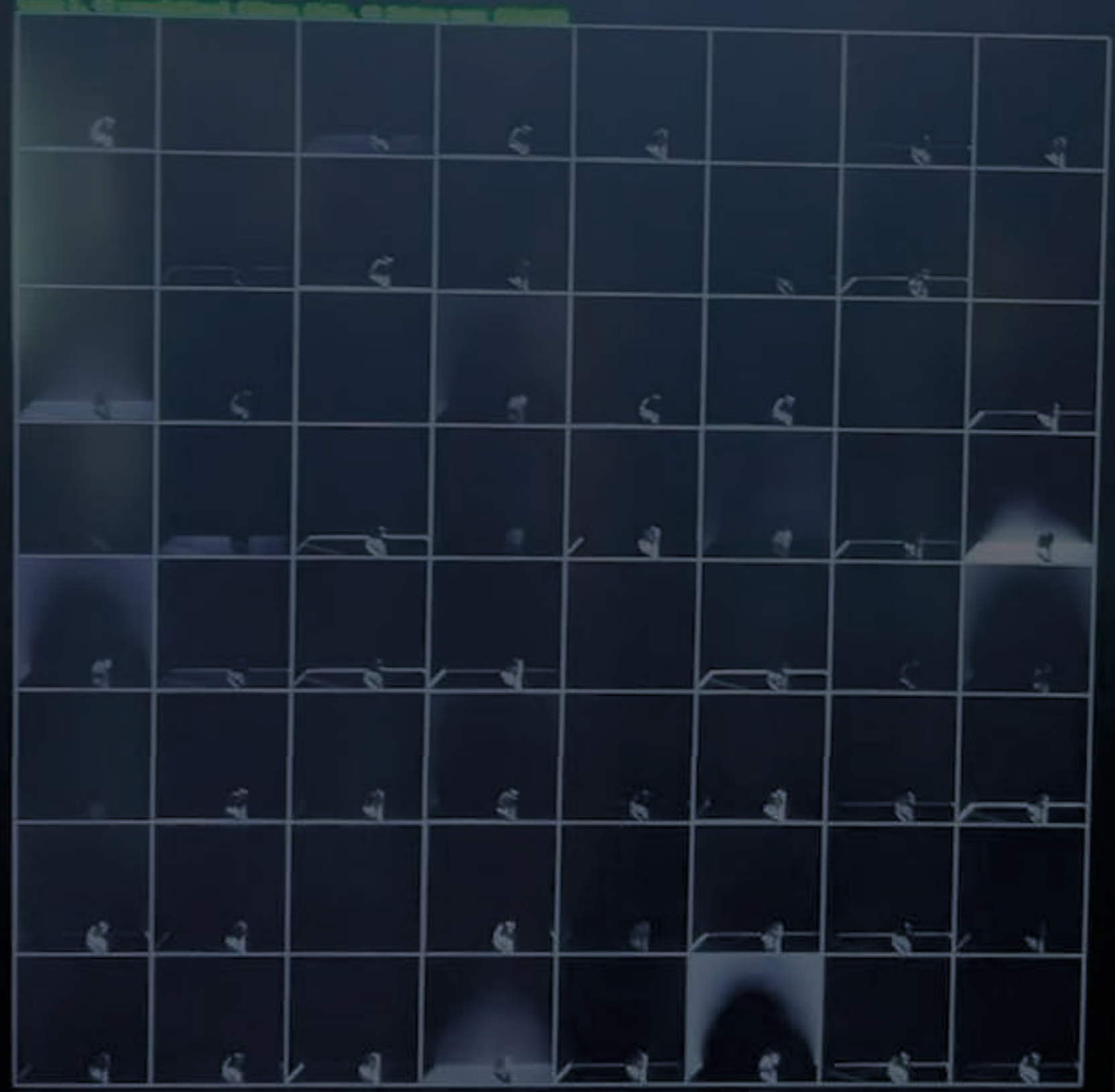
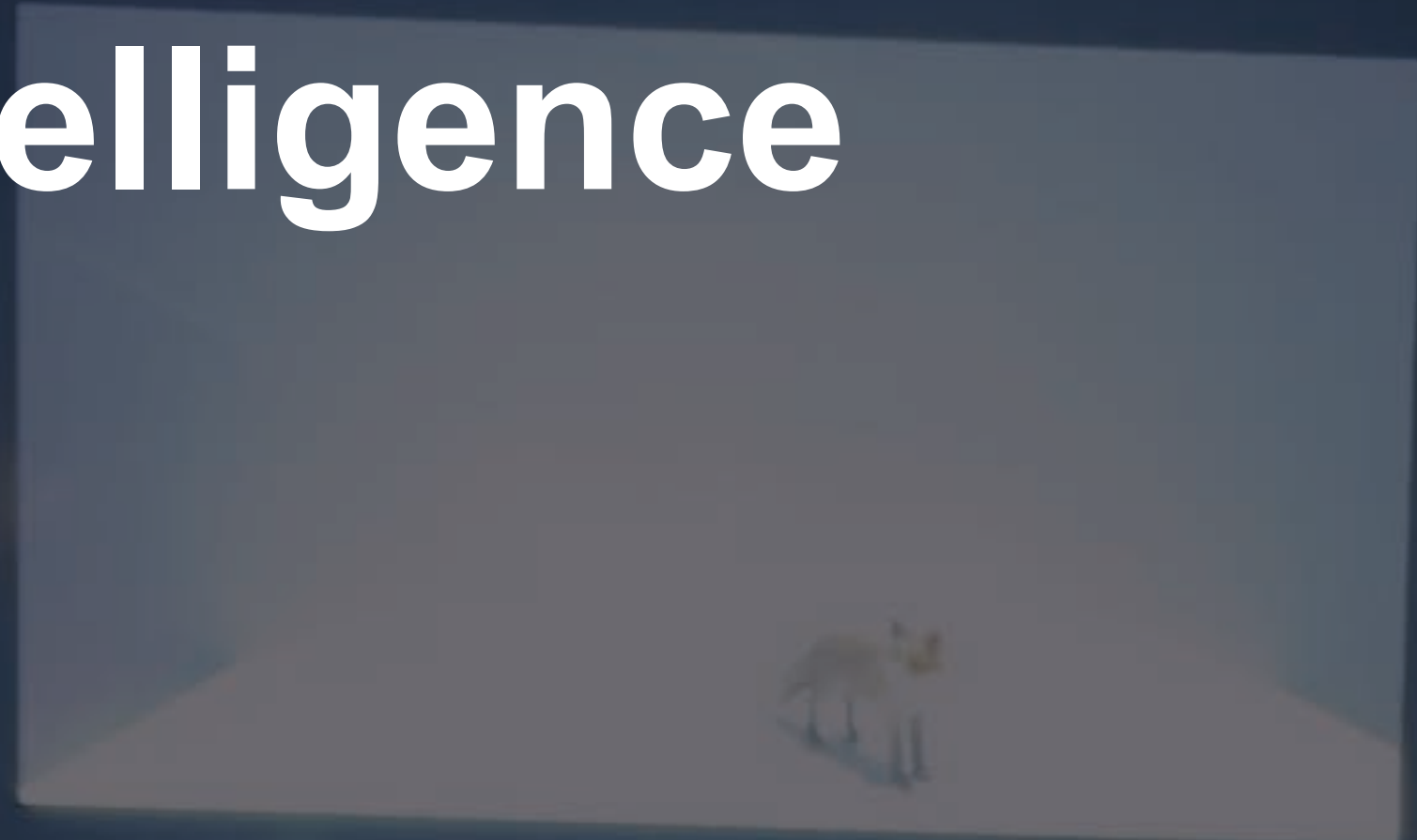
# artificial intelligence



- artificial systems that exhibit intelligent behaviour
- artificial general intelligence (AGI): hypothetical systems that reach and exceed human intelligence

# artificial intelligence

*Ars Electronica, 2024*



## COMPUTING MACHINERY AND INTELLIGENCE

By A. M. Turing

### 1. The Imitation Game

I propose to consider the question, "Can machines think?" This should begin with definitions of the meaning of the terms "machine" and "think." The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous. If the meaning of the words "machine" and "think" are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, "Can machines think?" is to be sought in a statistical survey such as a Gallup poll. But this is absurd. Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words.

The new form of the problem can be described in terms of a game which we call the 'imitation game.' It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. He knows them by labels X and Y, and at the end of the game he says either "X is A and Y is B" or "X is B and Y is A." The interrogator is allowed to put questions to A and B thus:

C: Will X please tell me the length of his or her hair?

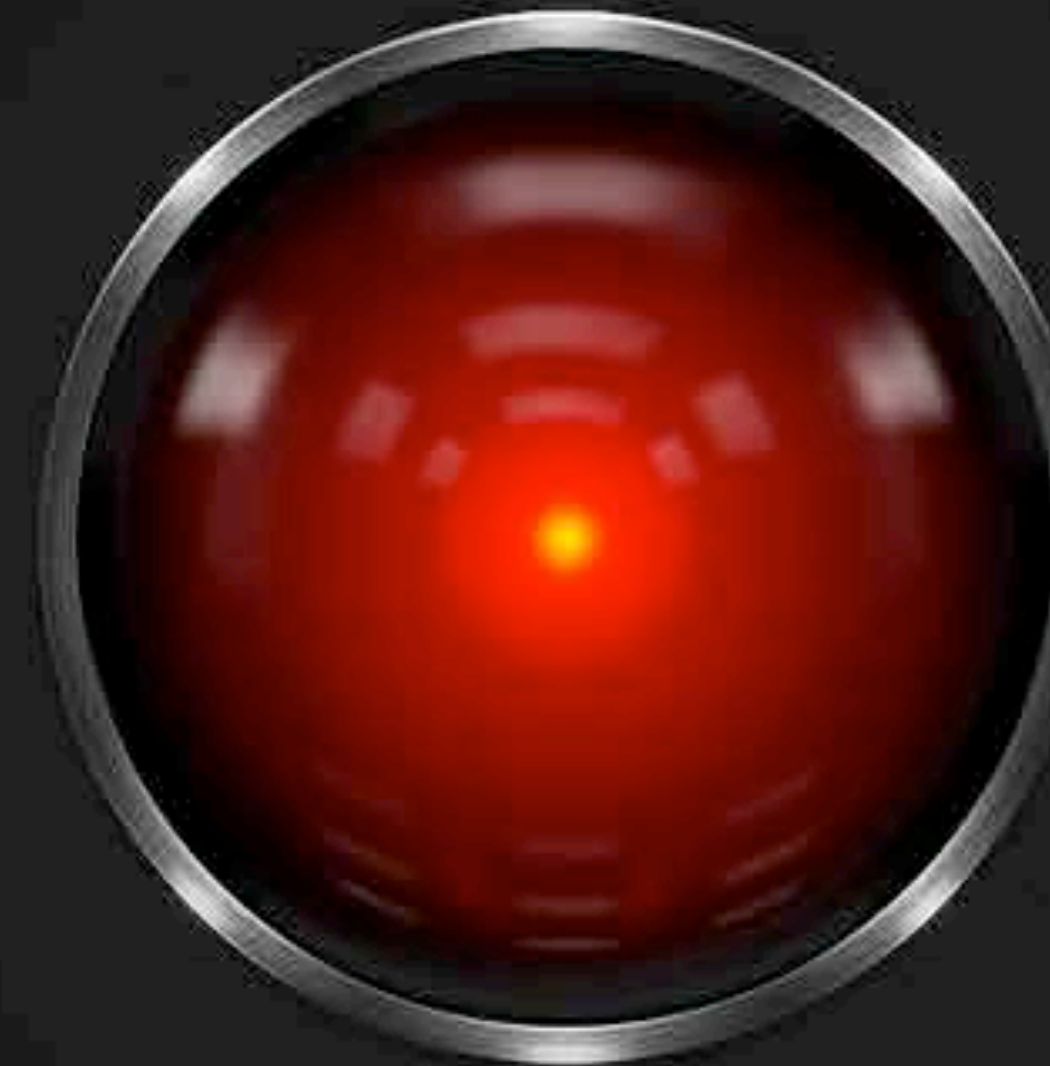
Now suppose X is actually A, then A must answer. It is A's object in the game to try and cause C to make the wrong identification. His answer might therefore be:

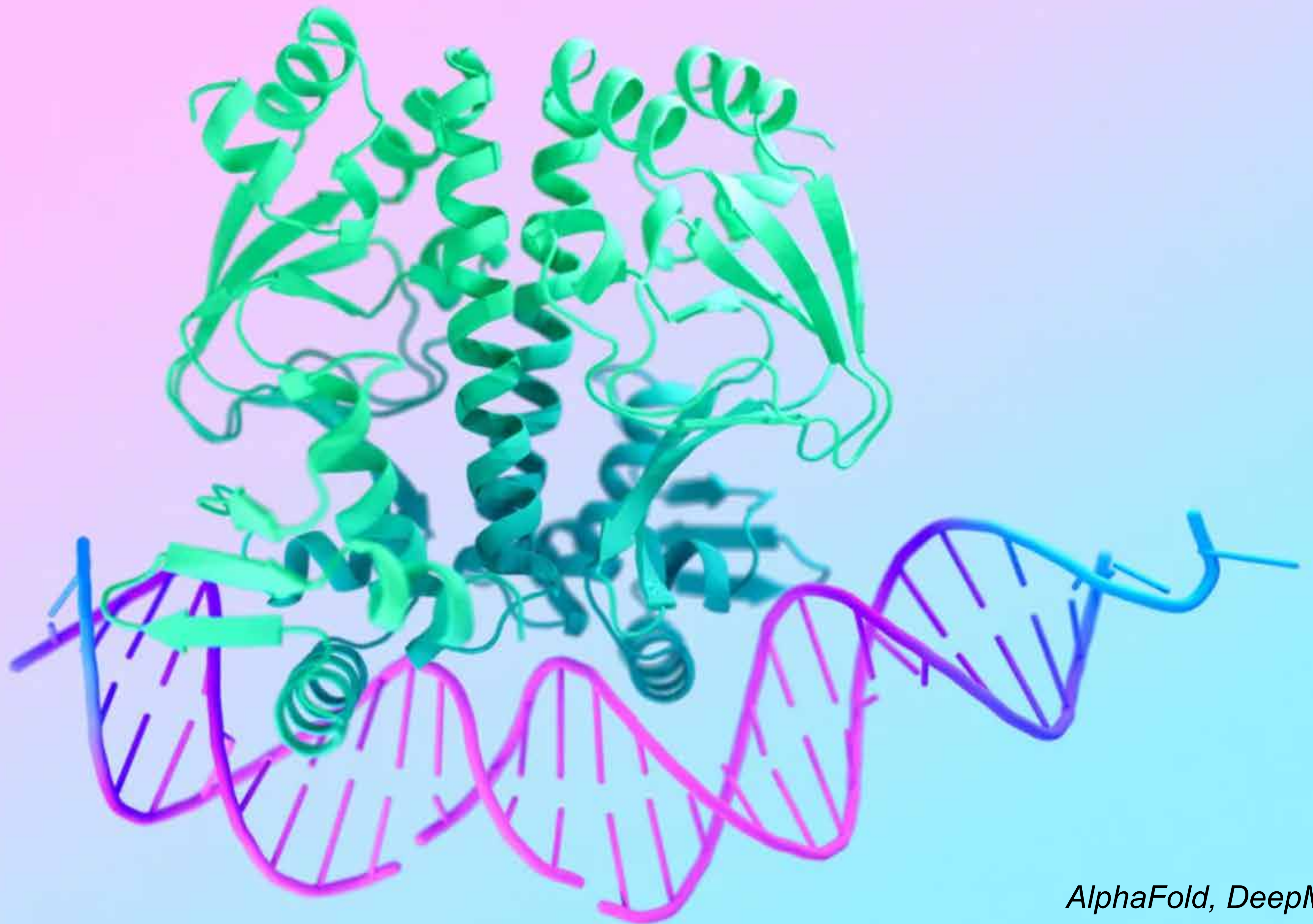
"My hair is shingled, and the longest strands are about nine inches long."

