

AI as agency / The nature of intelligence

Agency without intelligence?

the
Sociological
Review

theSociologicalReview.org



Luca Consoli and Michel Vitale, 09-12-2024

Outline

- AI as agency without intelligence?
- The nature of intelligence
- Explanation final term group paper

<https://openai.com/index/openai-o1-system-card/>

<https://www.wheresyoured.at/godot-isnt-making-it/>

[Kia will share insights from motorist data with governments and businesses](#)

AI as a new form of agency

Floridi argues that AI should **not** be viewed merely as a tool designed to replicate human intelligence. Instead, AI represents a new form of agency: autonomous systems that interact with the world and make decisions, even though they do not possess consciousness or intentionality like humans (and precisely because of that! Because of their not being intelligent!)



Luciano Floridi

Agency vs. Intelligence: This distinction is crucial. Floridi's theory suggests that AI's ability to perform tasks traditionally associated with intelligence does not imply that it has intelligence in the human sense. It may perform actions, make choices, and influence outcomes without understanding or awareness.

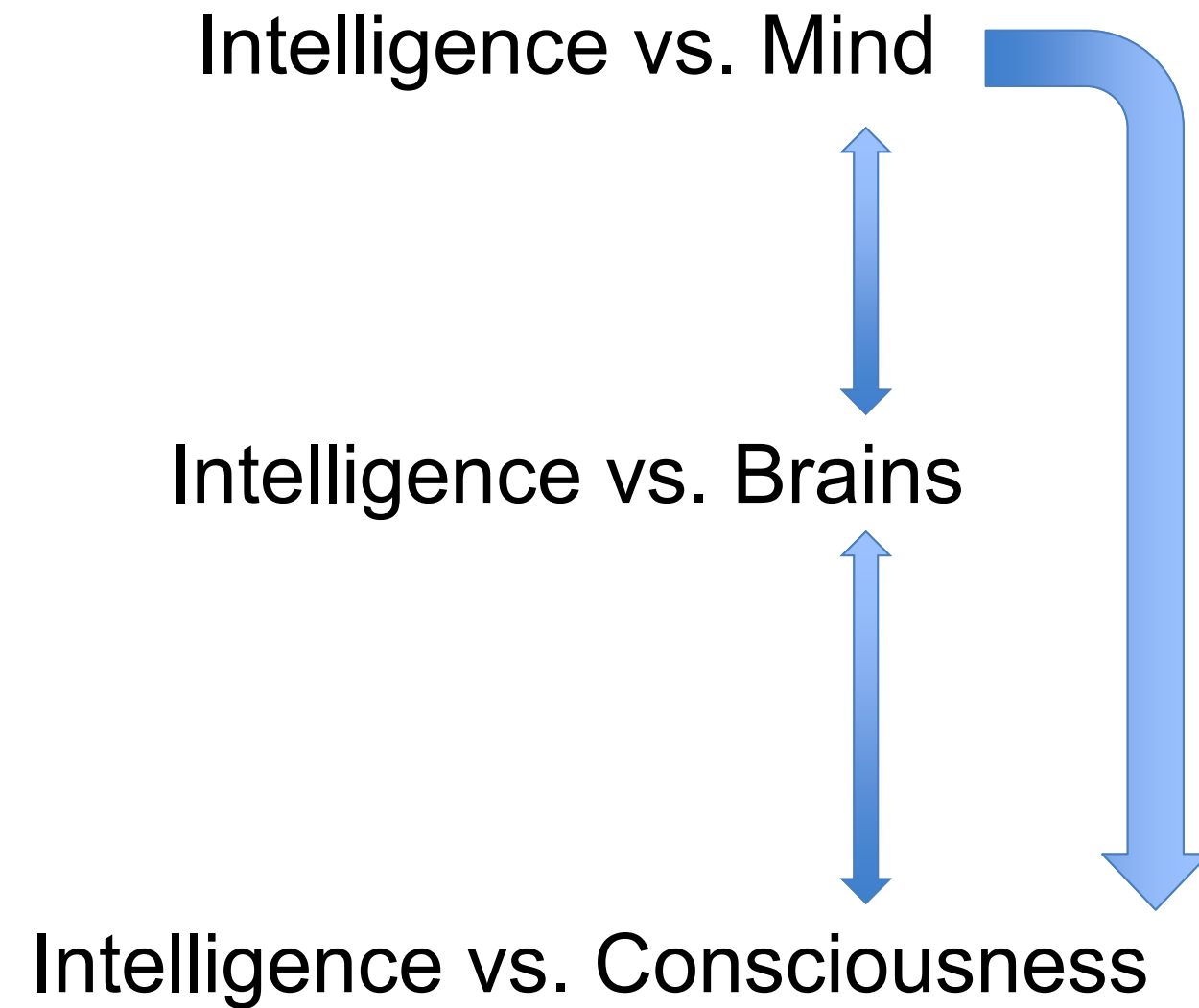
AI as a new form of agency

Does AI Require AGI (Artificial General Intelligence)?:

