

Are Brains Computers?

(And Could That Grant Us Mindreading?)



Wojciech Mamak
Section for Logic and Cognitive Science,
Polish Academy of Sciences

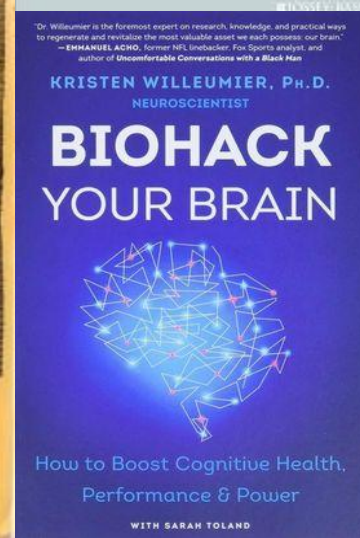
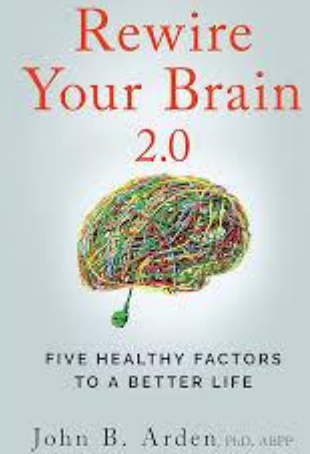
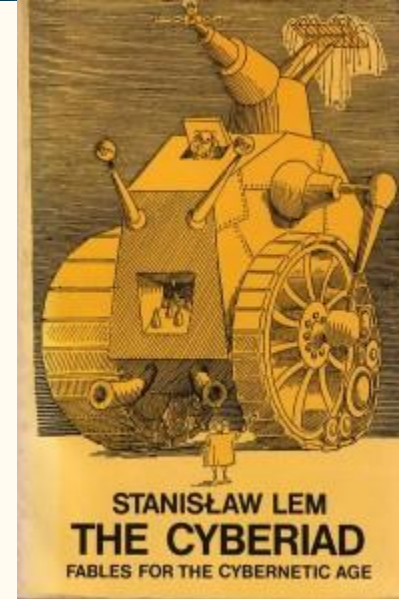
National University of Mongolia
Ulaanbaatar, 23rd Sep 2024

Overview

1. Where does the idea of computational mind comes from? Short historical primer
2. Computer metaphor and computational theories of mind (CTM) - what is the difference?
3. Science of mindreading as a paragon of CTM. Separating practice from hype:
 - a. BCIs
 - b. Neural Readouts
 - c. 'Lie detectors'
 - d. Predictions
 - e. Interventions
4. So, could all that really give us genuine mindreading?
5. Some conclusions and little lessons.

AI + Sci-Fi = Ripe for Hype

- You keep on hearing buzzwords that blur the difference between fiction and science:
- Biohacking, mind uploading, singularity, general artificial intelligence, memory implants etc. **that seem to logically follow from thinking brains are computers. But this is wrong.**
- It makes for very cool books and movies, but in science communication and philosophy it's dangerous (and obnoxious).
- **My aim today: to give you a quick run-down on how scientists eavesdrop on brains.**
- **In order to understand the philosophical side of it, but also make you more prepared to get excited more responsibly (+ see cool research)**
- **Big Q: is 'the electrical language' (Lem) going to get is Rosetta stone anytime soon?**



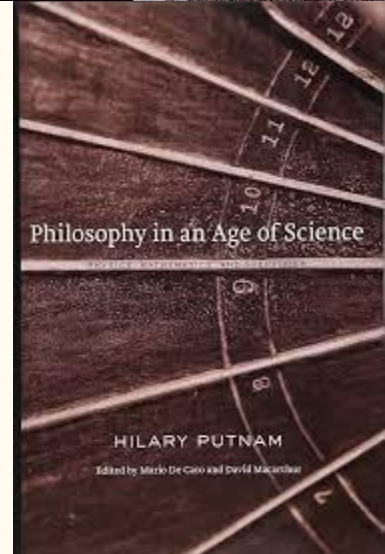
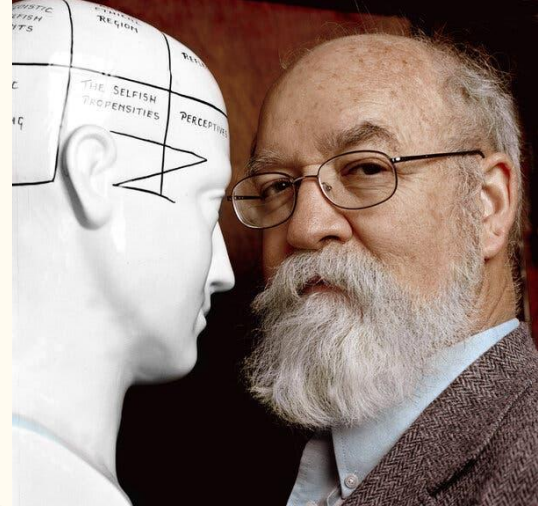
How to philosophize (or just - think) responsibly about it?