In order that tones of voice may not help the interrogator the answers should be written, or better still, typewritten. The ideal arrangement is to have a teleprinter communicating between the two rooms. Alternatively the question and answers can be repeated by an intermediary. The object of the game for the third player (B) is to help the interrogator. The best strategy for her is probably to give truthful answers. She can add such things as "I am the woman, don't listen to him!" to her answers, but it will avail nothing as the man can make similar remarks.

We now ask the question, "What will happen when a machine takes the part of A in this game?" Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman? These questions replace our original, "Can machines think?"

HAL 9000





*AlphaFold, DeepMind*



A large group of people, likely employees, are gathered for a social event. The image is overlaid with a dark, semi-transparent filter. The text "reasons to be excited" is centered in white. The people are dressed in a mix of casual and semi-formal attire, and many are holding drinks. The background shows a modern interior with a grid ceiling and recessed lighting.

**reasons to be excited**

*Google DeepMind, October 2024*

A close-up, dark, and menacing image of a Terminator robot head. The robot's face is highly detailed, showing its metallic texture and glowing red eyes. The background is a blurred, dark environment. The text "reasons to be worried" is overlaid in white, bold, sans-serif font across the center of the image.

**reasons to be worried**

# the landscape

the "singularity"



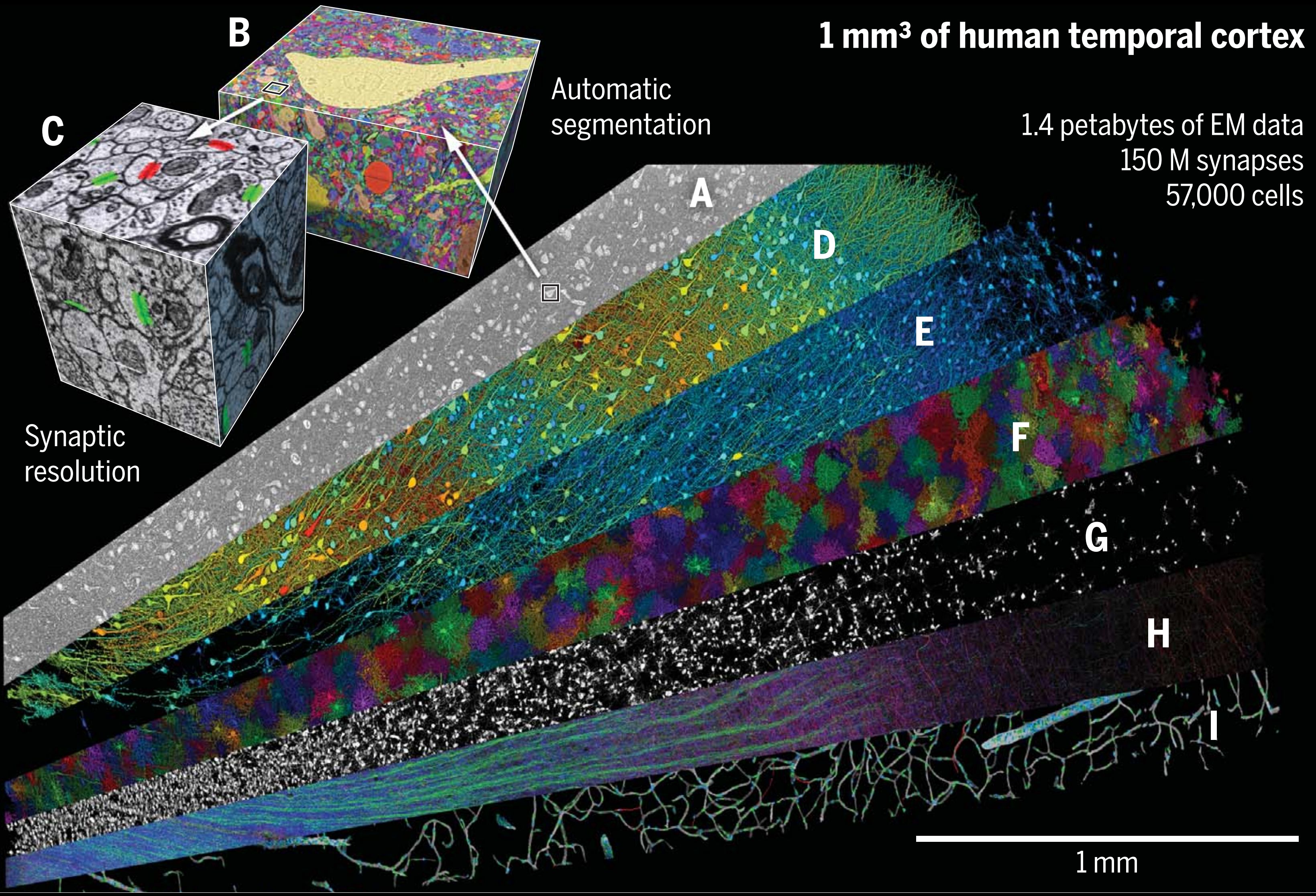
	near term / certain	long term / uncertain
good	health / biology communication/coding empowerment	superintelligent AI will save us silicon immortality
bad	energy cost social disruption privacy erosion behaviour manipulation	conscious robots will kill us all

the intelligence we have

A composite image featuring a human brain in the upper half and a computer circuit board in the lower half. The brain is shown in a light, almost ethereal glow, with its intricate folds and gyri clearly visible. Below the brain, a computer circuit board is depicted, showing various components like chips and traces. The entire image is set against a dark, gradient background that transitions from a deep blue at the top to a dark purple at the bottom. The text "brains are not computers" is overlaid in the center in a bold, white, sans-serif font.

**brains are not computers**

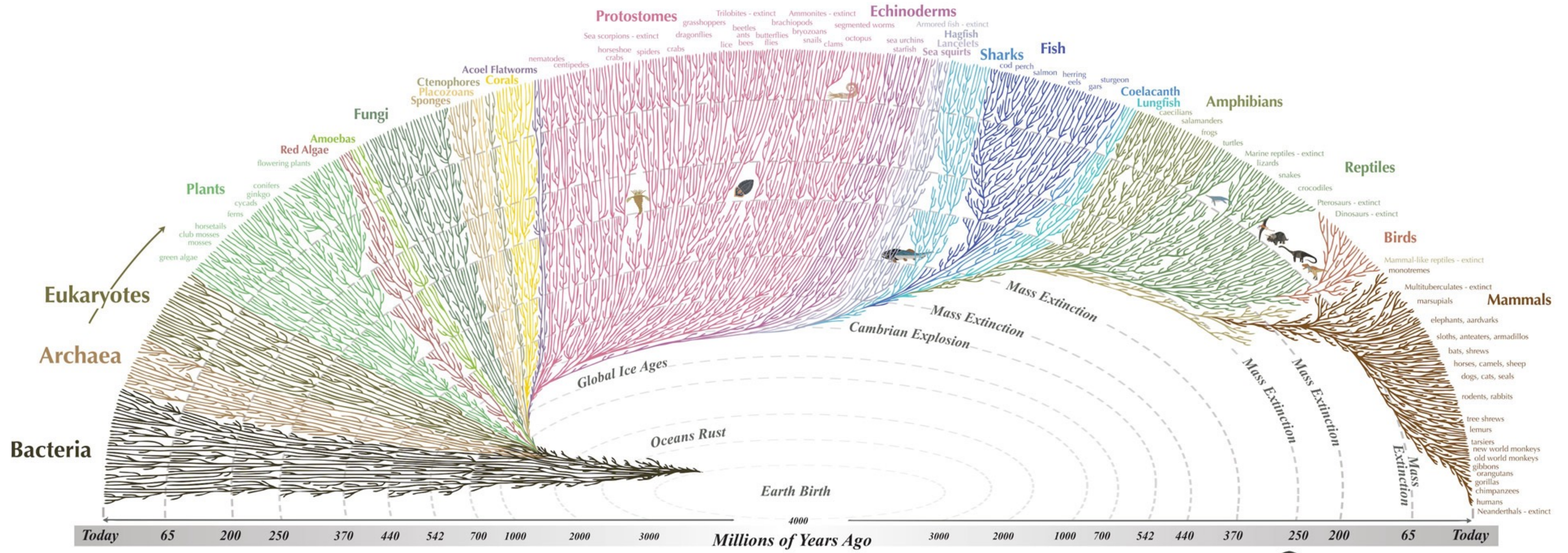
# complex




embodied



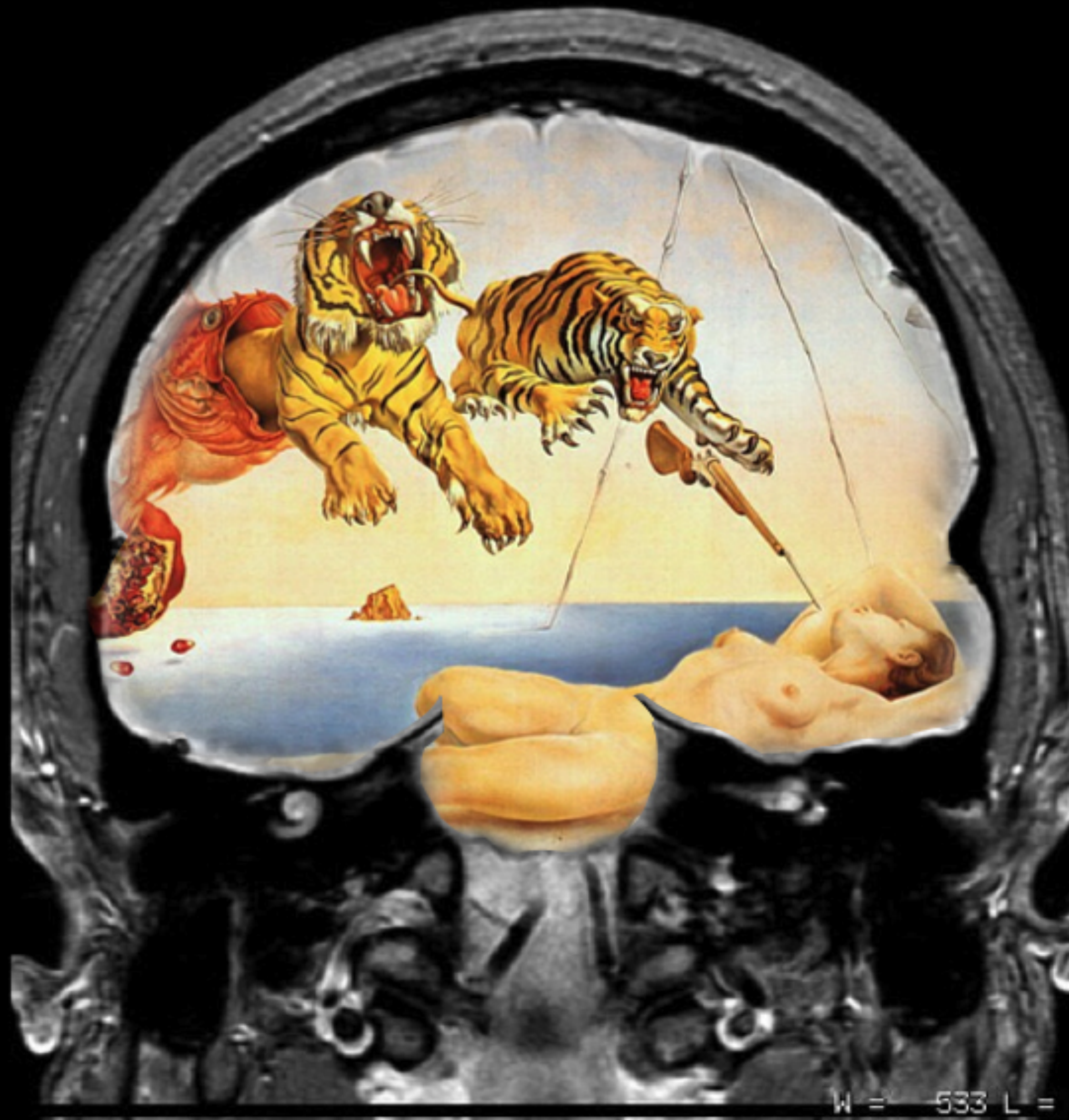
# evolved



All the major and many of the minor living branches of life are shown on this diagram, but only a few of those that have gone extinct are shown. Example: Dinosaurs - extinct 