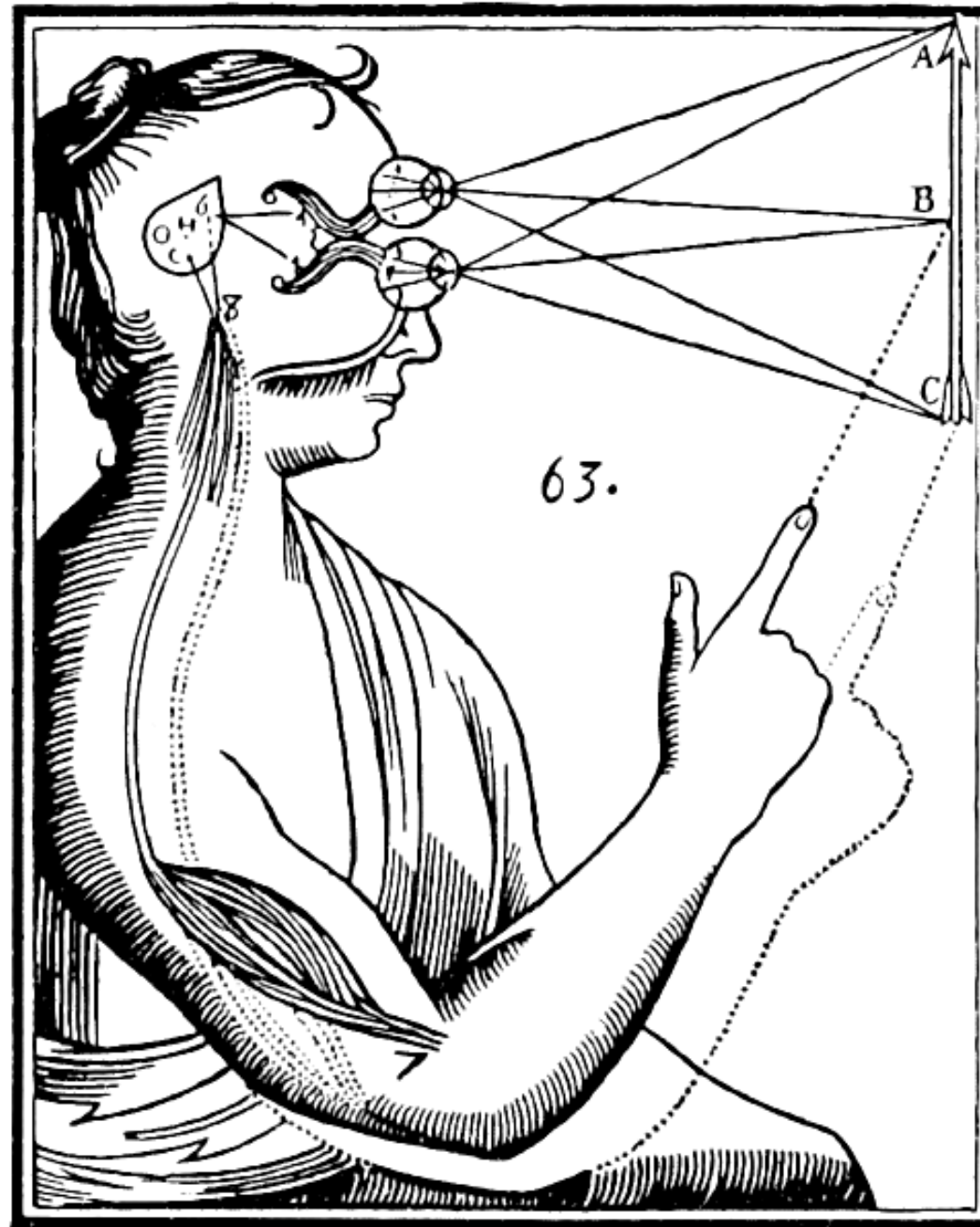
Some thinkers argue that proper artificial intelligence (comparable to human cognition) requires AGI, which would be capable of learning and adapting across different domains. Floridi counters this argument by suggesting that AI doesn't need AGI to function as a form of agency, it doesn't need intelligence in the first place. AI can perform specialized tasks, like diagnosing diseases or driving cars, without being “intelligent” in the general sense. We discovered that tasks that we always associated with intelligence, actually do not necessarily need intelligence to be carried out, but just some advanced pattern recognition technology, made possible by simple ML algorithms and impressive computing power...

What is intelligence?

Some important (?) distinctions



AI and the Mind/Body Problem



Grand Challenges in Science and Technology

- 1. Understanding the brain
 - Reasoning, cognition, creativity
- 2. Creating intelligent machines
 - Is this possible?
 - What are the technical and **philosophical** challenges?

Arguably, AI poses some of the most interesting challenges and questions in computer science today!

