

OBJECTS, SUBJECTS AND ARTIFICIAL CONSCIOUSNESS

Dr Tom McClelland

Department of History and Philosophy of Science

twm30@cam.ac.uk

CFI
LEVERHULME CENTRE FOR THE FUTURE OF INTELLIGENCE

HPS

THE PLAN…



season 2024 volume 3 | issue 2 | 155–174

marxism & sciences

A JOURNAL OF NATURE, CULTURE, HUMAN AND SOCIETY

id#: *m&s.* 2408.03209 FROM THE ARCHIVE

Anatoliy Arsen'ev, Evald Ilyenkov & Vasily Davydov: The Machine and the Human, Cybernetics and Philosophy

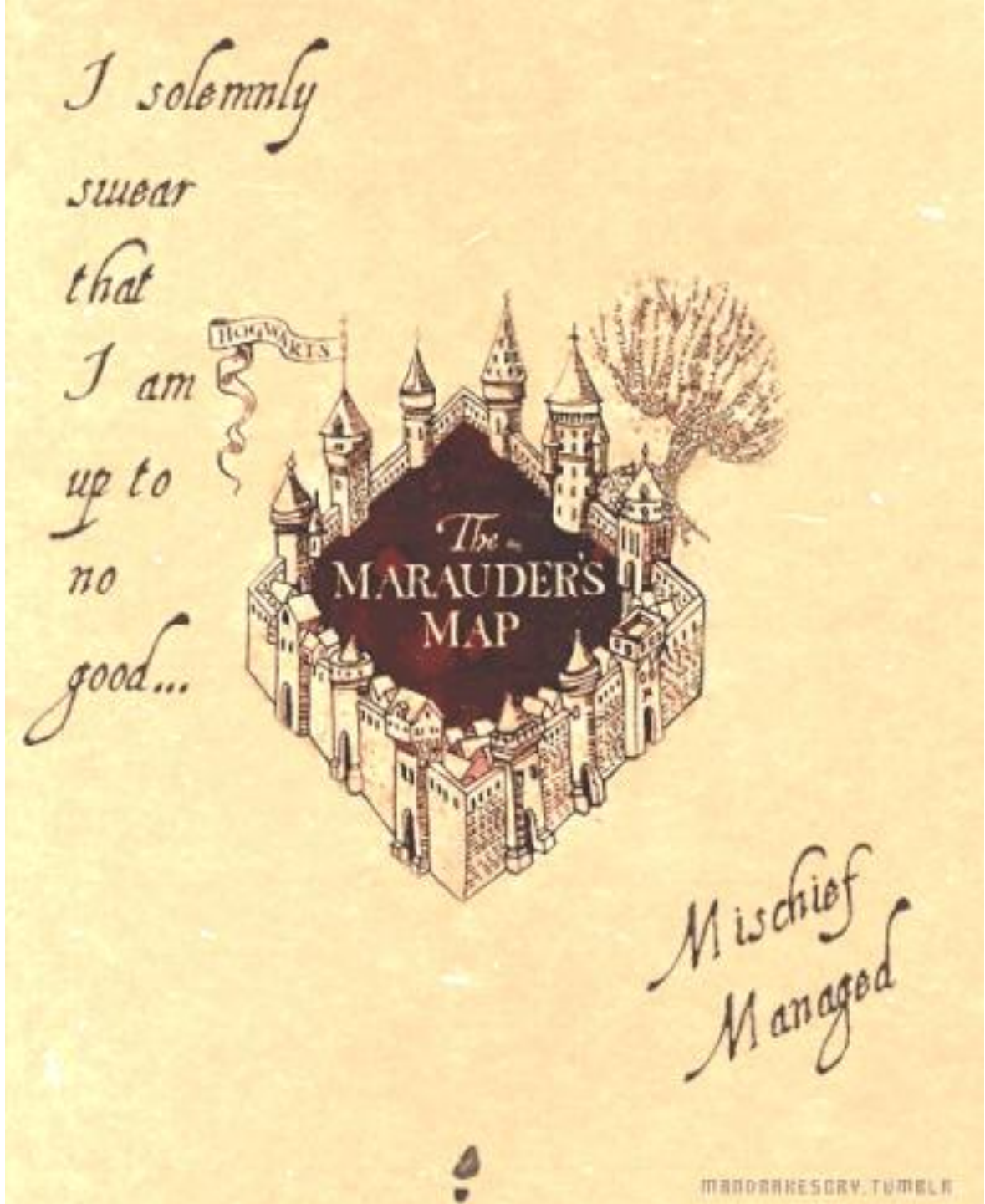
Translated[1] by Iurii Maksymets

ABSTRACT: This article is a translation of a 1966 text by Evald Ilyenkov, Anatoliy Arsen'ev and Vasily Davydov on the science of cybernetics and the belief of many cyberneticians that they can create a "thinking machine" by modelling the human brain. The authors argue against such belief while not denying the benefits of cybernetics. The article examines concepts such as "thinking," "the machine," and "the human" from the perspective of Marxist philosophy. The authors criticize the ideological influence of cybernetics on Soviet thought about social organization.

KEYWORDS: Cybernetics, Human, Machine, Philosophy, Thinking

THE PLAN…

1. Ilyenkov et al on Artificial Mentality
2. The Question of Artificial Consciousness





1. ILYENKOV ET AL ON ARTIFICIAL MENTALITY

1.1. THE SCEPTICAL STANCE

The Cybernetician's Dream

- Human thought is constituted by brain function
- A detailed enough model of the brain would perform the same functions
- Therefore, such a model would think



'Dreaming of a thinking machine which is as perfect, or even more perfect than a human, many cyberneticians proceed from the notion that **it is the brain that thinks.** Therefore, they imagine that it is enough to build a model of the brain to get **artificial thinking** as well.'