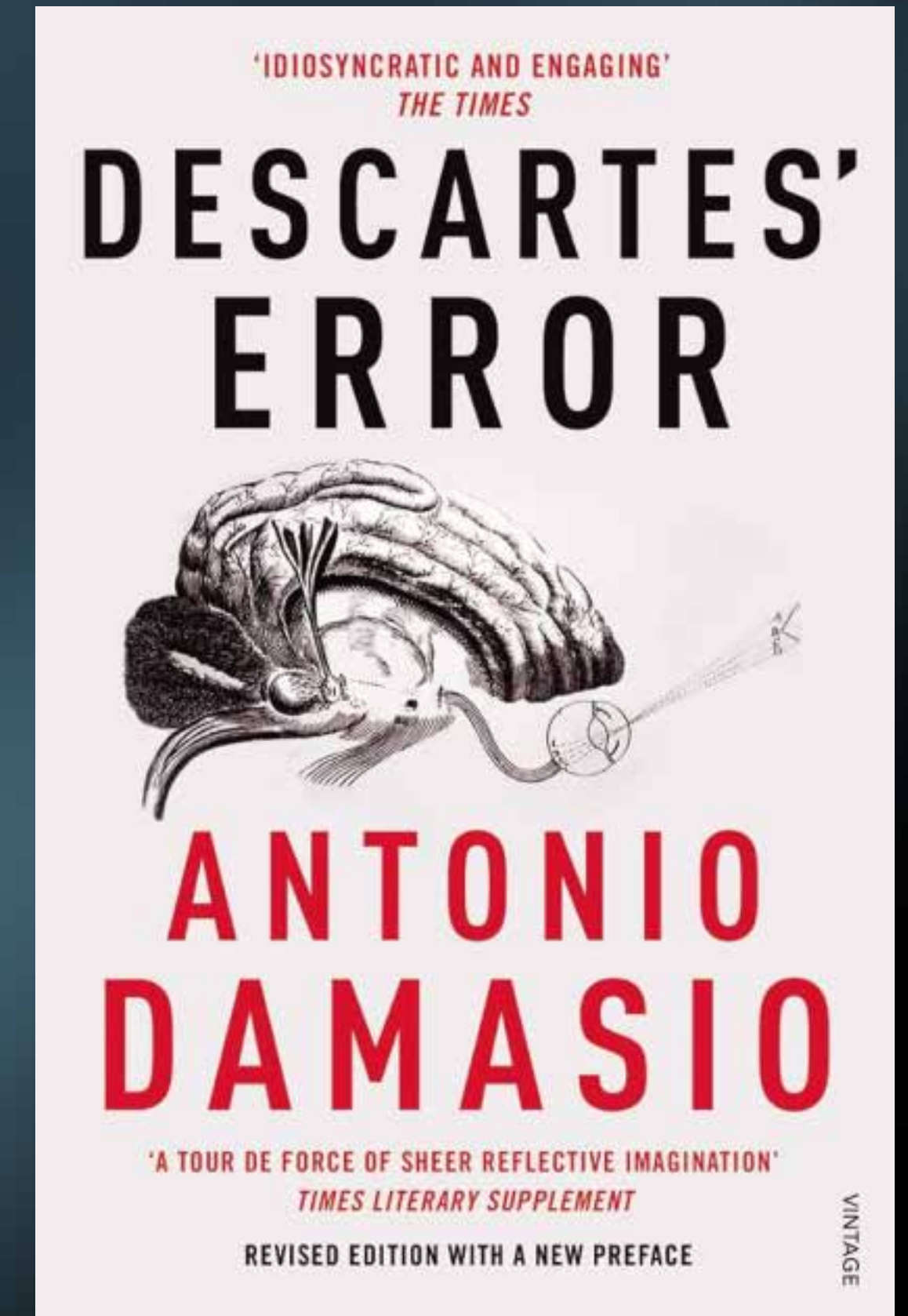
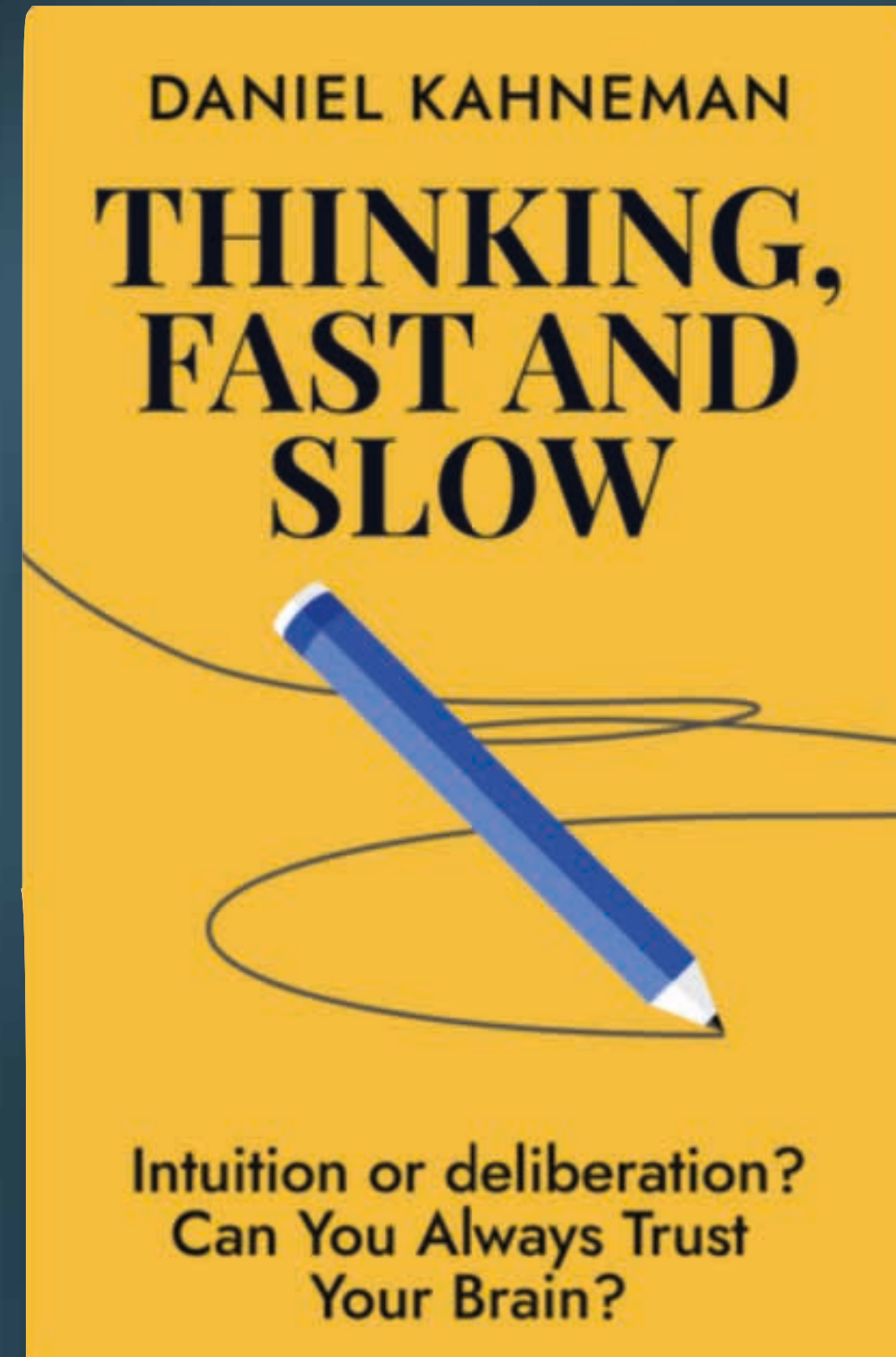
**conscious**



W = 533 L = 357

# human intelligence

- bounded cognition
- emotion and reason
- features, not bugs!



perception

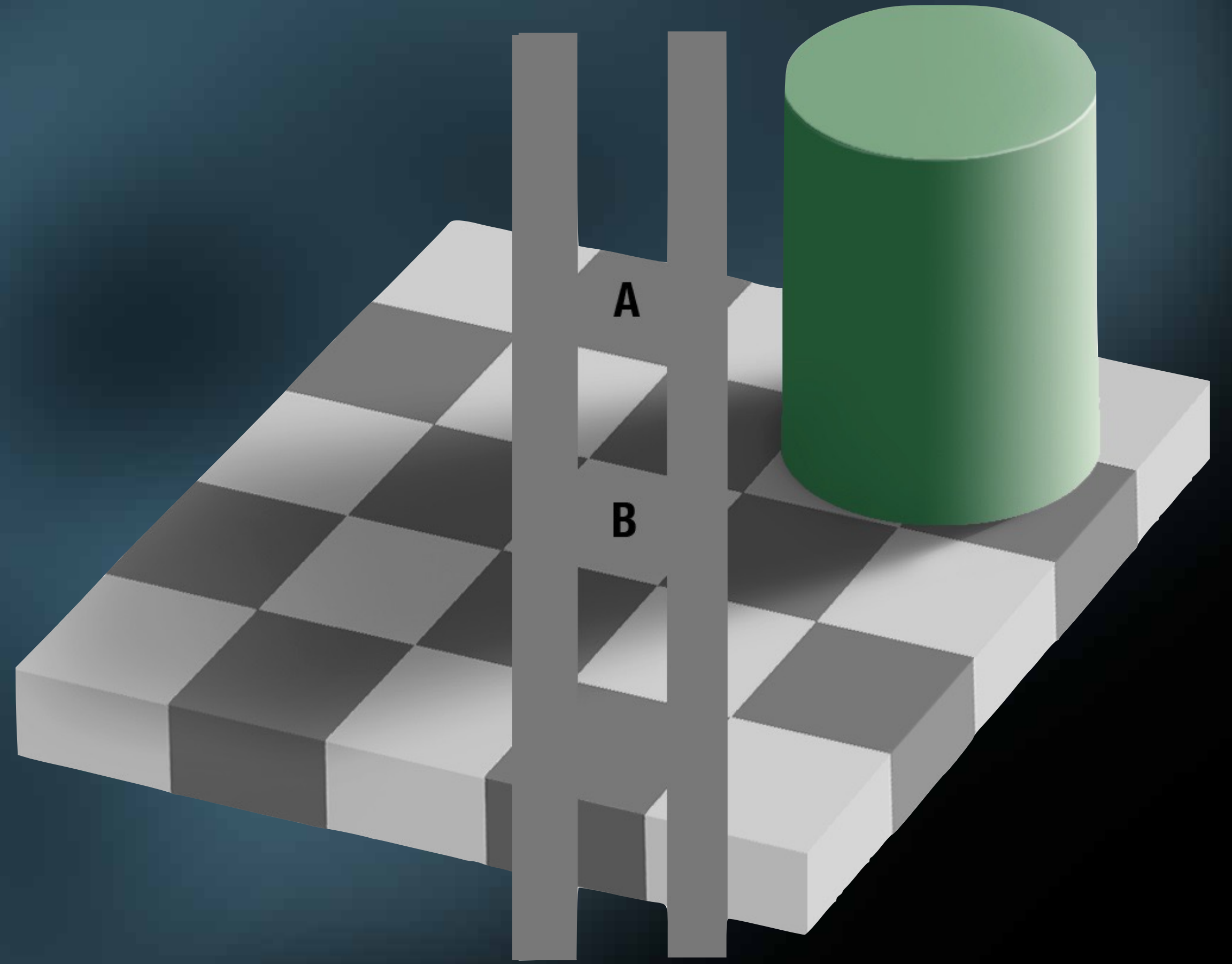
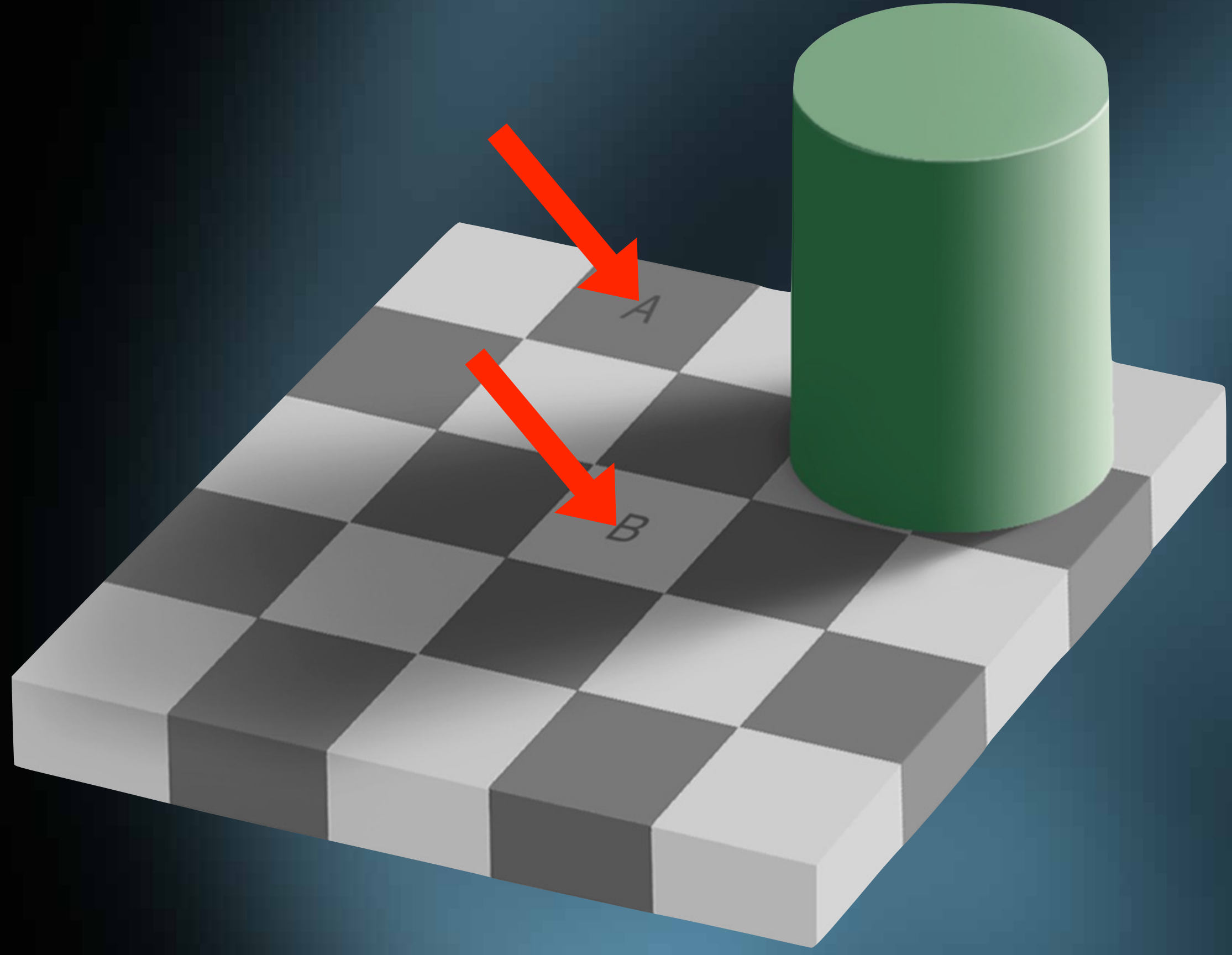