Understanding the brain: the Human Brain Project



112 Partners in 24 countries in Europe and around the world

Budget 2013-2023: 1.2 billion EUR

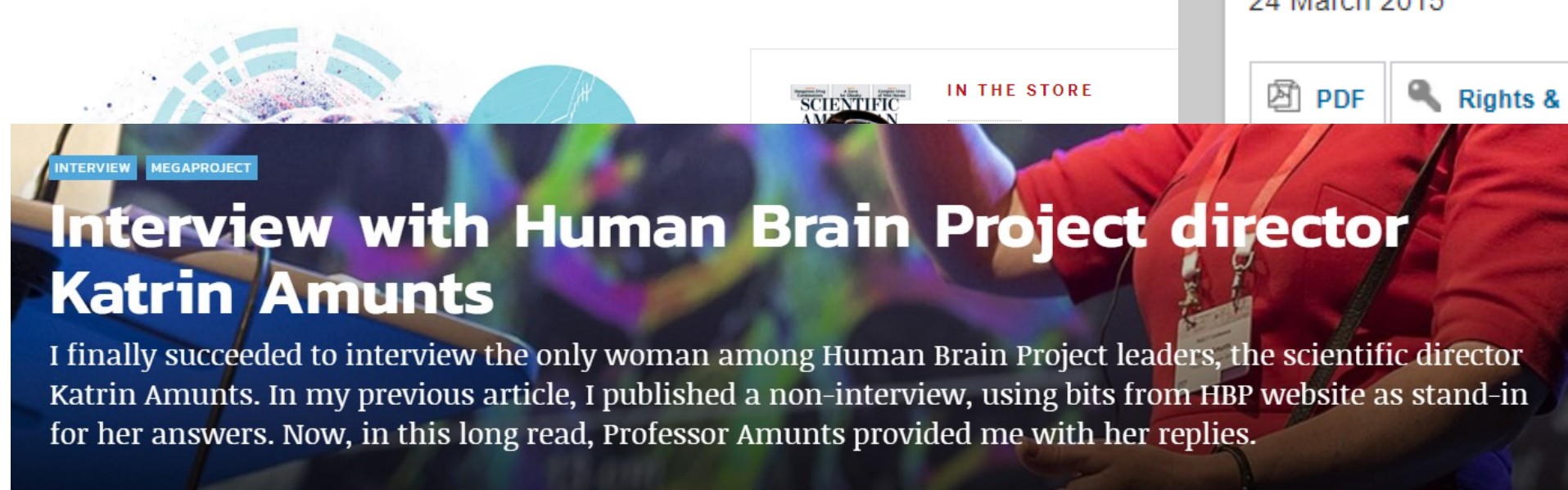
Human Brain Project

NEUROSCIENCE

Why the Human Brain Project Went Wrong—and How to Fix It

Two years in, a \$1-billion-plus effort to simulate the human brain is in disarray. Was it poor management, or is something fundamentally wrong with Big Science?

By Stefan Theil on October 1, 2015



BY LEONID SCHNEIDER
SEPTEMBER 14, 2018

COMMENTS 6

I hope we can agree that maps, models and simulations are necessarily a reduction, both regarding the spatial scale and the modalities you choose. When we talk about brain modelling and simulation, this does not mean that we create a 1-to-1 virtual duplicate of “the brain”.

NATURE | EDITORIAL

عربي

Rethinking the brain

Critics of the European Human Brain Project were justified, the project. Both its governance and its scientific direction