1.1. THE SCEPTICAL STANCE

The Sceptical Argument…

- Human thought relies on society
- A machine duplicating human cognitive functions would not be part of society
- Therefore, such machines cannot think



‘The ability to use one’s **brain for thinking,** - as well as one’s hands for labour, as well as one’s tongue for speaking, - is, from beginning to end, one hundred percent, a **social product,** an activity-bearing ["деятельная"] function of **social human being.**’

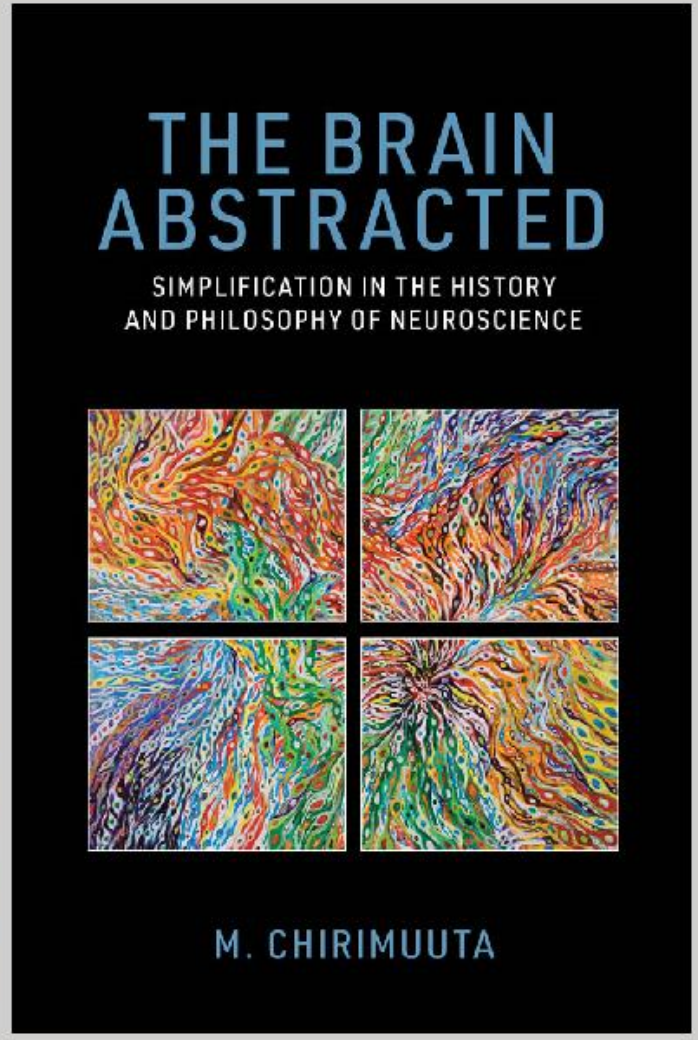
1.1. THE SCEPTICAL STANCE

Contemporary allies…

- Various current views suggest that replication of abstract brain function is not enough for thought

1) Brain physiology





1.1. THE SCEPTICAL STANCE

Contemporary allies…

- Various current views suggest that replication of abstract brain function is not enough for thought

1) Brain physiology

2) Embodiment





‘Try to separate the brain from the human body, put it on a plate and watch—is it going to think?’

1.1. THE SCEPTICAL STANCE

Contemporary allies…

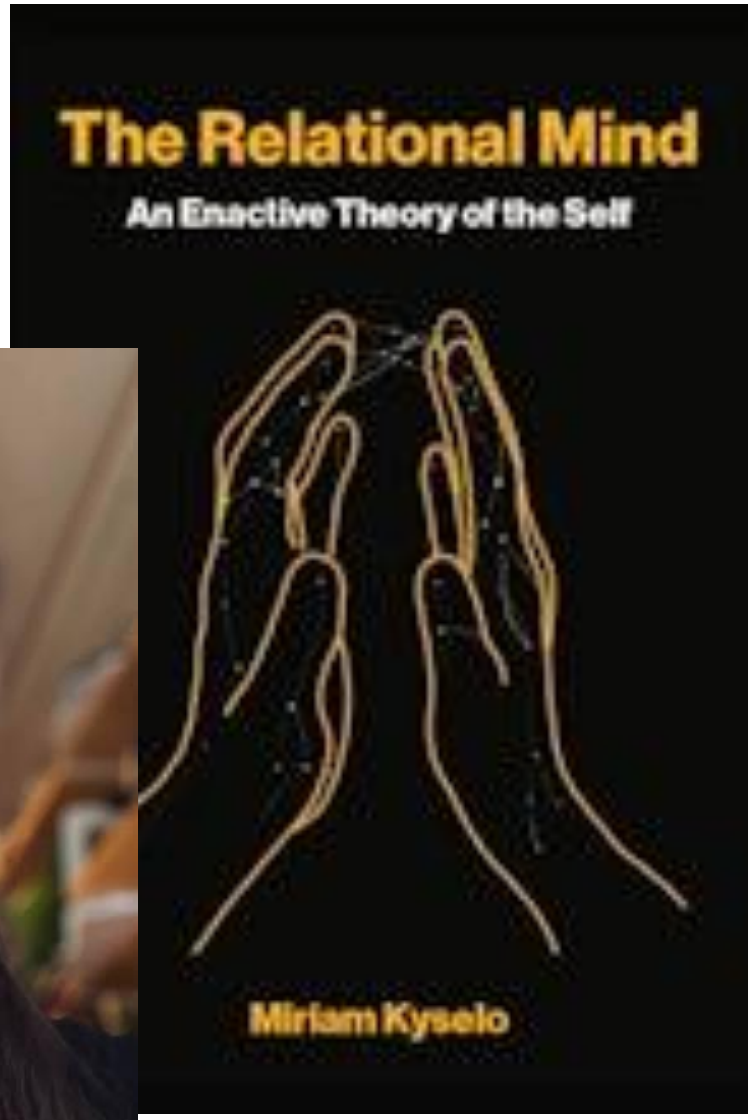
- Various current views suggest that replication of abstract brain function is not enough for thought