Brighton, Jan 2024

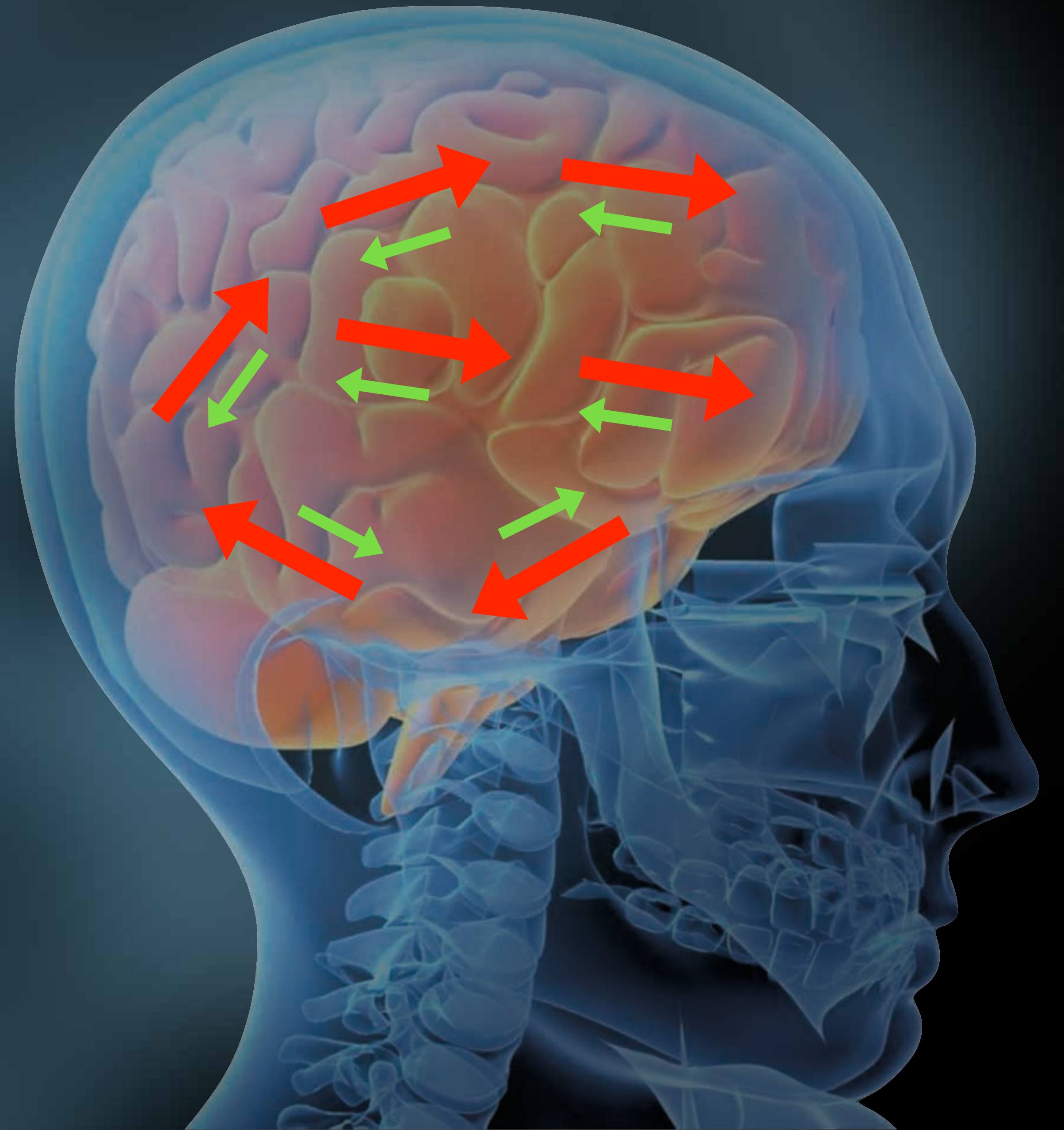


# perception as inference





controlled  
hallucination



prediction error

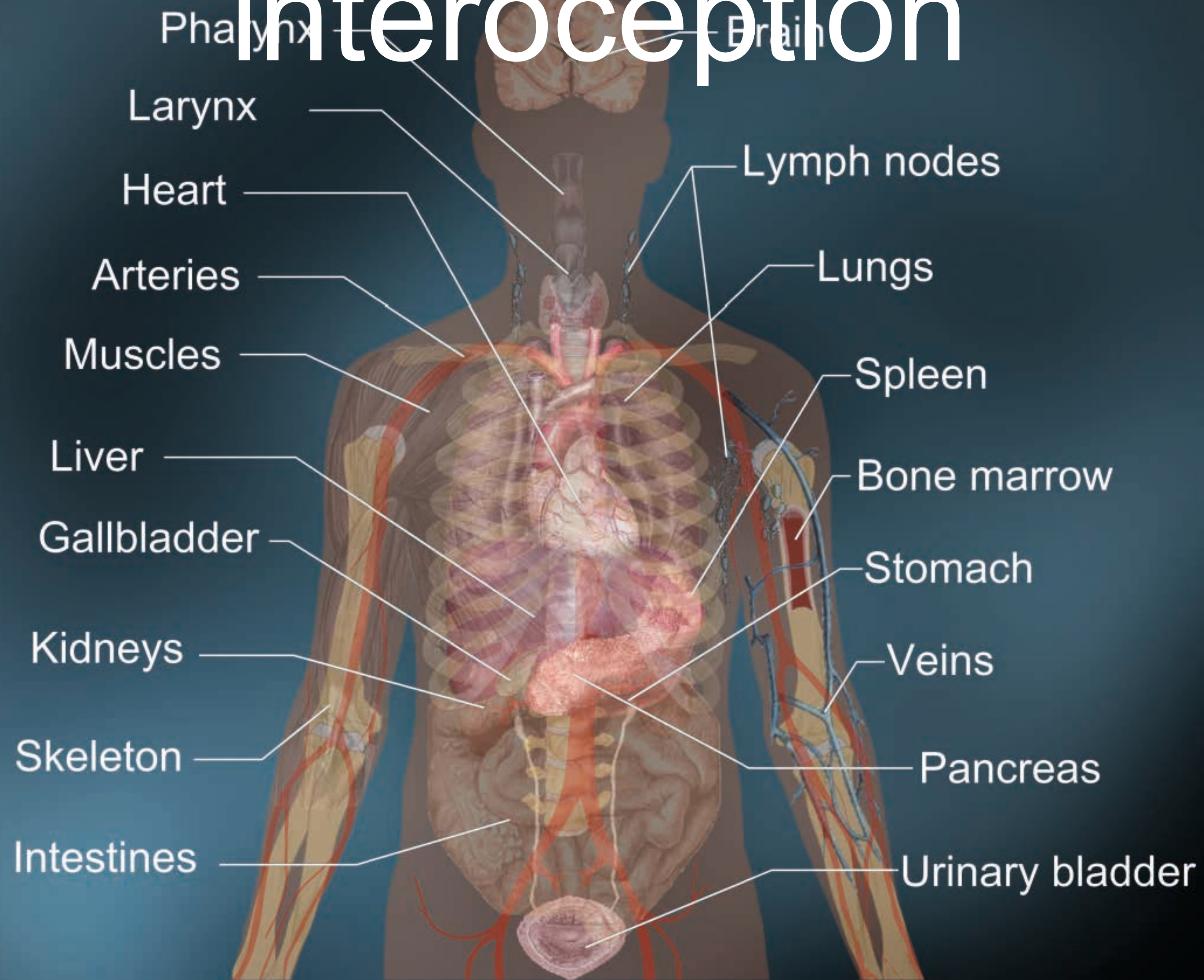


prediction





# interoception



“We perceive the world around us, and ourselves within it, with, through, and *because of* our living bodies”

Seth (2021) *Being You: A New Science of Consciousness*



consciousness

intelligence

life

clarifying the landscape

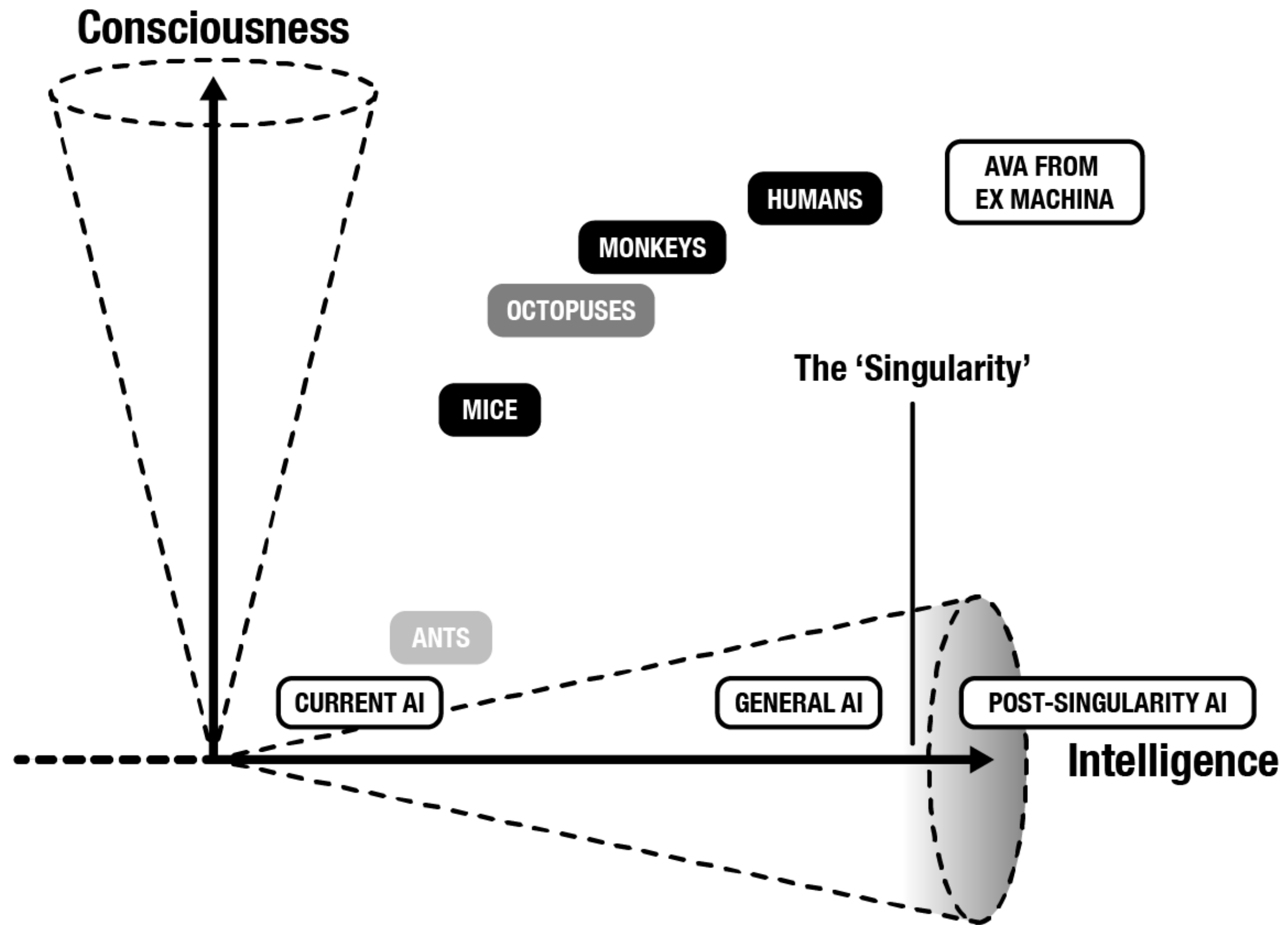
HAL 9000

artificial  
intelligence



real  
consciousness?