24 March 2015



PDF

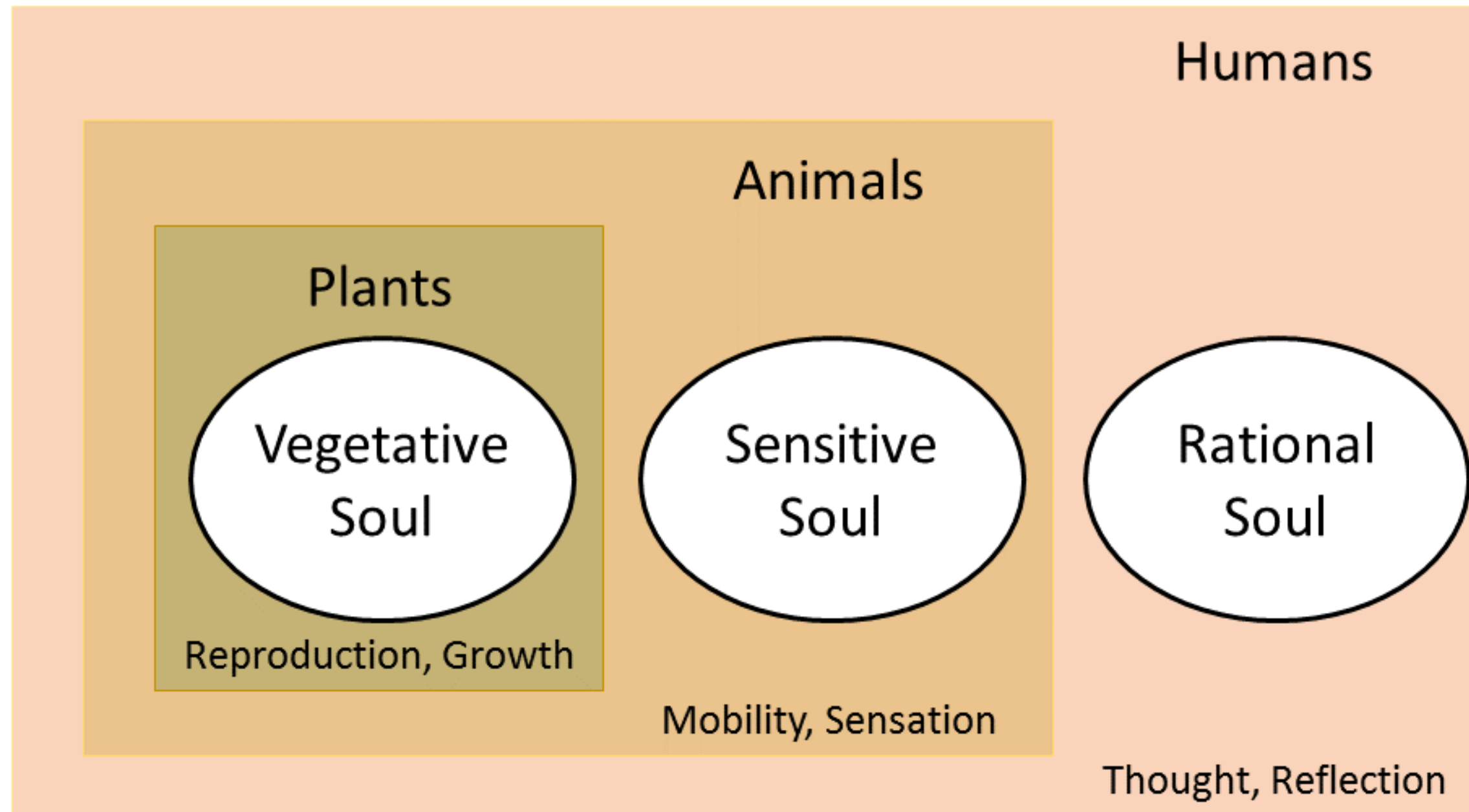


Rights & Permissions

itself, the European Commission's billion-dollar initiative. Launched 18 months ago, the m

Man as a Rational Animal

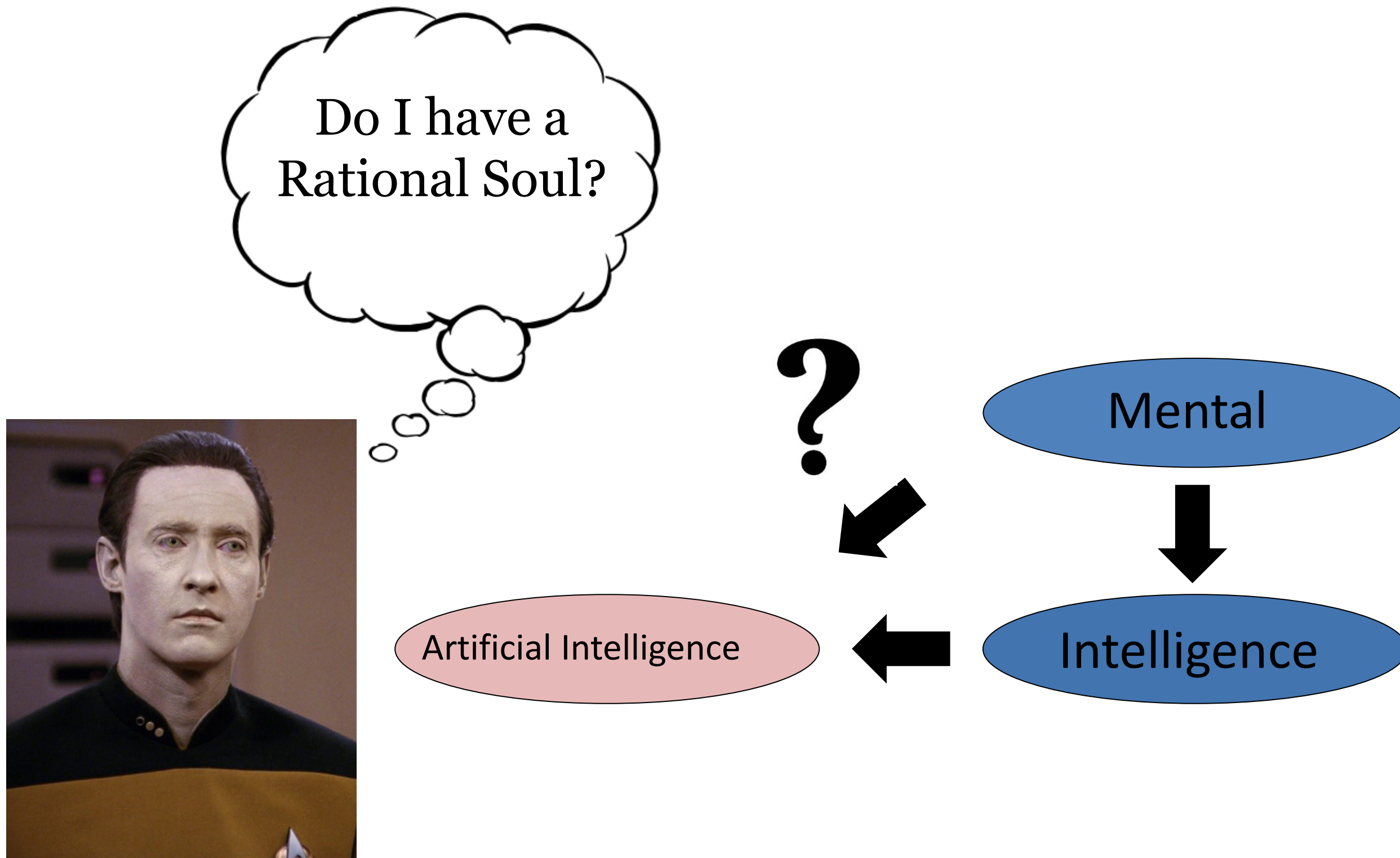
Aristotle 384-322 BCE (Before Common Era)