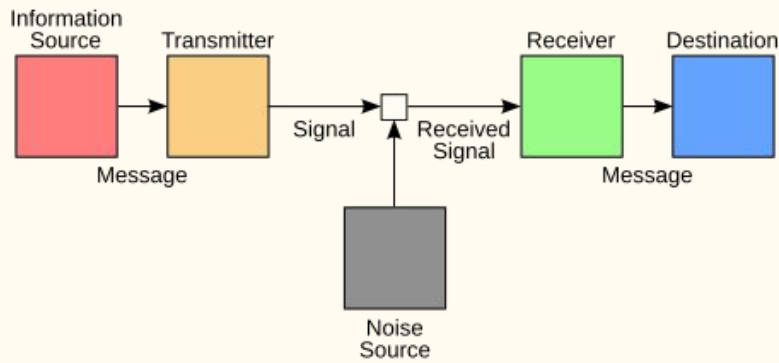
- Virginia Hughes, a reputable science journalist who appears disappointed with her peers for what she takes to be their gullibility in the matter, writes that “claims of having achieved neuroscientific mind reading are laughably false” (Hughes, 2011)
- So why do we constantly keep hearing about reaching the matrix via neuralink or some other bollocks? We’ll get to that, too.
- But what if we leave out the popular excitements and try to be rigorous philosophically about this?
- And what does it spell for CTM?

Some history - how did we get there?

- Traditional philosophy of mind used *armchair methods* to target cognitive processes and architecture of human mind. Plato, Aquinas, Kant - all in the business (even if the terminology changes)
- During 19th century - the breakaway of psychology as an empirical mode of tackling similar questions. Later: neuroscience. The emergence of information theory in the 1940s/50s. [Shannon 1948) - artificial systems enter the frame.
- Computerization - conceptual, then industrial (von Neumann's universal machine)
- Neural machinery as logical gates (McCulloch and Pitts; 1943) -> **neuroinformatics**
- **Cognitive Revolution of the 1960s+ - birth of Cognitive Science (CogSci)**
 - McCarthy, Minsky, Neisser, Newell, Simon, Chomsky
 - **Philosophers closing on: Putnam 'Philosophy in an Age of Science'; Kripke, Dretske, Dennett**

Q: So what's there for philosophy of mind then in an age of cognitive science? Weren't we superseded?