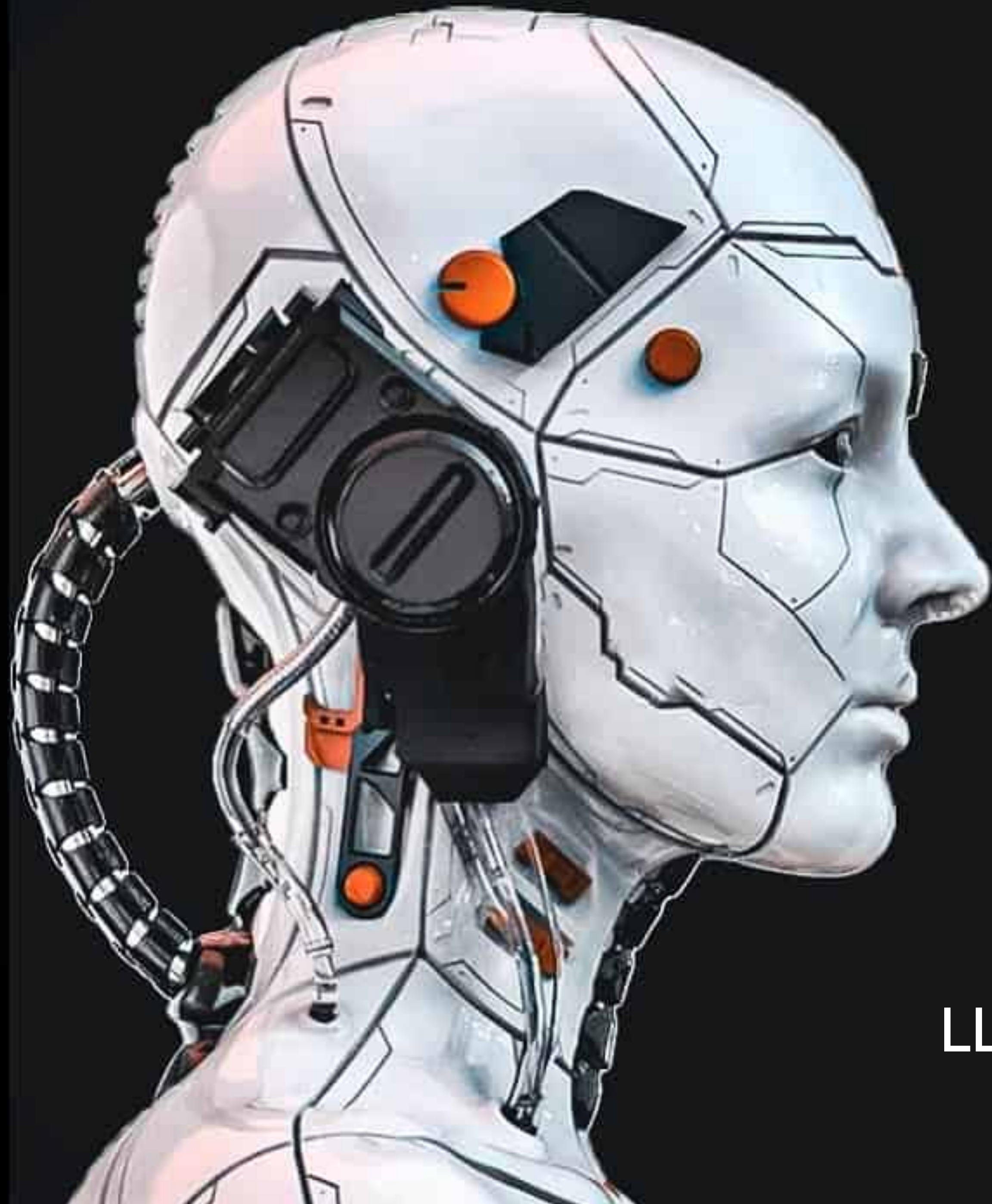




Seth (2024) *PsyArXiv*

Seth (2021) *Being You: A New Science of Consciousness*

IKEA by Spike Jonze, h/t Henry Shevlin



 OpenAI



**GPT-4**

LLMs don't **hallucinate**, they **confabulate**

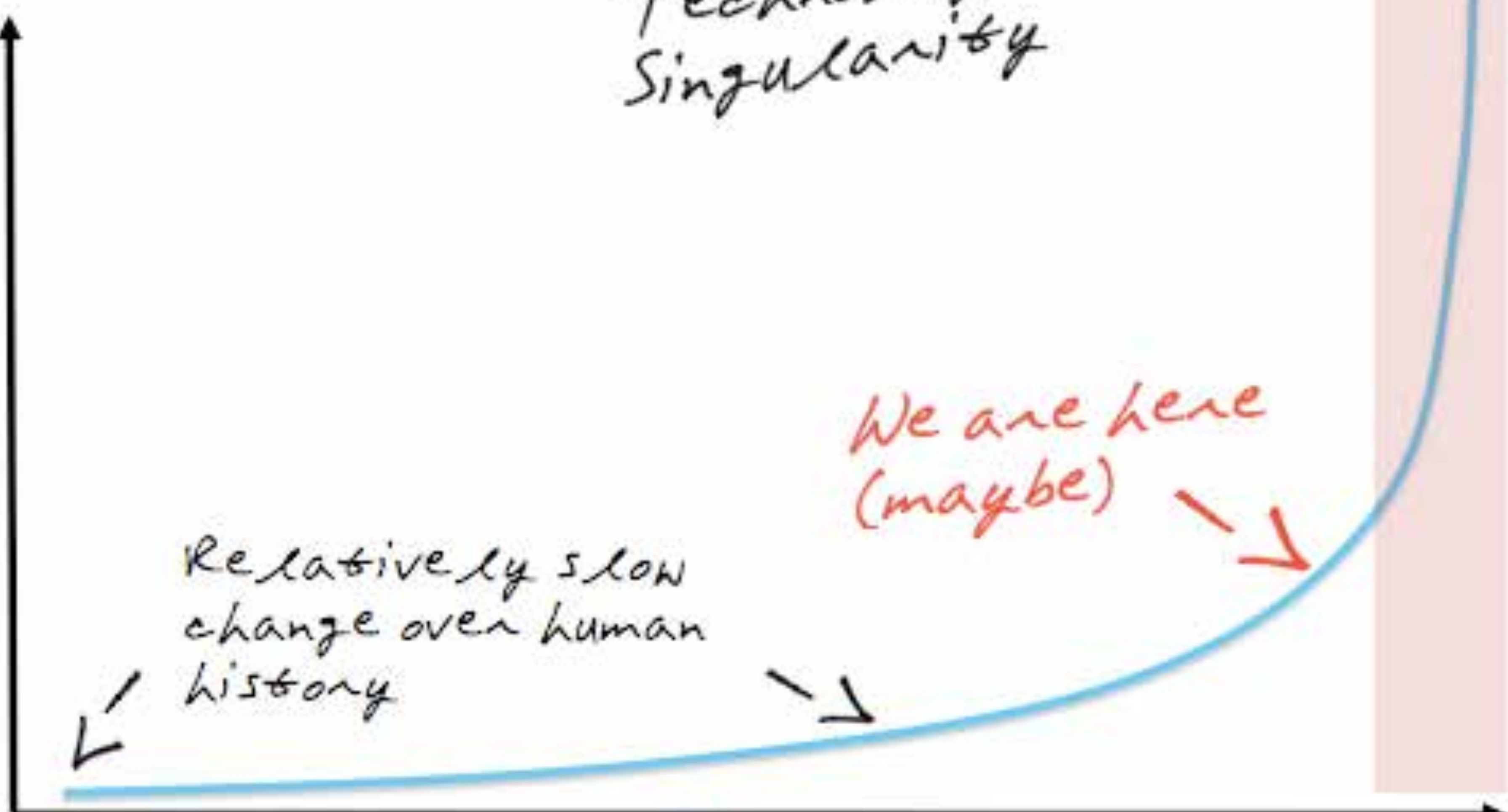


“Technology can make us forget  
what we know about life.”

Sherry Turkle



Level of Advancement



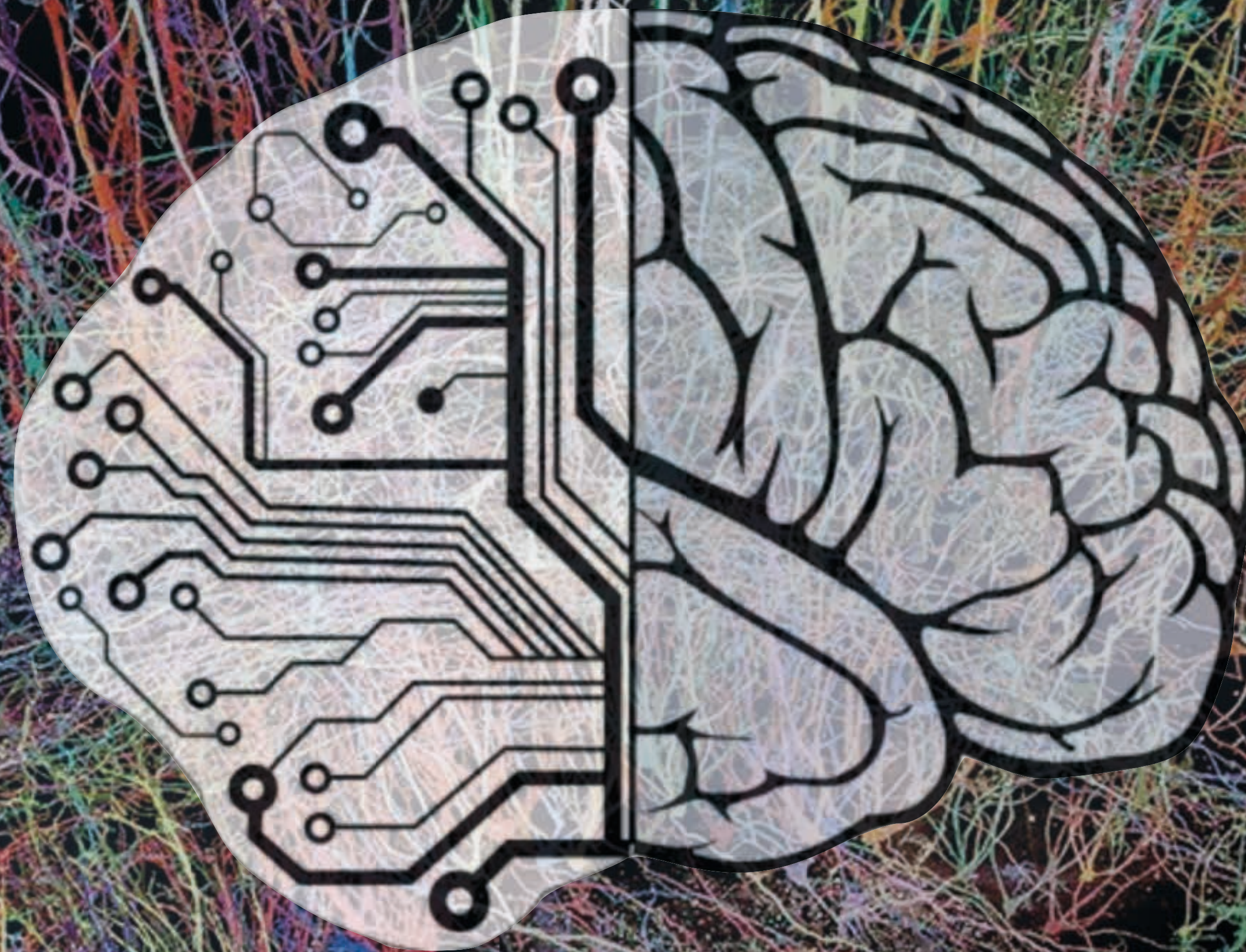
Time

Rapid change in the  
Technological Singularity →

Relatively slow  
change over human  
history

We are here  
(maybe)