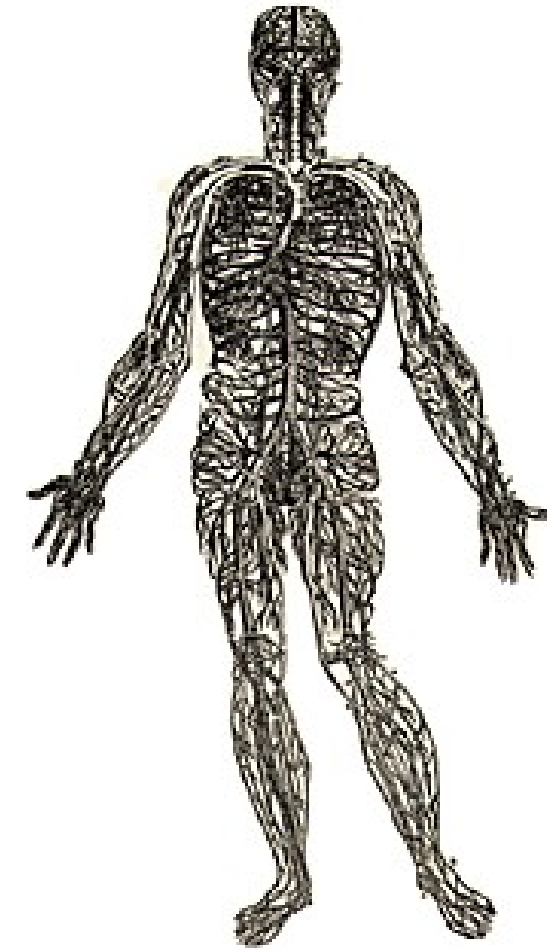
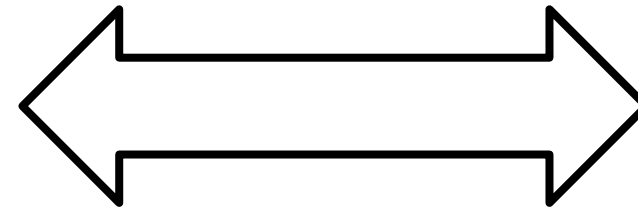
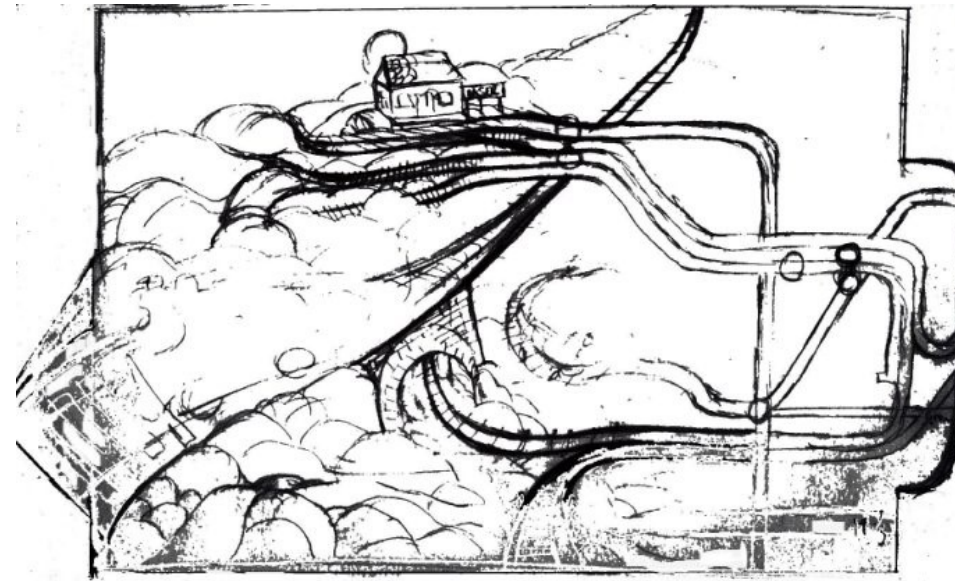
What is Intelligence?

- Intelligence:
 - “The ability to acquire and apply **knowledge** and **skills**” (Oxford English Dictionary)
 - Here, ‘apply knowledge’ = reasoning
- Question “What is knowledge?” → Epistemology
 - Very interesting question, but not our question today...
- Question “What is intelligence?” → Philosophy of Mind
- More generally, intelligence is a property of a **mind**; it is a so called **mental property** (just like reason, perception (in part), understanding, et cetera...)

Can we create something mental in a non-human body?



I. Simple framing of the problem



How are **mind** and **body** related?

OR

How are **mental phenomena** and **physical phenomena** related?