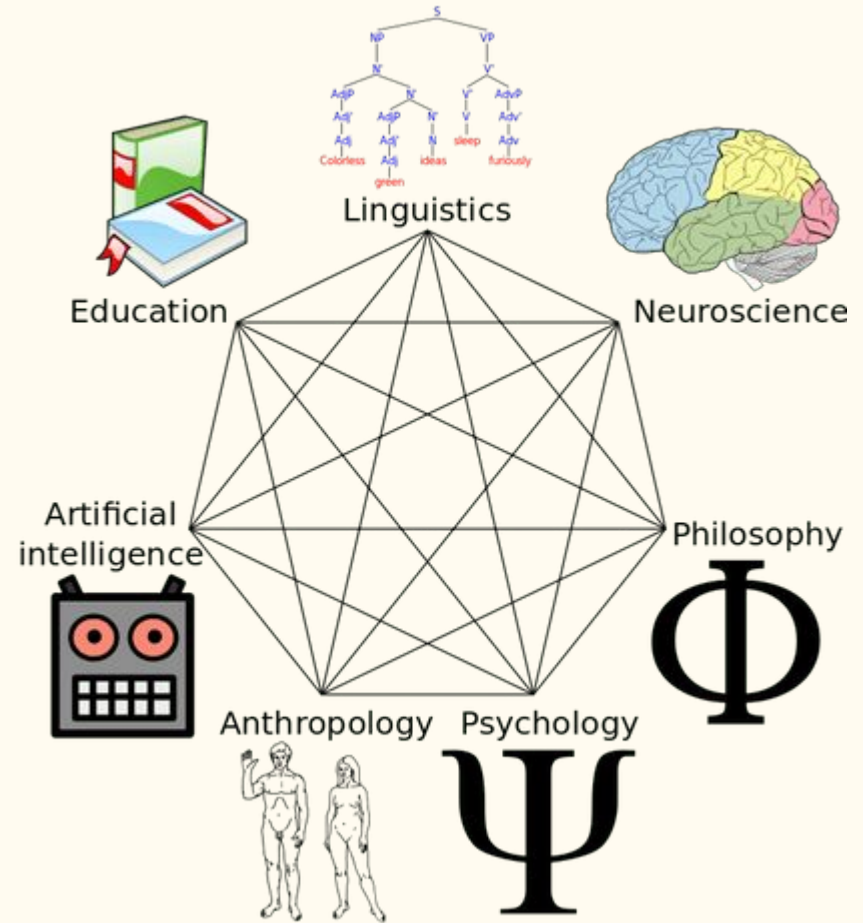
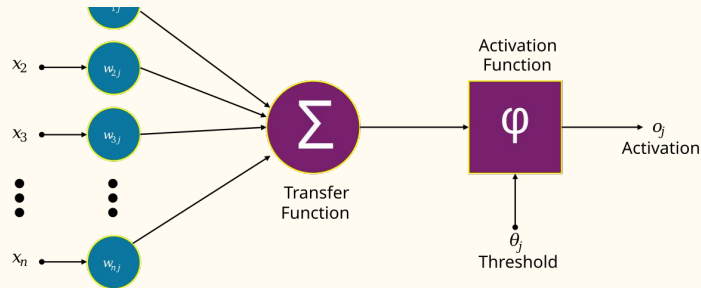




Top: diagram of Shannon's theory of information transfer

Bottom: McCulloch and Pitts' neuron

Right: 'map' of cognitive science as conceived during the foundational conferences of the 1970s



Naturalization of philosophy into CogSci

- Mind-brain identity theories + Post-war physicalism in vogue
- Even though they were eventually discarded, the idea that brains are information-processors remains virtually unquestioned and ubiquitous.
- Computational Theory of Mind (CTM; also known as Computationalism)



Cell Press

<https://www.cell.com/cancer-cell/pdf>

Can a biologist fix a radio?—Or, what I learned while ...

Y Lazebnik · 2002 · Cytowane przez 529 — I started to contemplate how **biologists** would determine why my **radio** does not work and how they would attempt to repair it. Because a majority of...

Explaining the Computational Mind



Marcin Milkowski

Computer metaphor or computationalist theory?

- Computationalism is sometimes bundled with ‘computer metaphor’ (Brette 2022) for the mind/brain, which posits that brains are computational machines *akin* to computers (but not necessarily *are* computers).
- But metaphors are different than explanatory theories! So, is it just a metaphor or a serious literal theory?
- All metaphors are wrong, but some are useful. Also, they are usually implicit (contrary to analogies)
- So, what are the main tenets of CTMs? There is *a lot* of variants. But generally:
 - Physicalism. Usually also reductionism but dualist positions also possible (hardware-software; see Wilder Penfield 1975)
 - Cognition is calculation (or more precisely: computation). Mental operations are defined over abstract symbols
 - Mental processes are algorithms that are instantiated in the neural tissue. Brains run programs.
- We can not only model, describe, approximate or simulate, but also *understand* the mind via understanding what computational mechanisms they realize.
- Theory is so much stronger than loose analogy metaphor. Because of that it’s also an easier target for criticisms (and that’s good!)

5 MINUTES AFTER INVENTION OF WHEEL

bro, literally tho bro, **the whole world is a wheel**, turning round and round, an endless cycle of seasons and lives and realms--it all keeps turning, and it all keeps coming back around



5 MINUTES AFTER INVENTION OF BOOKS

bro, literally tho bro, **the whole world is a book**, God is The Author of Being and he writes our names in the Book of Life. In the beginning was THE WORD, bro. You see bro??