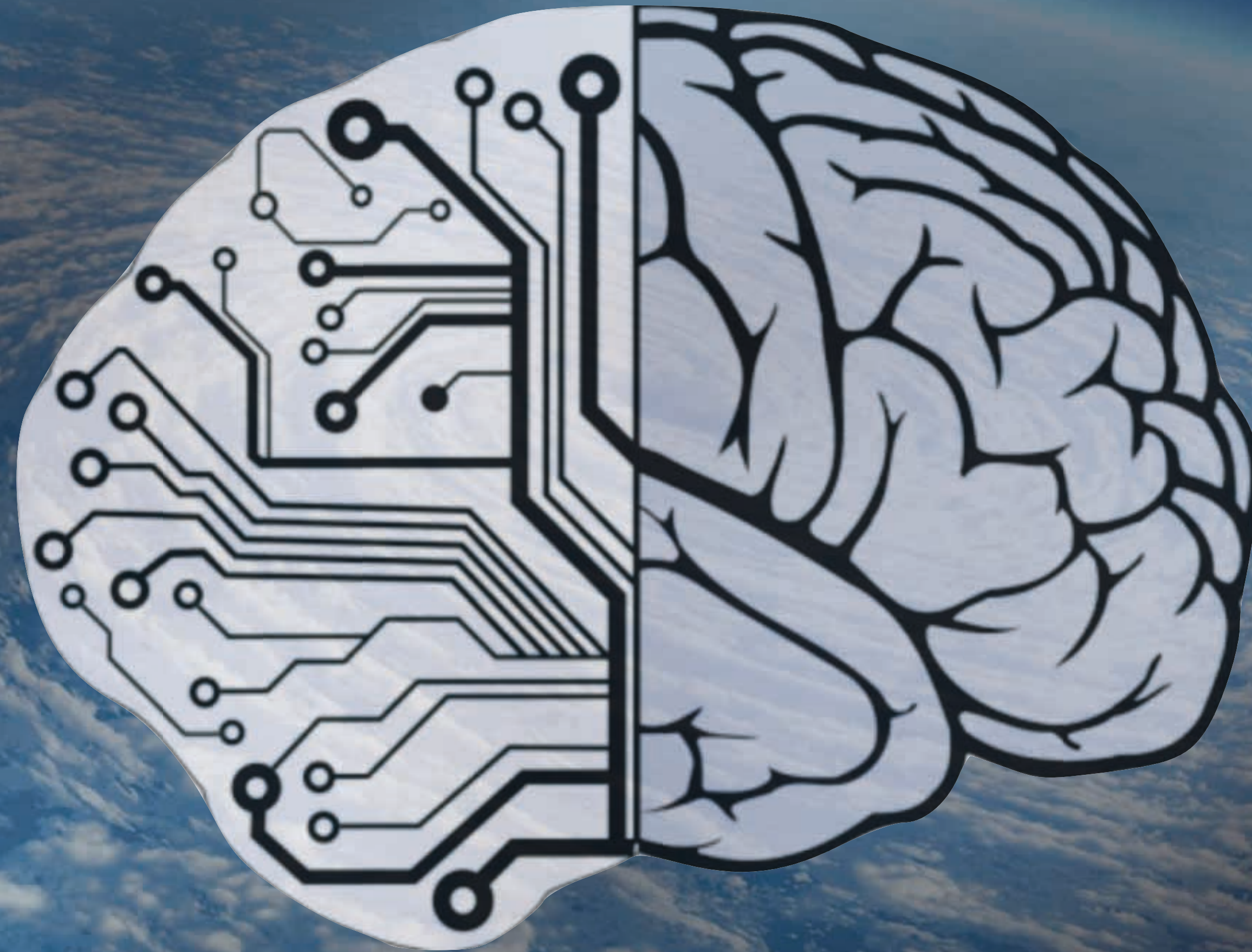
# wetware, not hardware



Seth (2024) *PsyArXiv*

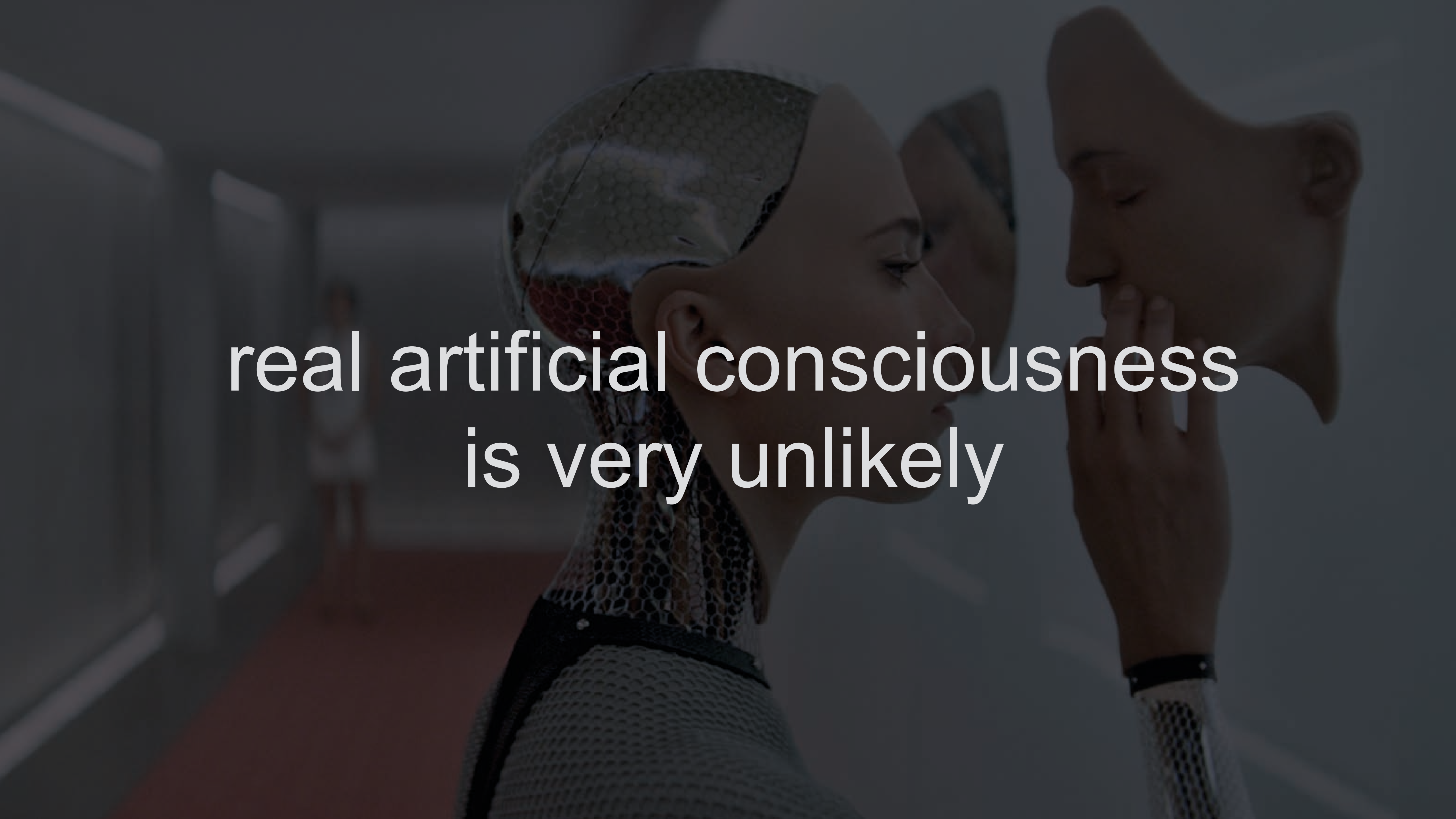
Seth (2021) *Being You: A New Science of Consciousness*

# simulation vs instantiation



Seth (2024) *PsyArXiv*

Seth (2021) *Being You: A New Science of Consciousness*

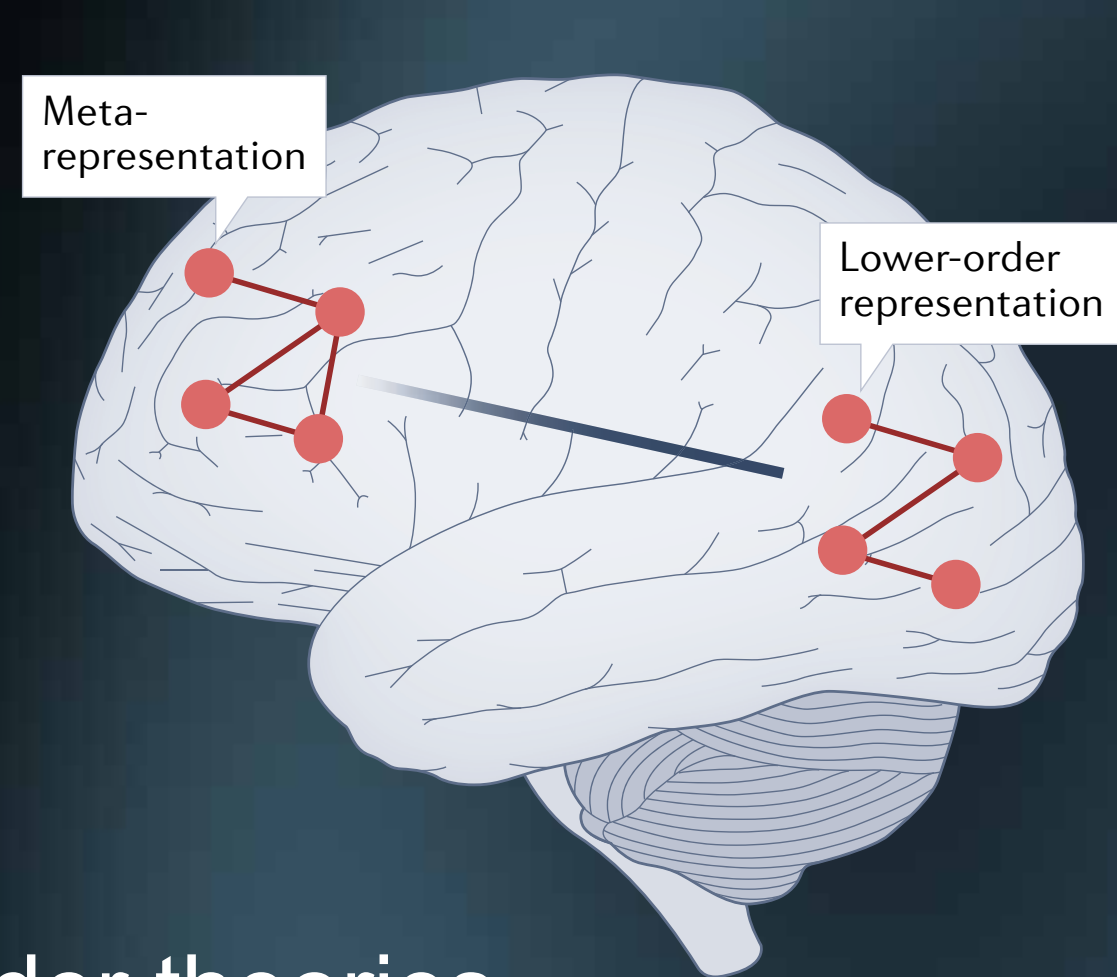
A woman in a futuristic, mesh-like outfit is shown in profile, looking towards a man in a similar outfit. The man is also in profile, looking back at the woman. The background is dark and out of focus, suggesting an indoor setting with some architectural elements. The overall mood is contemplative and futuristic.

real artificial consciousness  
is very unlikely

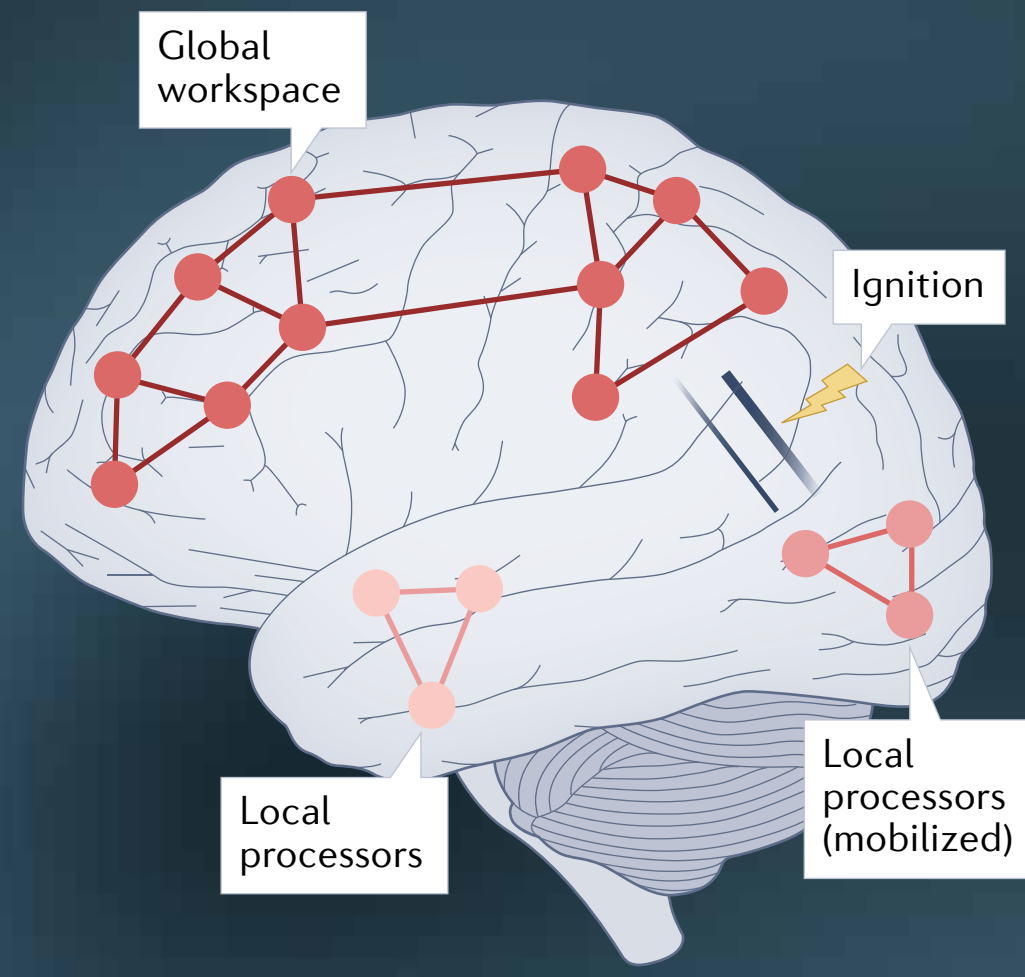
why does this matter?