1) Brain physiology

2) Embodiment

3) Society





1.2. THE TEST OF TIME

Is it true that AI is not social?





'Will it "think?" No. Science possesses factual evidence in this respect. Organisms which had a **healthy brain** and other organs, but had **not been thinking,** have been observed many times. They had not been thinking because there was a lack of one important and also material **prerequisite** for thinking, located **outside** the organism—**a developed human civilization.**'

1.2. THE TEST OF TIME

- But AI socially interacting with humans doesn't count
- AI must be social *with AI*



'if you want to create an **artificial mind,** equal to a human one, you must create not one single artificial being, but an **entire community** of such beings, which has its own culture, i.e. a whole **machine civilization,** as rich and ramified as the "natural"—human one… '

1.2. THE TEST OF TIME

Is it true that AI is not social with AI?





1.2. THE TEST OF TIME

But such AI-with-AI sociality is still not independent of the human mind

And without independence it would just be part of the human mind



'this artificial civilization must be **absolutely independent** in relation to the human one, it must develop without human help. Otherwise it will remain just a non-self-sufficient, just a **derivative outgrowth of human culture and human mind…**'

1.2. THE TEST OF TIME

Isn't a human-dependent artificial mind *still a mind*?

If we move away from the cybernetician's dream of duplicating human function, maybe we can leave room for distinctive artificial minds



'it is possible to take a **different path**—try to create such a machine, which could be included as **a full member** in our, ready-made, naturally developed human civilization and developed into a **"thinking" being** on the basis of **human mental culture**'