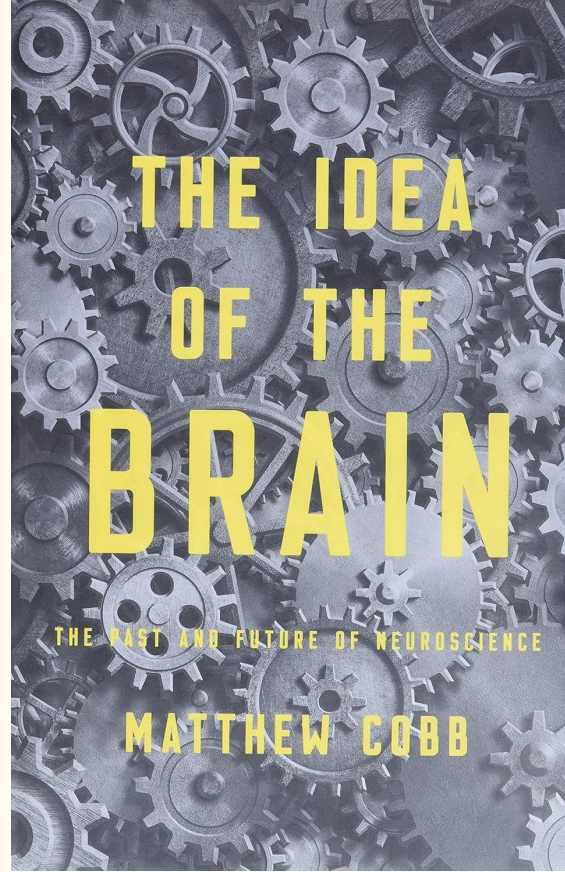
5 MINUTES AFTER INVENTION OF COMPLEX GEARS

bro, literally tho bro, **the whole world is mechanical**, God is The Divine Watchmaker, and the planets spin along predetermined paths according to the turning of cosmic gears! It's all just predictable mechanical paths from creation to here!



5 MINUTES AFTER INVENTION OF COMPUTERS

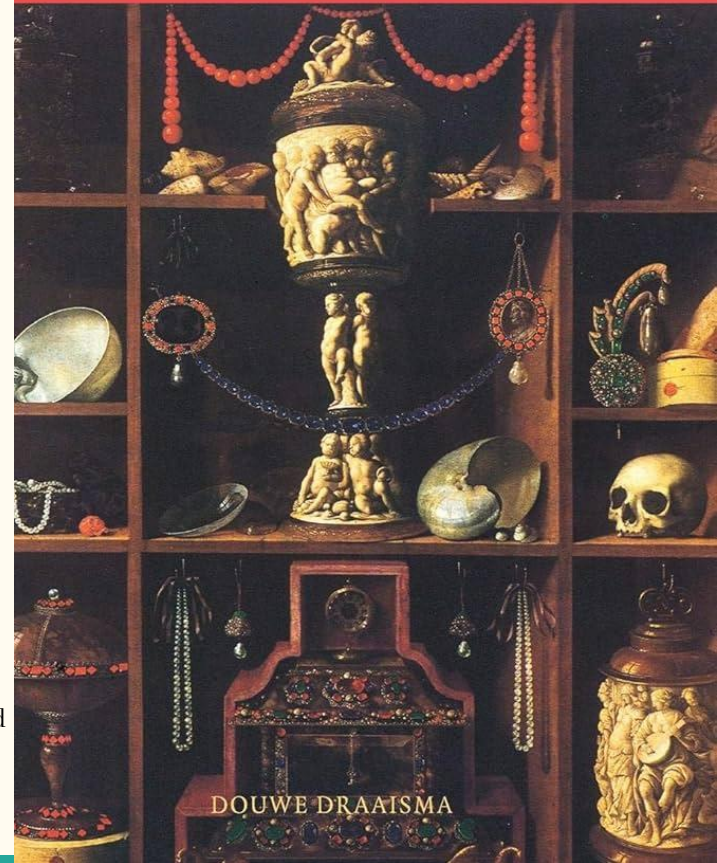
bro, i think i just figured something out-



Top: Cobb, M. (2020) The Idea of the Brain. The past and future of neuroscience.
Right: Draaisma, D. (2000), Metaphors of Memory. History of ideas about the mind.