II. A more precise framing of the problem:

Inconsistent Tetrad (Ludwig 2003):

1) Realism of the mental

Some things have mental properties

2) Conceptual autonomy

Mental properties are not reducible to non-mental properties

“mental phenomena \neq physical phenomena”

3) Constituent explanatory sufficiency

A constituent description of a thing entails a complete description of it

“A radio is nothing but a list of its parts and a description of how they fit together”

4) Constituent non-mentalism

Basic constituents of things do not have mental properties.

“A neuron, or a single motor protein, does not have mental properties”

Important: The truth of the conjunction of any three of these propositions entails the falsity of the fourth

III. Independent plausibility of (1)-(4):

(1) *Realism of the mental*

Just seems obviously true...

Introspect...

(2) *Conceptual autonomy*



If the mental can be reduced to the non-mental, then, conceptually, non-mental propositions must entail mental propositions

However:

(a) We can imagine non-material, mental things

- Gods, angels, spirits are at least conceivable
- Thus, the material is not conceptually *necessary* for the mental

(b) We can imagine material, non-mental things

- Perfect biological robot mimicking human behaviour (*Think of the ‘synths’ from the ‘Fallout’ series*)
- But no consciousness, no ‘real’ mentality
- Thus, the material is not conceptually *sufficient* for the mental

Thus, as a conceptual point, non-mental propositions do not entail mental ones

Thus, it does **not** seem that the mental can be reduced to the non-mental per se

(3) Constituent explanatory sufficiency

Look to the success of science:

Successful explanation of behaviour of complex systems is done in terms of constituents and the laws that govern them

This is how our best science works; it is usually enough!

(4) Constituent non-mentalism

Again: look at successful science

Physics explains the world without any appeal to mental properties

‘Mass’, ‘spin’, ‘velocity’, ‘size’...

Neurobiology explains the brains in terms of neurons and proteins without these having mental properties

So it looks like the basic constituents of the world do not have mental properties

So (1)-(4) are independently plausible, but jointly inconsistent

IV. Reject (4) – constituent non-mentalism

Deny that basic components cannot have mental properties

Substance Dualism *(Descartes 1596-1650)*



‘Substance’

Aristotle- “that which is neither said of a subject nor in a subject”

Medieval scholastics- “independent existents”

Substance as *property bearer*

Two kinds of substance:

Material (Corporeal)

Body: characterized by extension (categorical property)

Mental (Thinking)

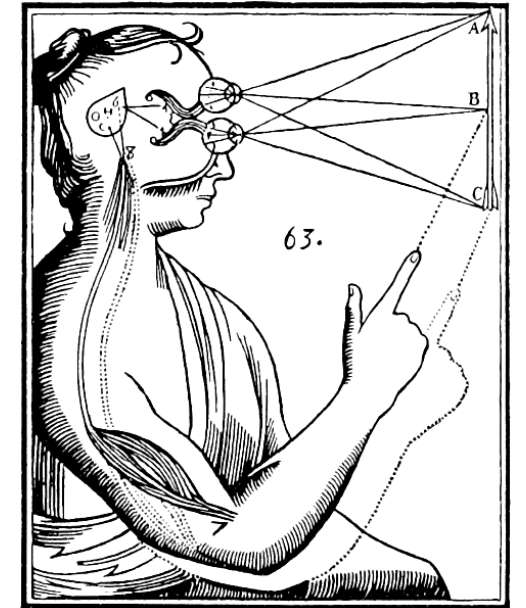
Non-material

Mind: characterized by thought (categorical property)

Mental causation

Not going into the details of Descartes' views, imagine we accept them, then how are mind and body causally related?

Descartes: *Causal interactionism*



Problem: We must then investigate empirically to see if it is *actual*.

...doesn't look so good.

V. Reject (3)- Constituent explanatory sufficiency

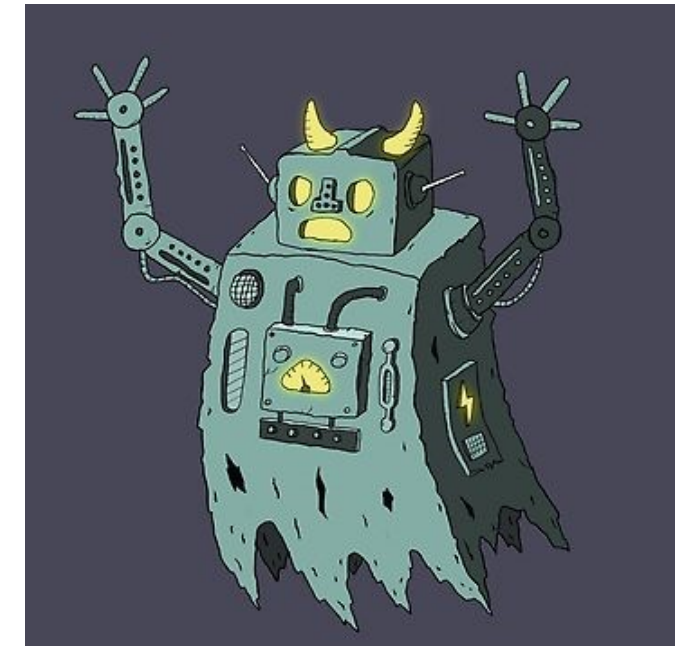
Deny that a constituent description of a thing is or entails a complete description of it. (*The whole is more than the sum of its parts*)

Emergentism (*Huxley 1901; Morgan 1923; Broad 1925*)

- Only collections of basic constituents (complexes) have mental properties
- These mental properties are *not* conceptually reducible to the physical properties of the complex
- Mental properties thus *emerge* from material complexes.
- So, even if we have a complete constituent description, we are not going to find the all of the mental there!