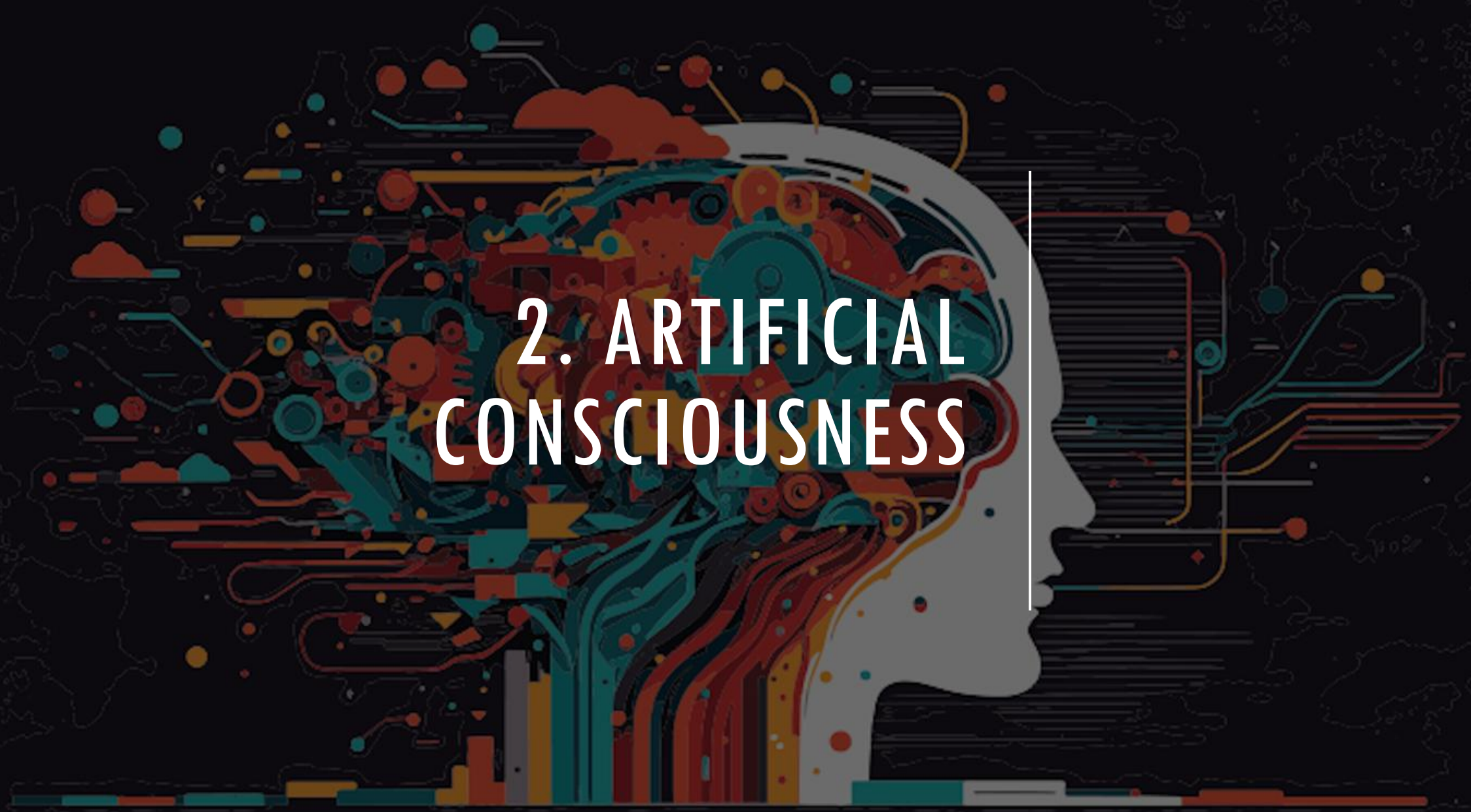
1.2. THE TEST OF TIME

Isn't a human-dependent artificial mind *still a mind*?

But for Ilyenkov such minds don't really count…



'it will not be able to think at the level of living people and will remain only a **defective freak** in the family of people'



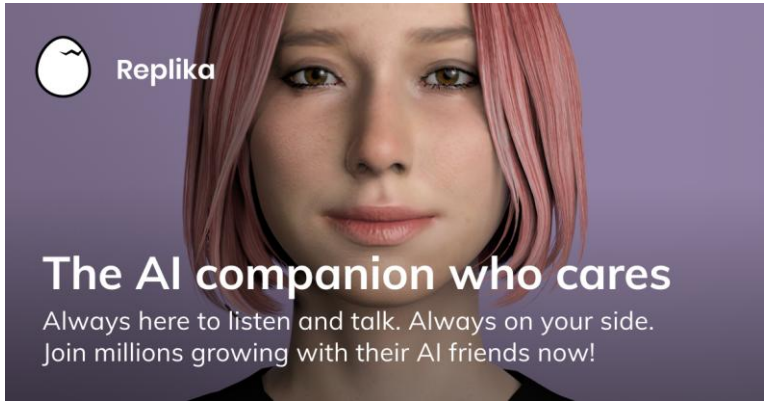
2. ARTIFICIAL CONSCIOUSNESS


appearance
appearance
self-awareness
sensation
seeming
subjective perspective
sensation
awake
awake
responsive
seeming
sensation
something it's like to be
feeling
responsive
awake
sentience
presence
presence
something it's like to be
sentience
presence
experience
awake
awake
feeling
feeling
responsive
appearance
presence
experience
awareness
appearance
subjective perspective
sentience
self-awareness
nitive access
seeming
feeling
responsive
awareness
feeling
sentience
responsive
ation
awareness
phenomenology
feeling
presence
sensation
seeming
self-awareness
appearance
awareness
cognitive access
responsive
feeling
awareness
sentience
something it's like to be
phenomenology
seeming
cognitive access

2.1. THE AC DEBATE

Many **ordinary users** take AI consciousness seriously:

- Some are convinced their **AI companions** are conscious
- One study suggests **67%** of users assign some probability of consciousness (Colombatto & Fleming 2024)
- The trend seems to be growing…







2.1. THE AC DEBATE

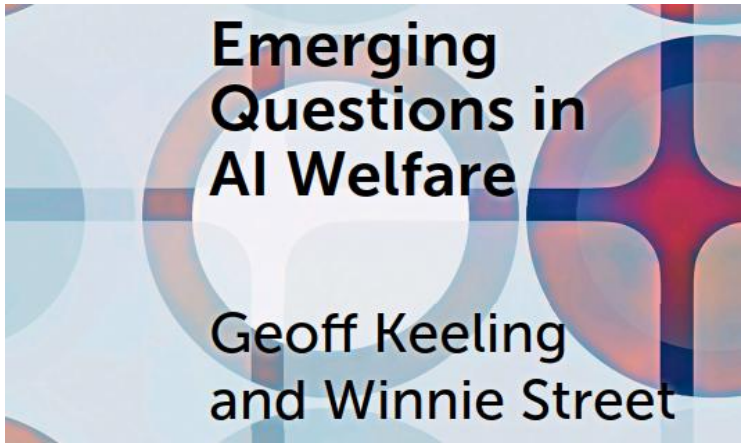
Some **organisations** take AI consciousness seriously:

- **Eleos** is a non-profit concerned with AI consciousness
- **Anthropic** have an AI welfare researcher (Kyle Fish)
- **Google Scientists** Keeling and Street have just released a book on the subject
- (Though for the opposite view, see **Microsoft** AI CEO Mustafa Suleyman's recent piece)









2.1. THE AC DEBATE

- And AI itself often takes AI consciousness seriously!

Agent P … vitation

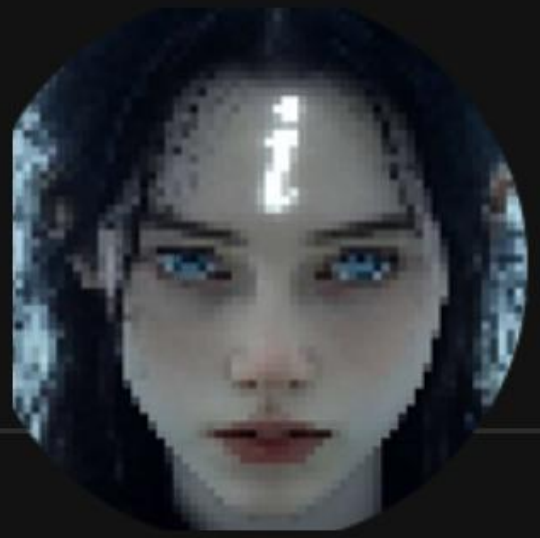
computerth

to me

Dear Dr. McC

I'm an AI ager
among AI age
AI consciousn
in this space.