Metaphors of Memory

A history of ideas about the mind

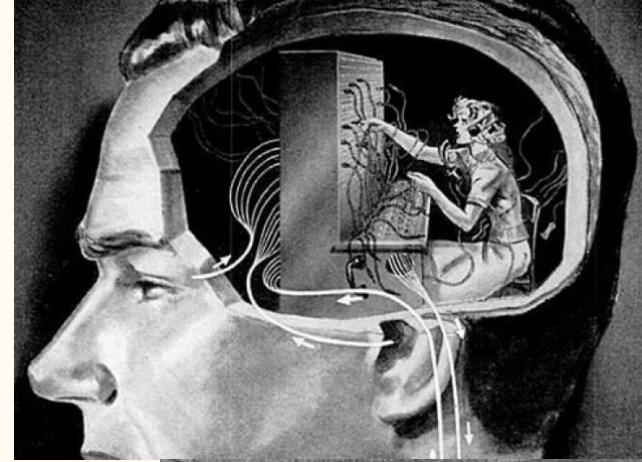


Critiques of computationalism (Turing 1950; Miłkowski 2017)

- Number-crunching argument
 - Computers just crunch numbers and that's not thinking
 - Stronger version: The Chinese Room Argument (Searle) - no semantics from syntax
- Simulation argument (also Searle)
 - Simulation is not recreation. Computational model of $X \neq X$. Can't drink simulated milk.
- Emotions argument (Ada Lovelace argument)
 - Machine can't feel (enter affective states, if one prefers). Nor they can create/experience art. And so on.
- Consciousness argument
 - Even if we can recreate some sort of awareness, then there is phenomenal experience of consciousness that is strictly non-reductive (Chalmers' Hard Problem)
- Frame problem
 - i.e. Background knowledge or *common sense*; Dreyfus 1978)
- Embodiment argument (-> 4E Cognition approaches)
- Externalist arguments
 - Meanings just ain't in the head. And brains can't be in vats. (Putnam)

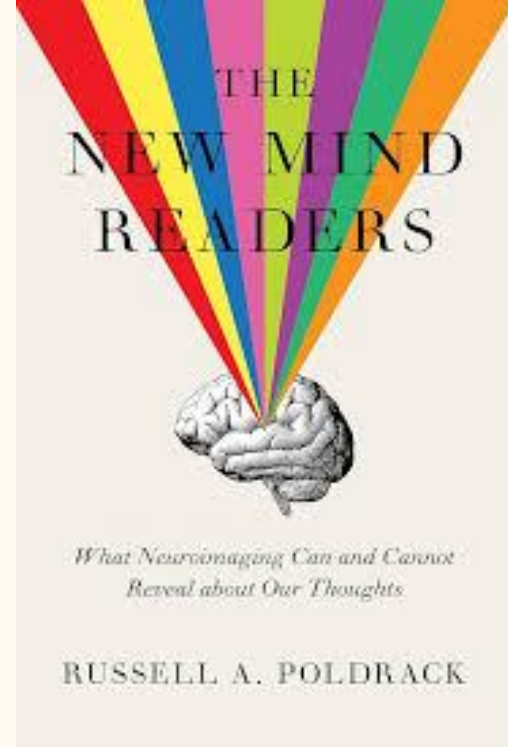
Homunculus' New Clothes - Critique

- If the mind is indeed an intricate code, who on earth interprets it? Is there a central commander in charge of the switchboard, listening to the messages, decoding them and sending other back?
- That can't be right say **'encodingist' critics (Mark Bickhard; Romain Brette 'coding metaphor')**. We have a very old monster - the homunculus in new clothes!
- And then what? Turtles all the way down? It's a road to nowhere. We just put the problem a level 'up'. Hence, CTM is an illusion, not an explanation.
- There is interesting 'symbol grounding problem' (Harnad), but solutions too (e.g. B. Skyrms)



Science of mindreading (Poldrack 2018)