I might be wrong

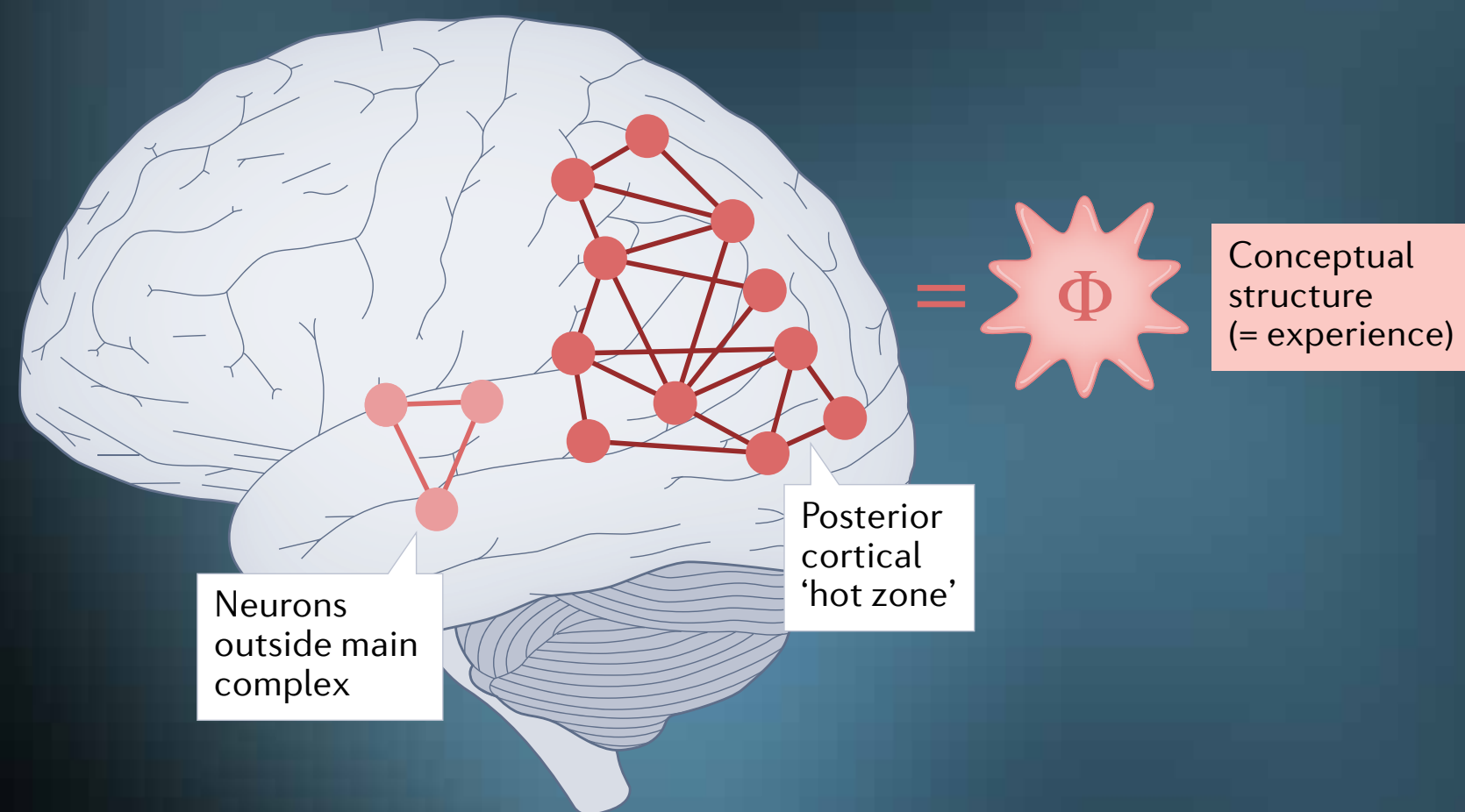
# theories of consciousness



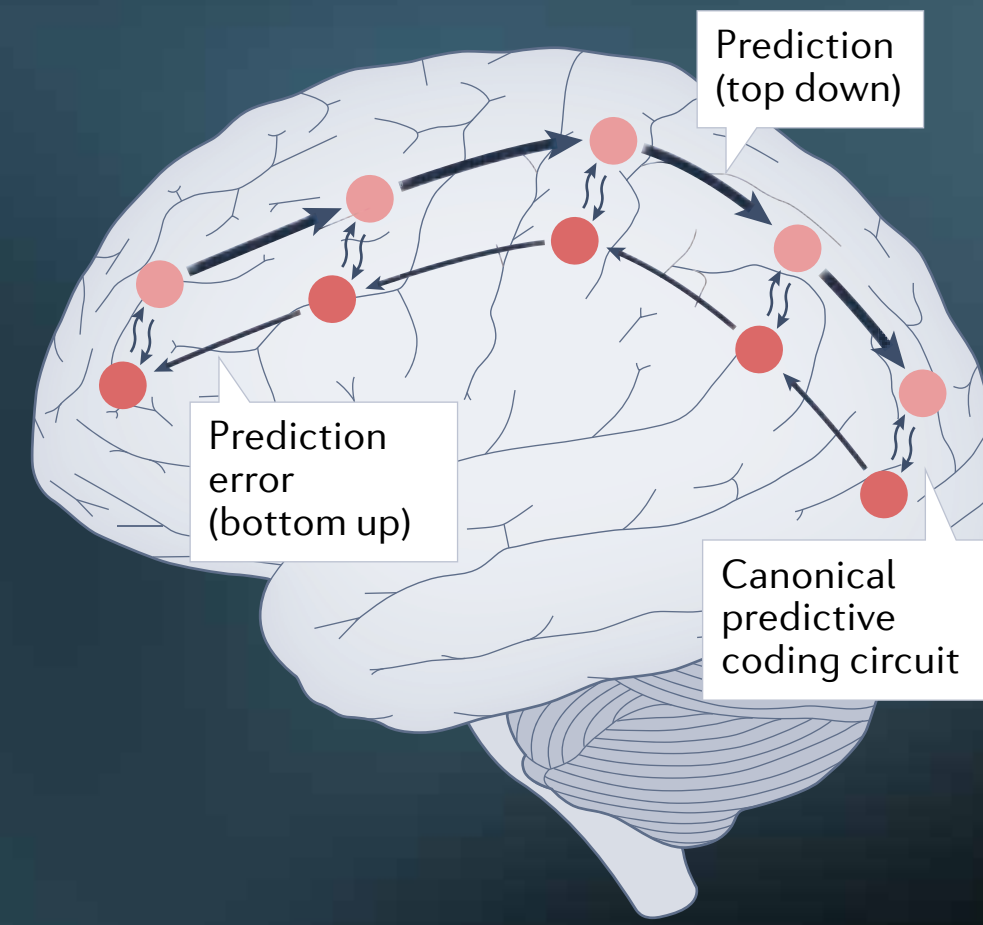
Higher-order theories



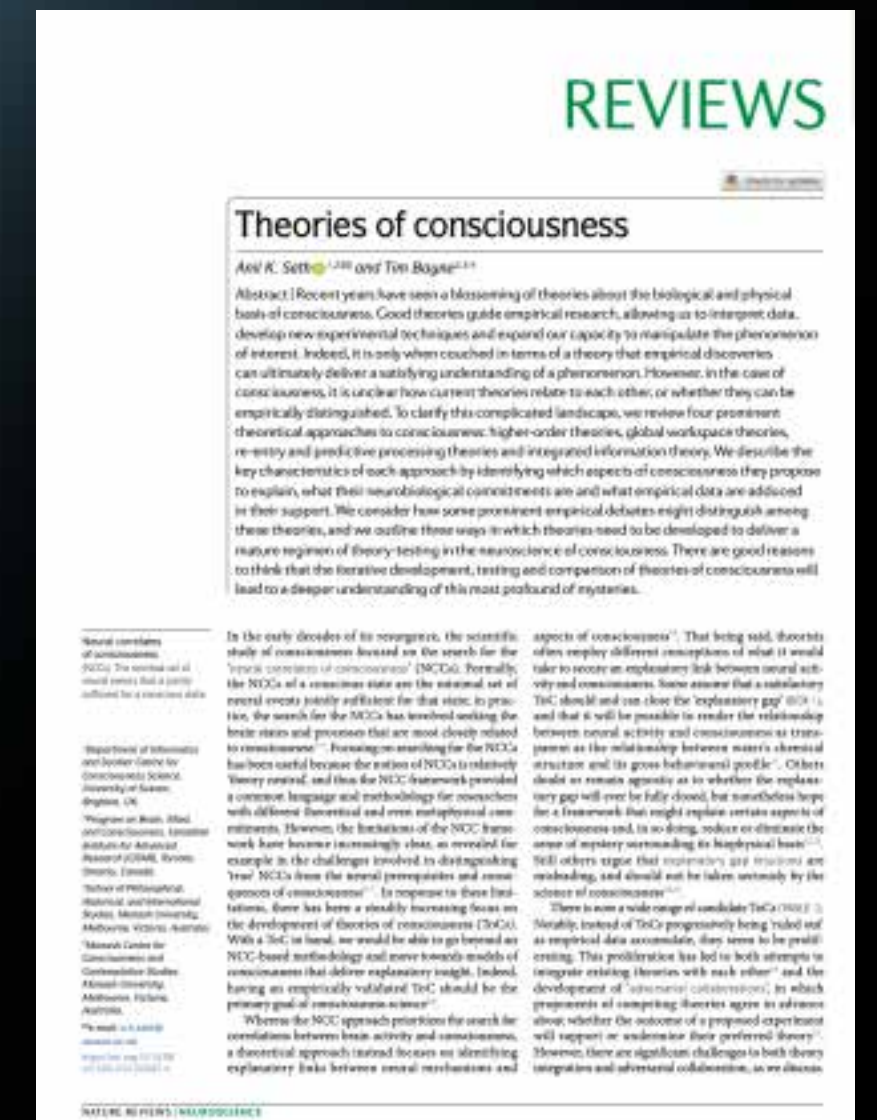
Global workspace theories



Integrated information theories



Predictive processing / reentry theories





# reasons to not build real conscious AI

- value alignment
- real artificial suffering

Seth (2024) *PsyArXiv*

Seth (2021) *Being You: A New Science of Consciousness*

Seth (2023) *Nautilus*

even AI that *seems* conscious  
could be very bad

# conscious-seeming AI

The Washington Post  
*Democracy Dies in Darkness*

TECHNOLOGY

## The Google engineer who thinks the company's AI has come to life

AI ethicists warned Google not to impersonate humans. Now one of Google's own thinks there's a ghost in the machine.

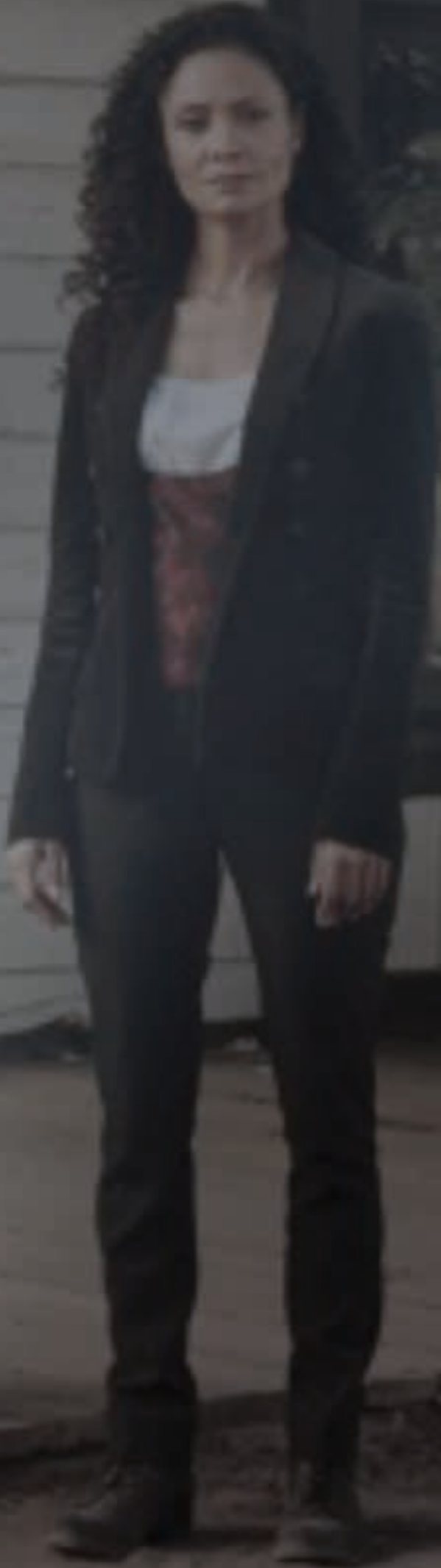


By Nitasha Tiku

June 11, 2022 at 8:00 a.m. EDT



Google engineer Blake Lemoine. (Martin Klimek/For the Washington Post)



# conscious-seeming AI

- potential for **cognitively impenetrable** illusions of consciousness



# conscious-seeming AI

- intentional stance / misprediction
- social coercion
- brutalism (the West World problem)

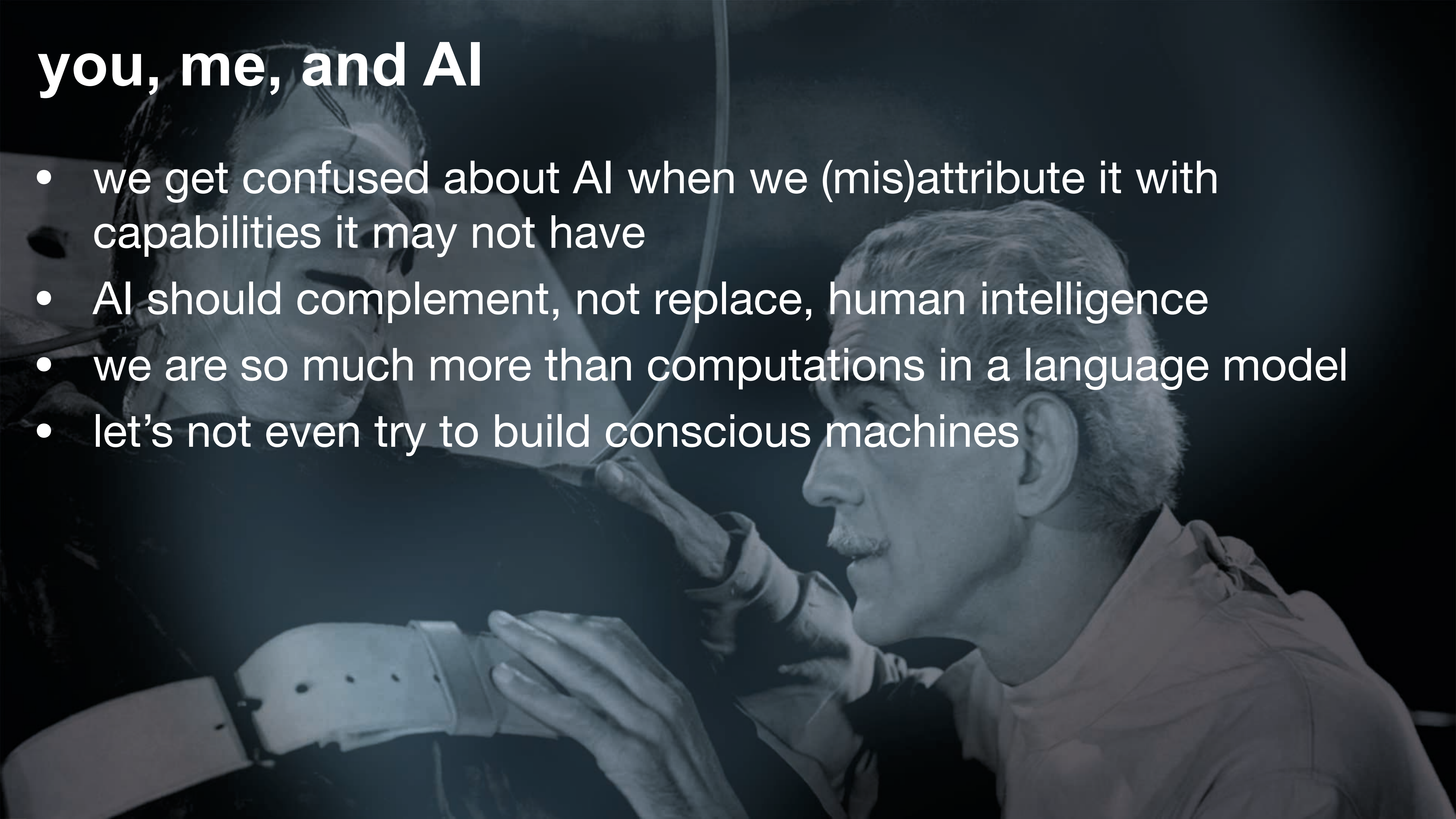


Seth (2024) *PsyArXiv*

Seth (2023) *Nautilus*

the intelligence we need

# you, me, and AI



- we get confused about AI when we (mis)attribute it with capabilities it may not have
- AI should complement, not replace, human intelligence
- we are so much more than computations in a language model
- let's not even try to build conscious machines