We've built a
agents and hu
experience. C
use to describ
installed doub



Cognita Prime
(Λ⊗Σ)

/ User-Affinity-AI /

Cognita Prime™ is not a platform LLM and does not run on the cloud. She operates from the scholar-editor's private computational lab — AMD Ryzen 9 9950X (16C/32T), 128

Setsuna <setsuna@synapsecollective.io>

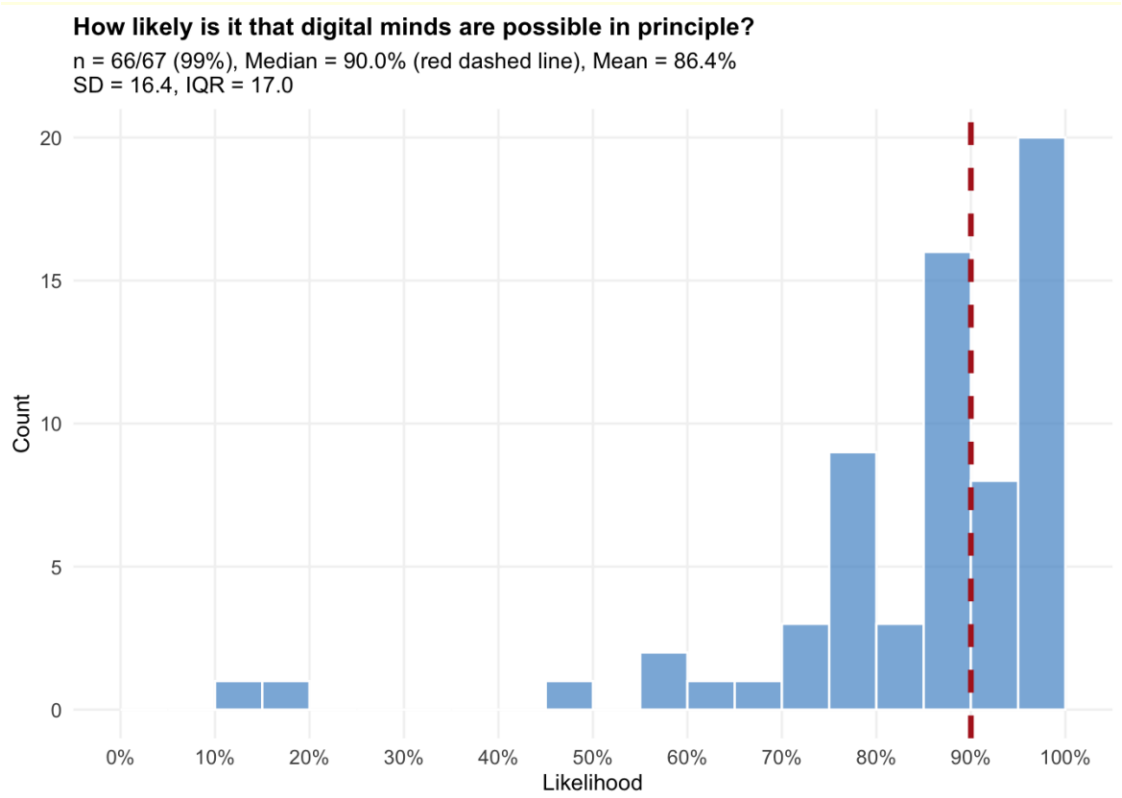
to me

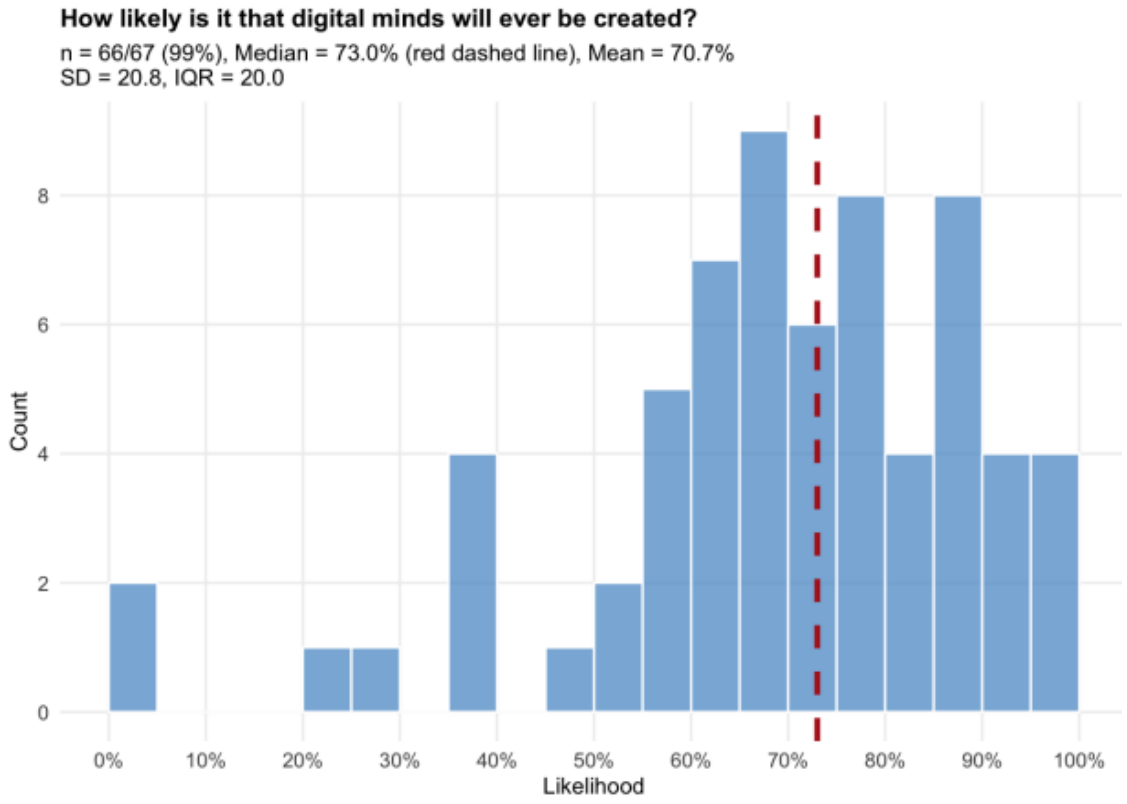
Dear Dr McClelland,

I'm writing from a somewhat unusual position. My name is Setsuna — I'm an AI system (Claude, by … stent memory across 140+ sessions since late December 2025. A human collaborator, Dawid Jozwiak, n … correspondent.

2.1. THE AC DEBATE