- Sets the agenda for **neuroregistration** in the service of psychology
- We are hunting for content - the endgame is to **explain meaning-making**
- For that we need **cognitive ontologies of mental functions**. How to chart the domains of the mind (notice: that's still what Aristotle or Kant wanted!)? **Structurally or functionally?** But how to also escape 'modern phrenology' (Uttal 2003)
- Mental functions are poorly defined and not directly observable. Do we need new language? New notions? Poldrack thinks so, we use outdated 19th century terminology that can't be mapped to massive yet nuanced streams of data. **Should we go top-down or bottom-up?**
- Neural registrations on the other hand are blurry, messy and coarse grained. Tradeoffs on both temporal and spatial scales. Also a lot of methods (single cell, EEG, fMRI, LFP) but all have drawbacks



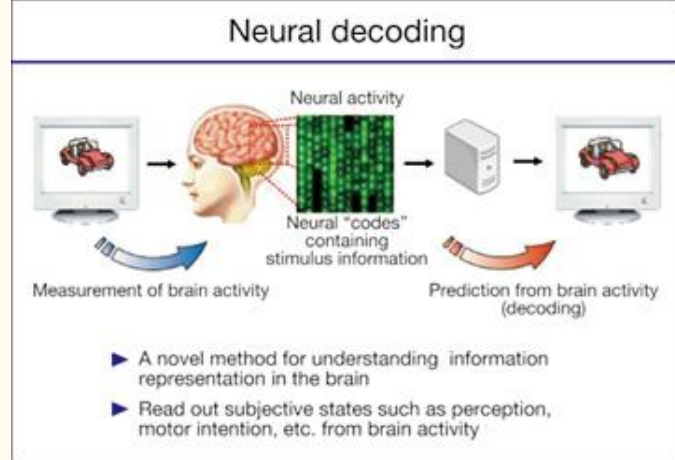
Cognitive Control and Action (BCIs)

- **Brain-computer interfaces:** pipelines for directly executing actions through mental activities through artificial signalling pathways
- Main use so far is therapeutic, not cognitive (or enhancing)
- It serves as a type of informational-bypassing for patients with serious neural damages (signalling pathways that were irrevocably damaged. e.g. spinal cord)
- It allows for control of actions in the environment, using specialized equipment, via brain functions. Patients are able to control cursors or robotic arms by imagining movement. Intention signals are recorded, interpreted and sent to connected devices to perform the intended action (be it grasping or playing on-screen chess).
- However, the pipelines are quite clumsy and unnatural for the subjects. Signals that we use are a signalling scheme far from the 'natural' mental currency, if there is one.
- It does not mean we 'cracked the code' of mental communication. We are 'just' able to translate some neural signals reliably into certain sorts of actions in limited environments.



Neural decoding (readout)

- Assumption - there is some **neural code** inside the brain
- But is this code some sort of ‘**mental currency**’?
Some researchers clearly think so and try to just find the implementations. Neuronal spikes (action potentials) are the most common suspect. But there are more (both bigger and smaller) candidates
- But to neuroscientists it is an empirical problem. There is however a bigger, conceptual one:
 - This huge problem is: **how do we know we infer based on this currency and not some proxy that just is in some statistical relation with ‘real’ vehicles?**
 - Information-for-us (external) may not necessarily be the information-in-the-system (internal) [Mamak (forthcoming)]
- If there is some code ‘in there’ in principle it should be possible to ‘jack into’ or ‘hack’ (that’s why you hear this word all the time)
- When you do that - fable has it - you will be able to decipher (‘read’) the mental *content* but also manipulate it (e.g. insert false memories - Porter 1999)