Problems: Emergentism and mental causation

Downward causation vs. epiphenomenalism



VI. Reject (2)- Conceptual autonomy

Deny that mental properties cannot be reduced to physical properties. (*In other words: mental properties = physical properties*)

Logical behaviourism (*Carnap 1931; Hempel 1935; Ryle 1949*)

Logical positivism and behavioural psychology

All mental talk can be translated into behavioural talk.



‘B is in pain’ just means that B is exhibiting behaviours such as wincing, groaning... (or is disposed to given the right circumstances)

- How do we test for (verify) mental ascriptions?
 - observe behaviour
- Thus the *meaning* of mental ascriptions are behavioural

Identity theory *(Place 1956; Feigl 1958; Smart 1959; Lewis 1966, 1972)*

- Mental states are brain states:

To be in pain, is thus just for a certain event to occur in the brain.

- Resistance to mind/brain identity:

Leibniz's Mill

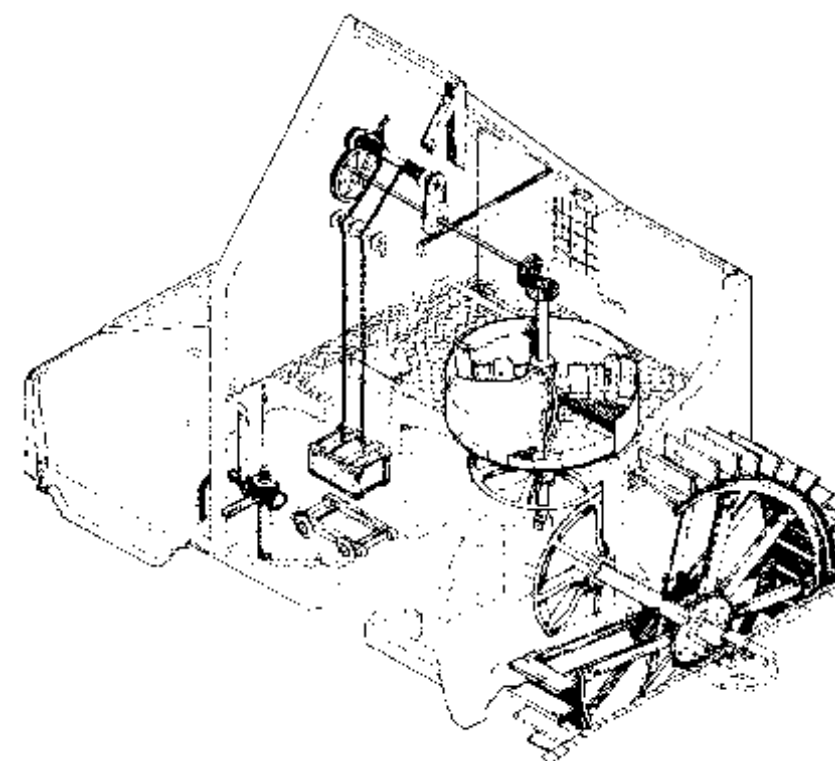
- Yet, think of other empirical discoveries:

lightning = electrical discharge

gold = element with atomic # 79

clouds = masses of water droplets

mental states = brain states



How are mental states and brain states identical?

- **Token Identity** – One particular mental state token is identical with one particular brain state token

(Like Donald Trump is identical to the President of the United States of America)

- **Type Identity** – All mental state tokens of type X are identical to all brain state tokens of type Y

(Like all water molecules are identical to all H₂O molecules)

Problems:

- Token Identity Theory lacks generalized explanatory power
- Type Identity Theory is chauvinistic...
 - Imagine the mental state 'pain' is type identical to brain state 'c-fiber firing'
 - So, pain simply is c-fiber firing
 - An octopus does not have c-fibers, but d-fibers...
 - ... an octopus cannot feel pain?



Functionalism (Putnam 1960; Fodor 1968; Block & Fodor 1972; Block 1978; Sober 1985)

Mental states can be equated with *functional roles*.