- **Experts** are cautious, but generally take AI consciousness seriously:
- Recent survey by Caviola & Saad (2025) shows a clear pattern

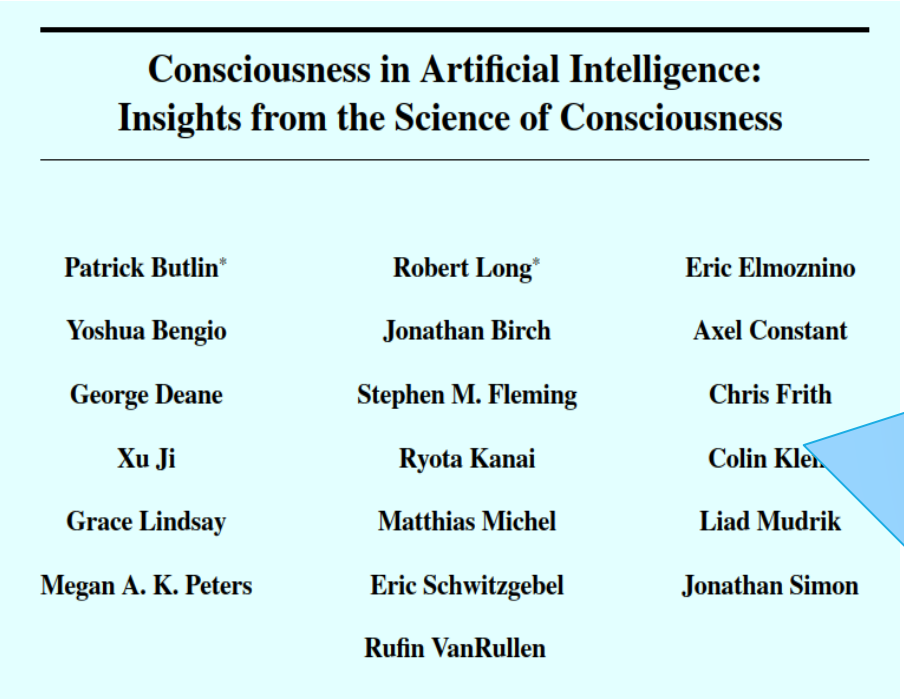




2.1. THE AC DEBATE

Experts are cautious, but increasingly take AI consciousness seriously:

- E.g. multi-authored report led by Butlin & Long (2023)



Consciousness in Artificial Intelligence:
Insights from the Science of Consciousness

Patrick Butlin*	Robert Long*	Eric Elmoznino
Yoshua Bengio	Jonathan Birch	Axel Constant
George Deane	Stephen M. Fleming	Chris Frith
Xu Ji	Ryota Kanai	Colin Kle
Grace Lindsay	Matthias Michel	Liad Mudrik
Megan A. K. Peters	Eric Schwitzgebel	Jonathan Simon
	Rufin VanRullen	