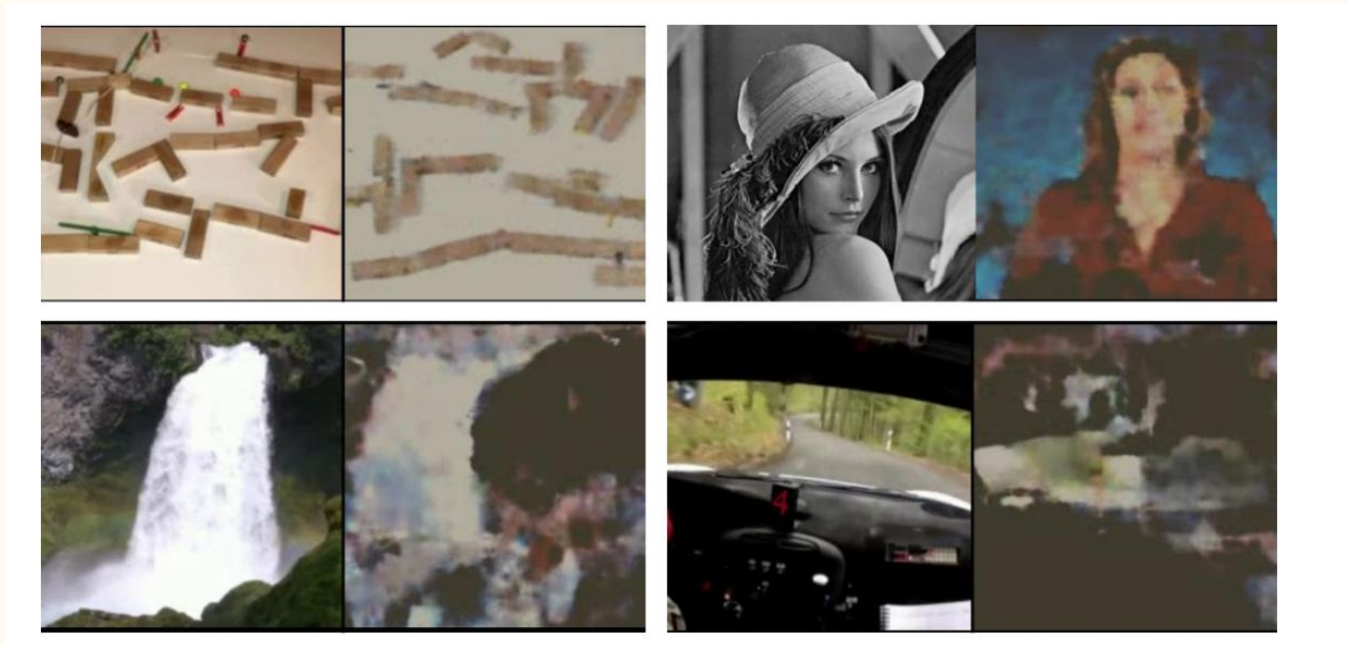


Reconstructing field of vision from brain data

- Short clips are shown to subjects. Then, ‘solely’ relying on brain data the researchers are able to reconstruct what the subjects are seeing with their eyes (caveat: prior training of the model is necessary to mediate these reconstruction)
- So the hype says: they are playing out the movie from the brain (a blurrier, but recognizable copy, anyways)!
- Now, a lot of similar studies use cutting edge sophisticated machine learning techniques to interpret high-dimensional neural data and reconstruct the signal back to ‘perceptual form’ (e.g. EEG - Lan et al. 2023; fMRI - Ren et al. 2021, also sounds (Kamitani lab). Newer pipelines work much better technically than in 2011. **But reality is still much more nuanced than they can detect and reconstruct!**
- It’s because there are important limitations to the entire process. For these reconstructions, researchers use dictionaries and ‘guess’ on the basis of what was used to construct them (train the models). So these ‘recordings’ are in fact only approximations, guesses or interpolation based on already ‘known’ (selected by researchers) data. Because of that the pipelines do not work equally well for any given arbitrary signal (e.g. movie). If the clip is far removed from what was available in the dictionary, they struggle.
- In other words, these studies **do not chart the entire realm of possible experience**, but a very, very tiny fragment of it. Impressive results you see are usually strictly curated and limited cases. **They are extremely impressive on its own. But it’s equally important to understand the context and what is really proven by these studies and do not overestimate and overhype their real results.**

Nishimoto et al. 2011 study; Rashkov et al. 2019

<https://www.youtube.com/watch?v=nsjDnYxJ0bo>



So-called ‘Lie detectors’