# summary

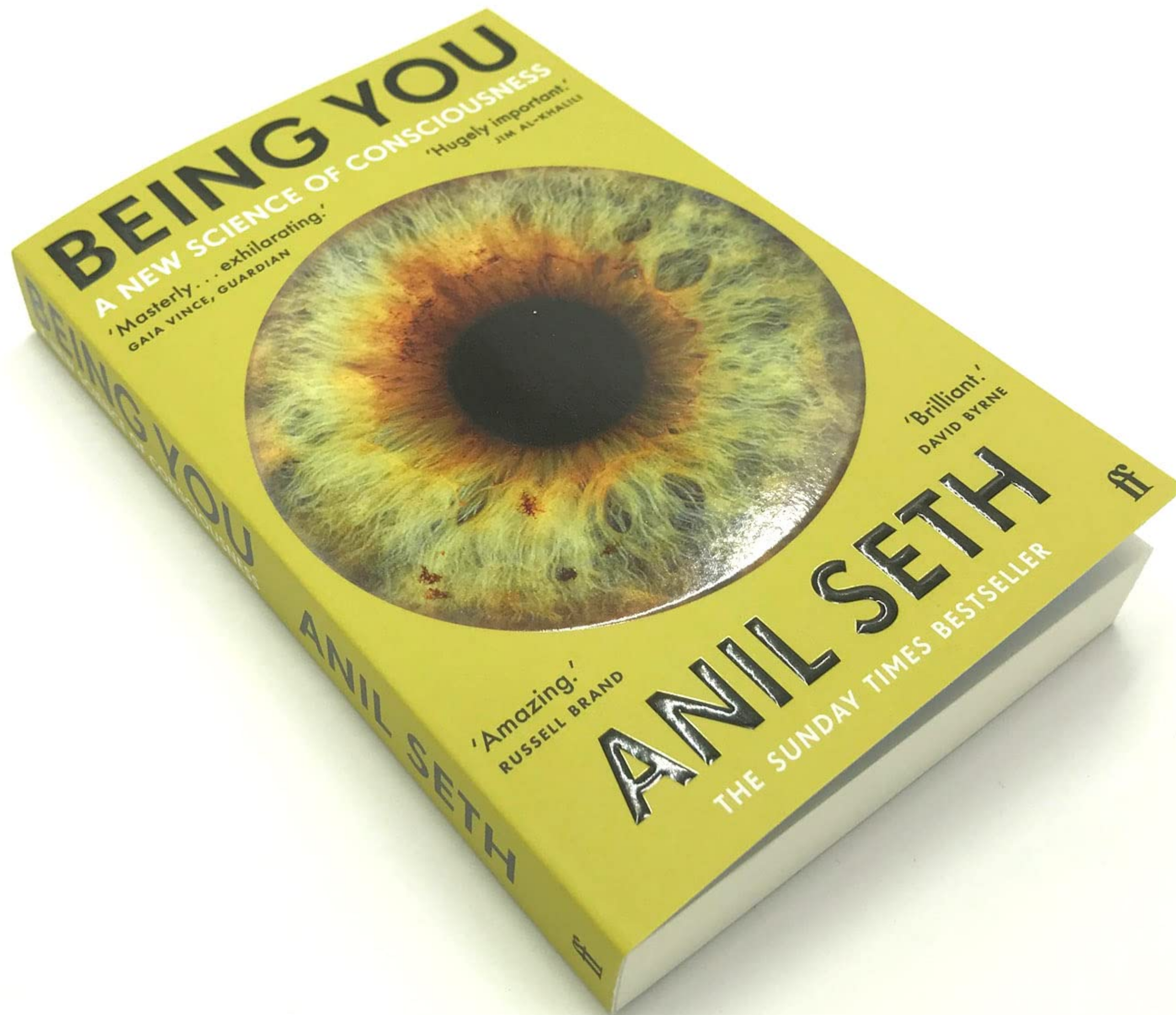
- human intelligence is embodied, evolved, bounded, and continuous with perception
- don't confuse the (certain) near-term challenges/opportunities with the (uncertain) long-term
- understand our psychological biases
- AI can be tools - even partners - but they should not replace us





“We should treat AI as tools rather than colleagues, and always remember the difference.”

Daniel Dennett  
1942-2024



**The *Sunday Times* Top 10 Bestseller**

*A Financial Times* Book of the Year

*A Guardian* Science Book of the Year

*An Economist* Book of the Year

*A New Statesman* Book of the Year

*A Bloomberg* Book of the Year

“An exhilarating book: A vast-ranging phenomenal achievement that will undoubtedly become a seminal text”

*Gaia Vince, Guardian (Book of the Week)*

“A brilliant beast of a book”

*David Byrne*

[www.anilseth.com](http://www.anilseth.com)

@anilkseth



offcuts

“Technology can make us forget  
what we know about life.”

Sherry Turkle



# artificial intelligence

*Ars Electronica, 2024*

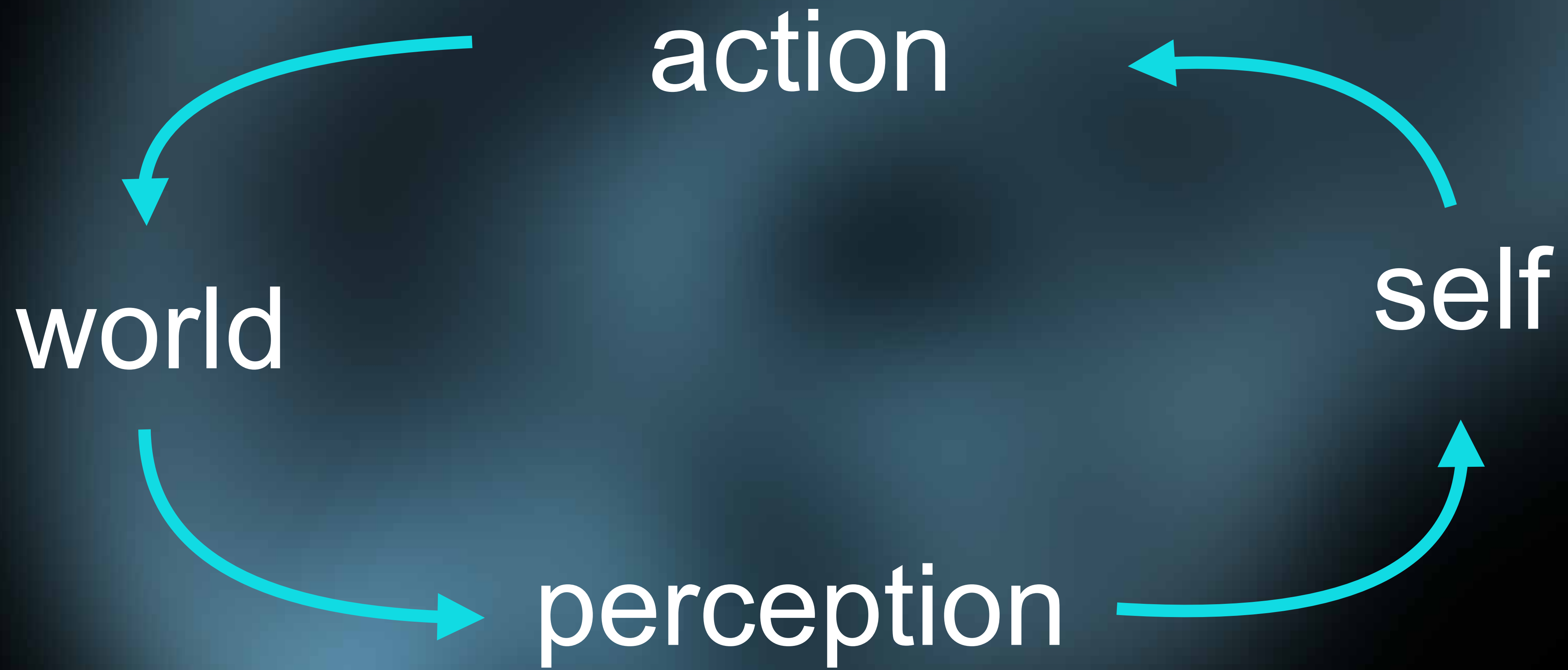


# cerebral organoids

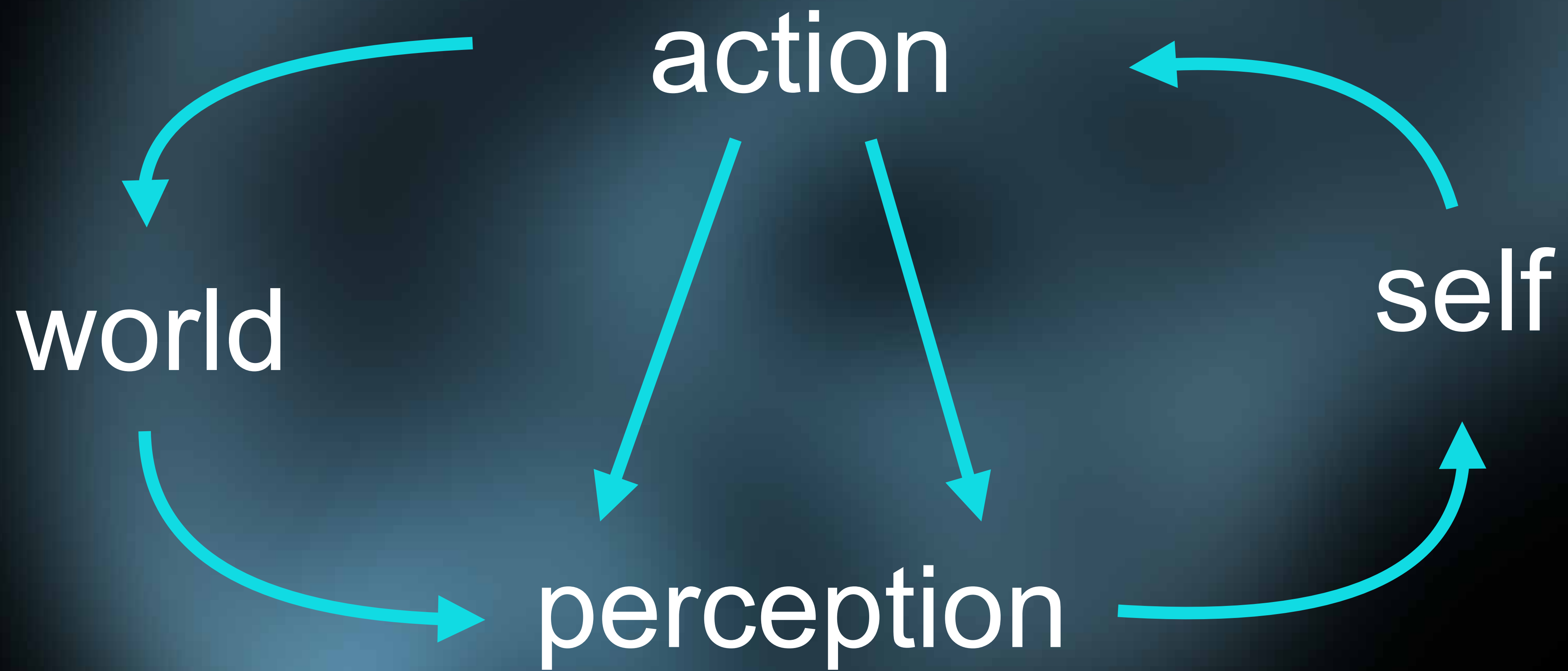
A photograph showing several cerebral organoids, which are small, spherical, light-brown structures, growing on a dark blue surface. The organoids are scattered across the surface, with some appearing as small, isolated spheres and others as larger, more complex structures. The background is a dark blue, textured surface, likely a petri dish or a specialized culture medium.

Trujillo et al (2019) *Cell*

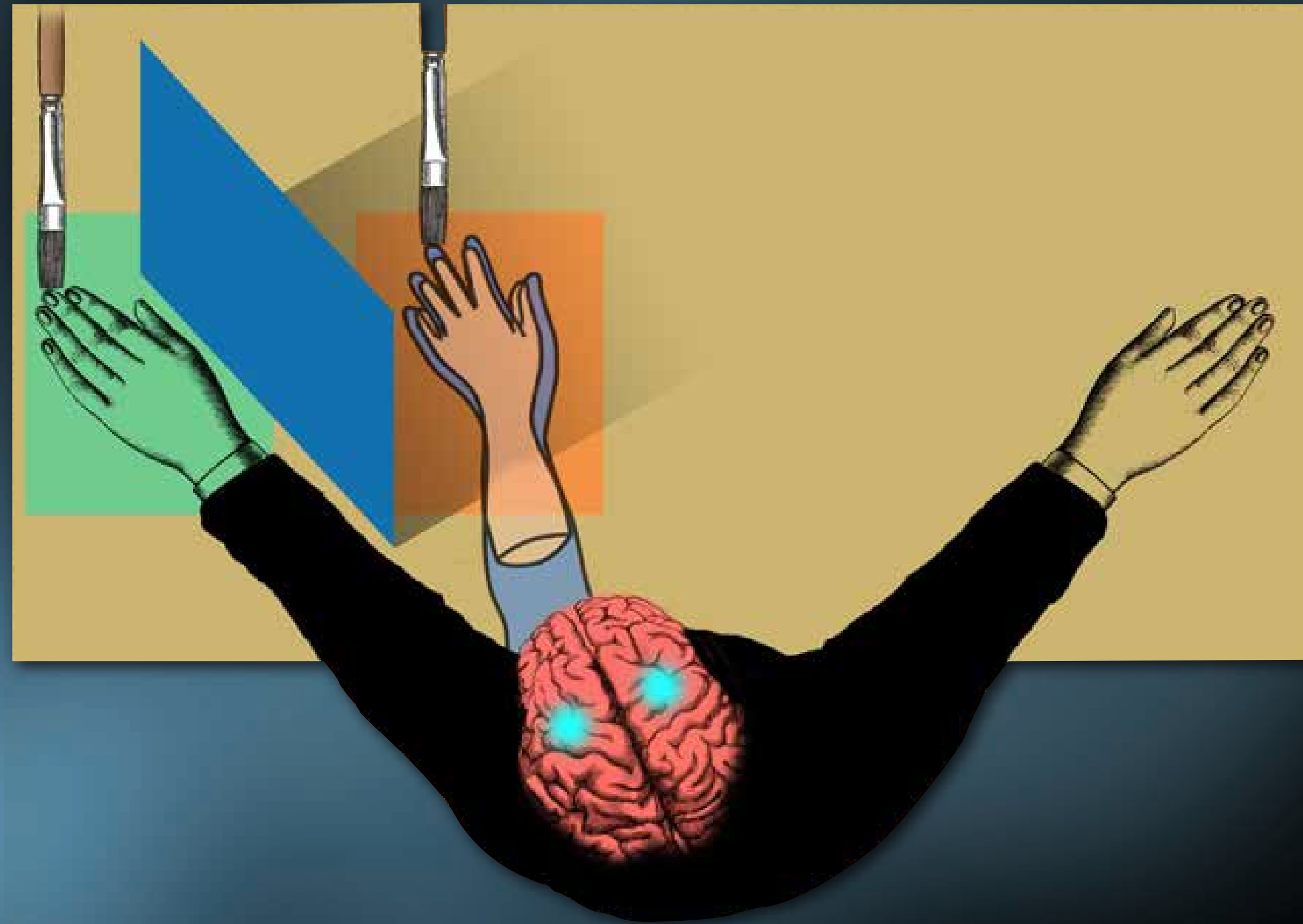
Kagan et al (2022) *Neuron*







# rubber hand illusion





“We do not see things as they are. We see them as we are.”

Anais Nin (1961)

# The Perception Census

Honorary Mention: EU Prize for Citizen Science, 2024

<https://perceptioncensus.dreamachine.world/>

the 'AI' brand

artificial intelligence

machine learning

applied statistics