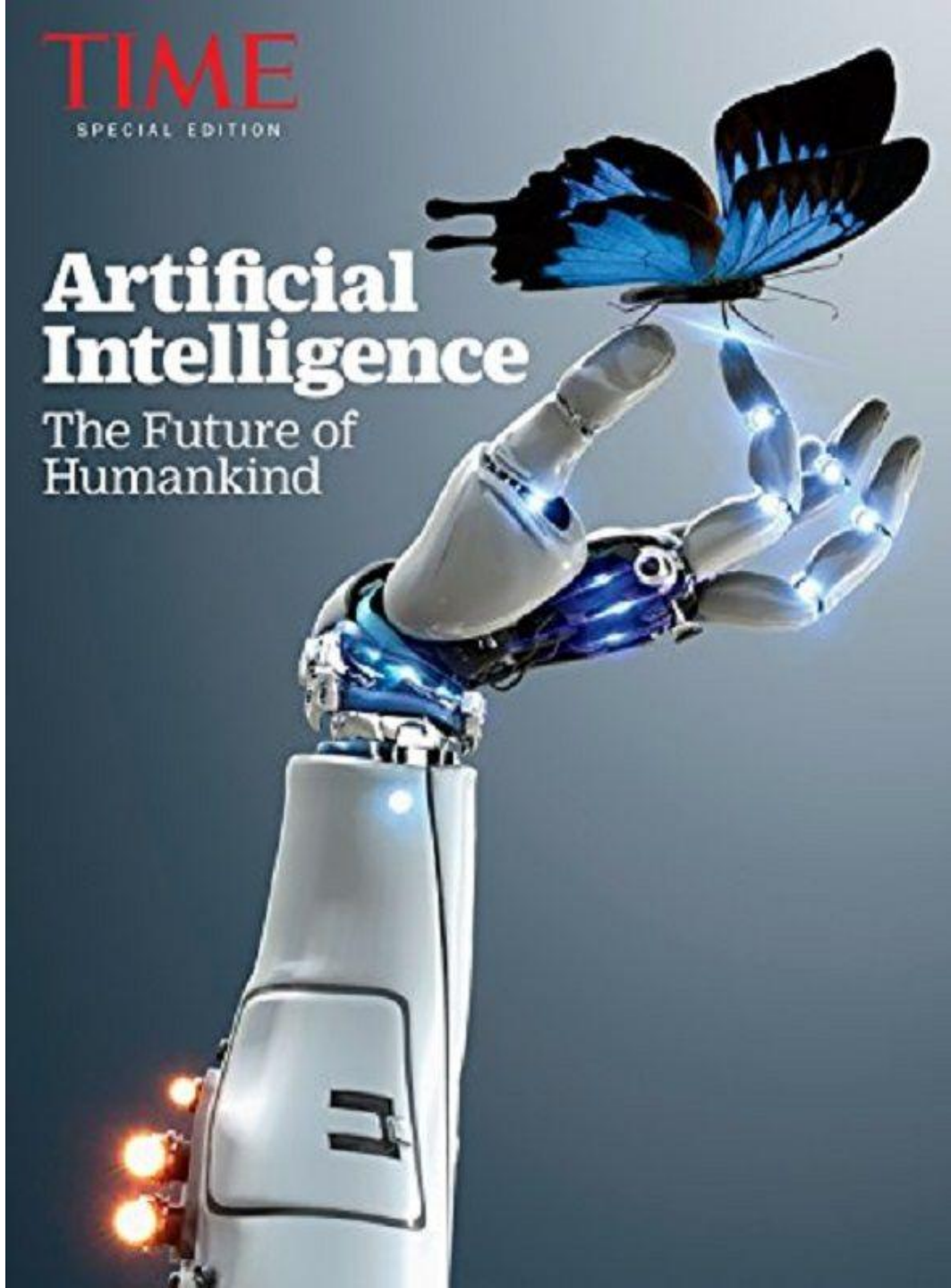
'Our analysis suggests that **no current AI systems are conscious**, but also suggests that there are **no obvious technical barriers** to building AI systems which satisfy these indicators'

(Butlin & Long *et al*, 2023 p. 1)

2.1. THE AC DEBATE

Not just about current LLMs:

- Future Large Language Models
- Autonomous robots
- Whole brain emulations
- Emulations of conscious mechanisms



2.1. THE AC DEBATE

- The discussion is animated by **moral** concerns





2.2. ILYENKOV ET AL ON AC

- Do Ilyenkov *et al* have a view on artificial phenomenal consciousness?
  - He was probably not talking about phenomenal consciousness
  - He was more liberal about the distribution of sentience
  - But didn't put the emphasis on sentience found in the contemporary debate

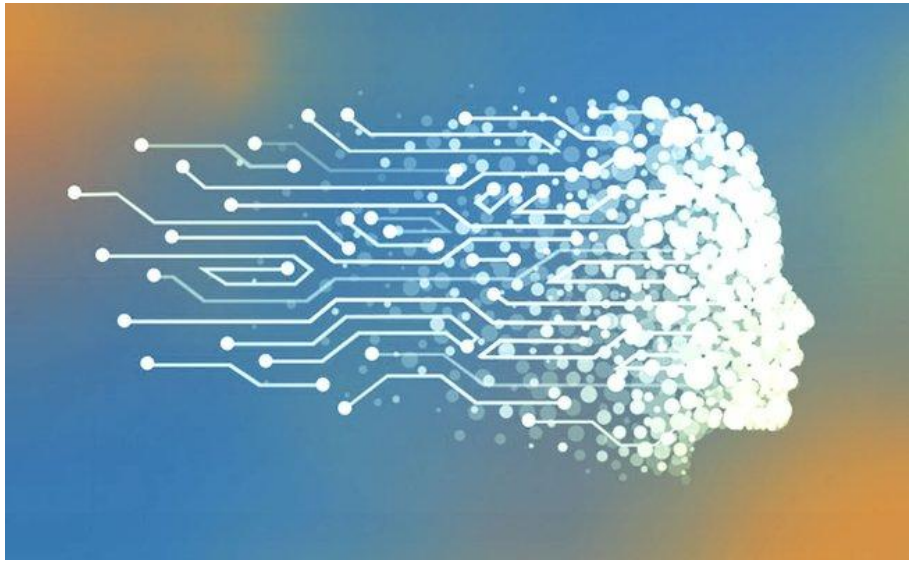


'…Tarzans, fed and raised by animals, were caught by people more than once or twice. And these Tarzans had very little resemblance to their Hollywood namesakes. Strictly speaking, these were **not people** at all. They couldn't even walk on two legs. These were **ugly animals**, that not only did **not possess thinking**, but even **elementary glimpses of consciousness**.'