According to functionalism, to be in a mental state M (e.g. pain) is to have an internal state which does the M-job



functional role (*pain role*)

- What matters is function and not the particular set-up
(*Think of a abacus versus a calculator: different material, same function, namely to facilitate calculation*)
- Advantage: a mental state M can be realized in multiple ways again
(*This was a problem for type identity theory*)
- Problems: China Brain, Blockhead, Chinese Room → **Next lecture**

VII. Reject (1) – Mental realism

Deny that there are mental phenomena

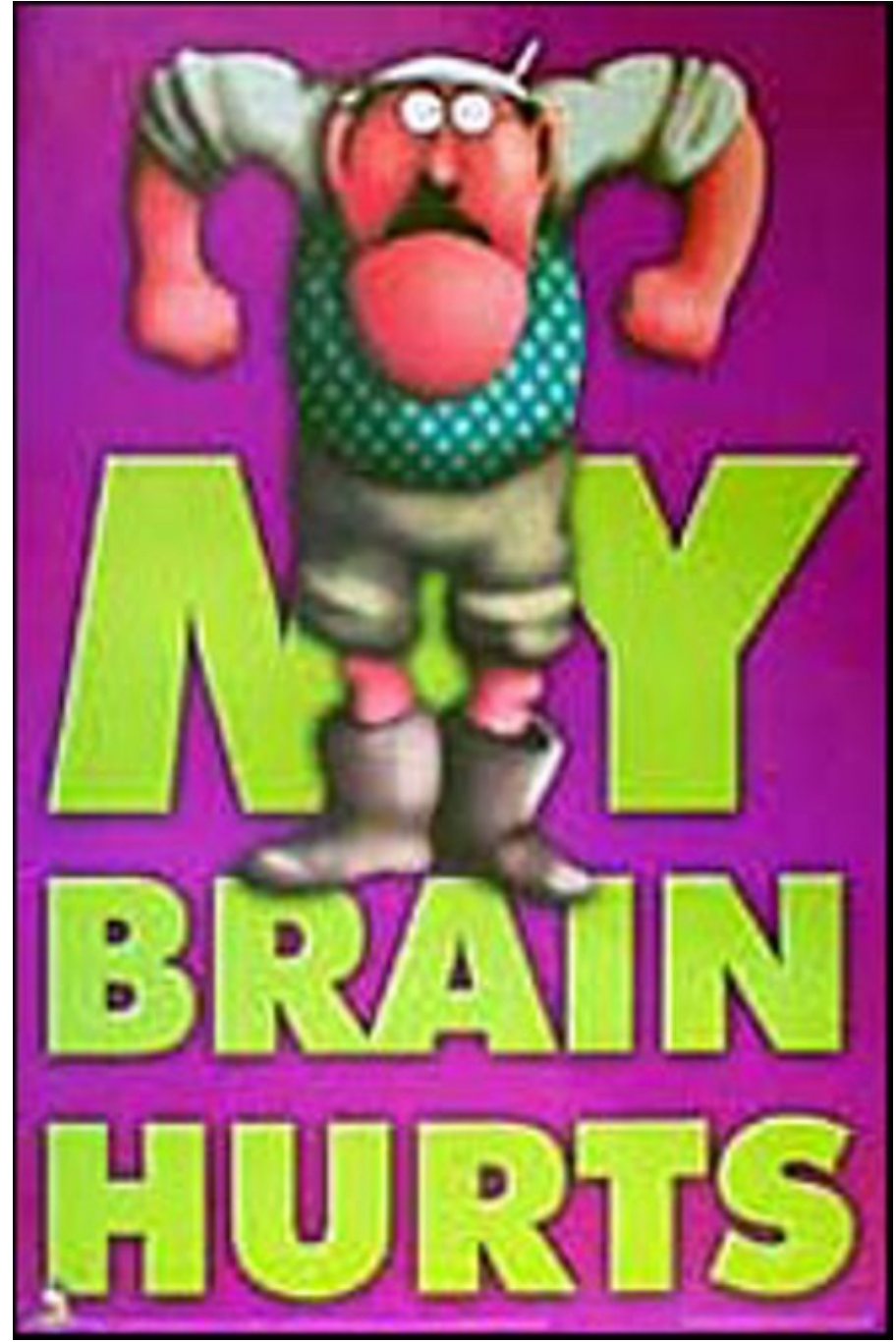


Eliminativist Materialism (*Churchland 1981; Stich 1983*)

Deny the reality of mental properties, hence 'folk psychology' mental language is useless.

How?

- Mental terms (the concepts they express; the properties they pick out) are symptomatic of our pre-scientific theories of behaviour...
- With the advent & development of neuroscience and cognitive psychology, these terms are to be replaced with neurophysiological ones.
- This change is not merely semantic...



Final term paper

The final term paper (group, summative) is a short 'research paper' that you have to write in order to track and evaluate how you can apply the 'philosophical way of thinking' we have discussed during the course. The idea is to choose a topic - you are free in the choice as long as it has to do with ethics and philosophy of AI / computer science / information science -, formulate a small 'research question' and answer it in an academic way.

The paper should have a clear structure: an introduction, the statement of your research question, the answer, a conclusion, a literature list. **The literature list is a crucial part of the assignment!** It shows that you are able to place your paper in a broader context.

Deadlines (strict!): 26-01-2025 or 06-04-2025 (resit)

Length: ~3000-3500 words (~8 pages), max 4000 words!

Structure:

- Introduction (10-15% of text): What is your question? Why is it relevant? What is your answer? What are you going to do in the rest of the text? (Not necessarily in this order!)
- Body of text: Contains your analysis and answer.
- Conclusion (10-15% of text): Briefly restate your question, what you've done, and how this results in your answer.
- Literature list: Don't forget this! It is also important to mark in the text where you have used which source. You are free to use any of the main reference styles, see this link: <https://www.rlf.org.uk/resources/mla-apa-harvard-or-mhra/>