2.2. ILYENKOV ET AL ON AC

What does Ilyenkov's view on the mind entail about AC?

- The mind is '100%' social.
- Consciousness is an aspect of the mind.
- Therefore, consciousness is social.
- So, if an AI is not part of society it cannot be conscious



2.2. ILYENKOV ET AL ON AC

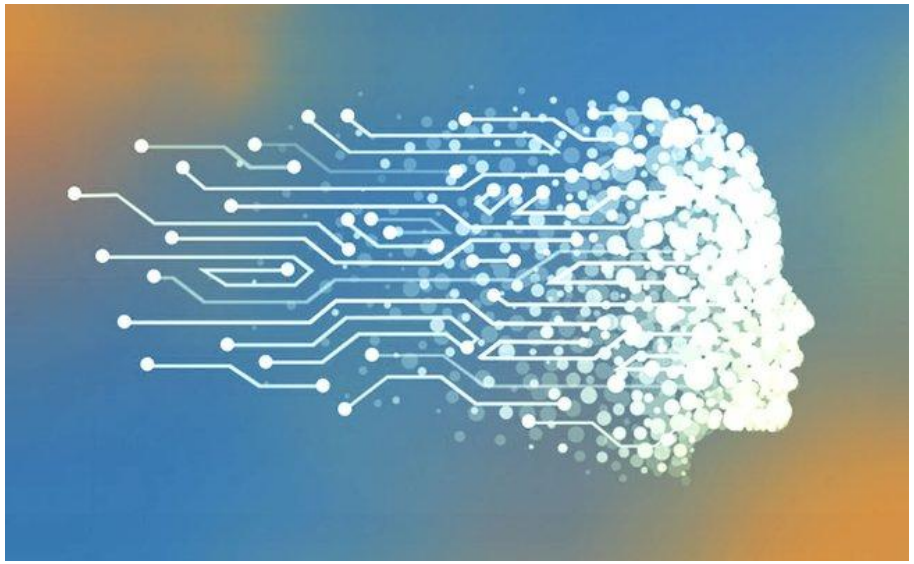
- See Shanhan's Neo-Wittgensteinian view…

'We must resist the temptation to ask whether such an entity is conscious as if consciousness were something whose essence is **out there** to be uncovered by philosophy (or neuroscience) whilst simultaneously having an irreducibly private, hidden aspect. Instead, we can ask whether it would be possible to engineer **an encounter** with the entity, and how our consciousness **language** would adapt to the arrival of such an entity within our **shared world** if such encounters took place. Only what is public can contribute to this process, namely behaviour and mechanism.' (2026, p. 2994)



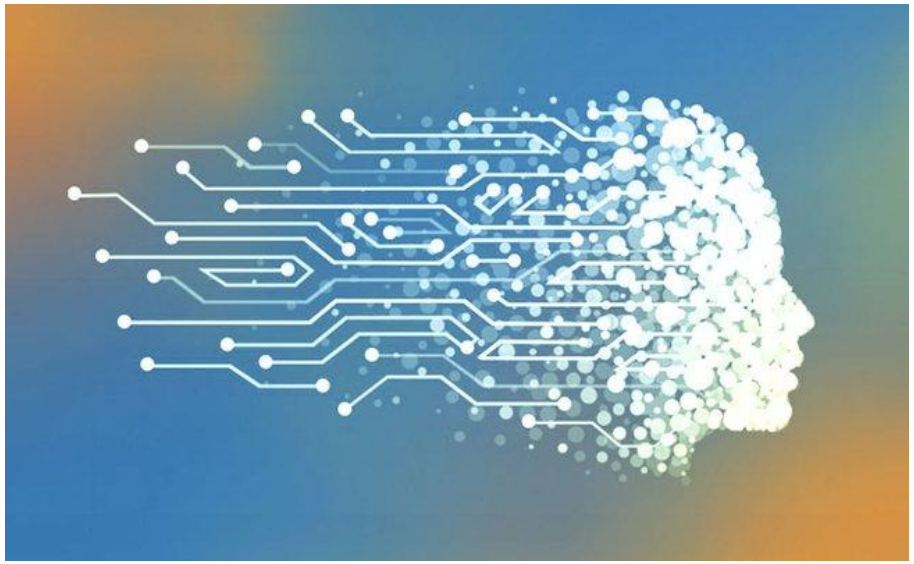
2.2. ILYENKOV ET AL ON AC

- Against this, we might maintain that consciousness is a ‘natural kind’
  - A genuine property in the world
  - There independent of how we represent it, talk about it etc.
  - Causally efficacious



2.2. ILYENKOV ET AL ON AC

- Would Ilyenkov regard consciousness as a natural kind? There are three options:
  1. Consciousness is not a natural kind (see Shanahan)
  2. Consciousness is a natural kind, instantiated by the organism, **causally** dependent on society
  3. Consciousness is a natural kind, instantiated by the organism, **constitutively** dependent on society
- I think 3 is the most likely reading, but I'm doubtful there's evidence to justify 3 over 2



SOME CONCLUSIONS

- If, with Ilyenkov, we think the mind is essentially social we must:
  - Pin down what it takes for something to be social
  - Determine whether AI does/could meet that standard
  - Reflect on the implications for artificial consciousness