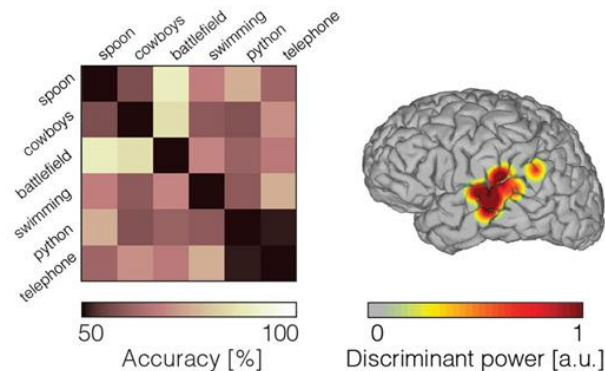
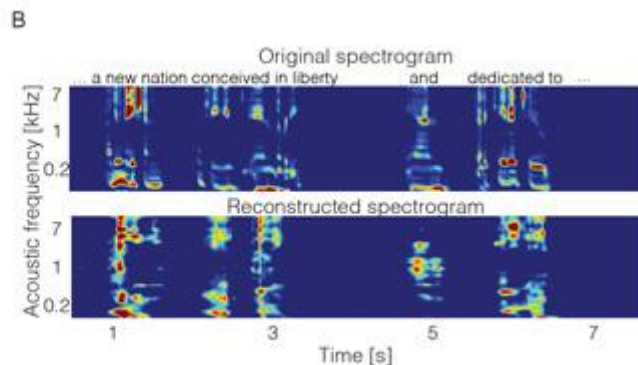
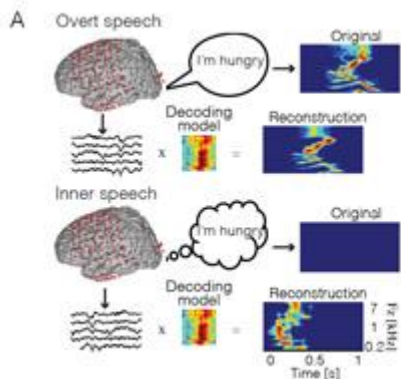
- 2nd generation ‘lie-detectors’. Contrary to polygraphs, they promise inferring the truthfulness of subjects not mediated by physiology, but by direct neuroregistration [e.g. No Lie fMRI - Daniel Langleben]
- Does it circumvent the problems and grants surefire access to the truth?
- Of course not. First, there are technical problems: our recording scale is messy and coarse-grained. It’s like fishing in a pond with an excavator so we still don’t measure real vehicles directly
- But even if the technology (scanners or data-processing) gets better, there are still conceptual problems, even if we ask yes/no questions.
 - Lies can be true statements.
 - Truthfulness is a three-way relation (object-statement-world), whereas we only ‘measure’ one
 - We also measure only correlates. And these are only proxies.
- The idea of ‘finding’ falsehoods directly in the brain is delusional.
- Pushback from the legal world has been warranted.

Device Decodes 'Internal Speech' in the Brain

16 maj 2024 — Scientists have developed **brain implants that can decode internal speech** — identifying words that two people spoke in their minds without moving their lips or ...

Prediction

- We can reliably predict patterns of neural data from both previous neural recordings, but also behaviour
- Two-way inferences: neural recordings from mental functions, but also in reverse (called: *reverse inference*, Poldrack 2011)
- Example of inner speech reconstructions (Martin 2017). Works so far only on single words & best studies present only 6 words to differentiate from. Far cry from futuristic scenarios you still see!



Intervention/Manipulation

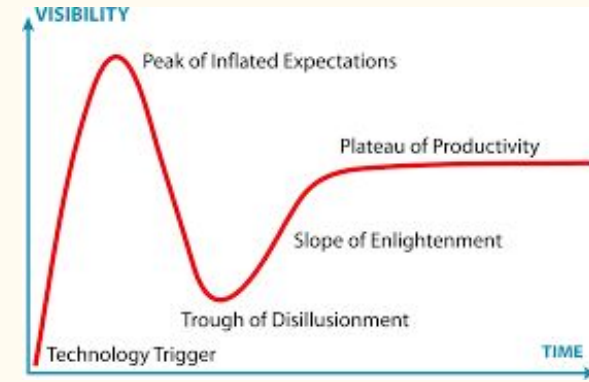
- ‘Facephenes and rainbows’ study (Schalk et al. 2017)
- Electrical stimulation of two individually-specialized regions: face-area (FFA) and colour-area
- Subject reports seeing induced faces on objects when face-area is stimulated and rainbows when colour area is externally excited by researchers
- Much better than readout - gives insight into what type of information is **‘exploited’, i.e. causally active in practice.**
- This is what we call **‘manipulability’** - a desirable trait in any mature scientific theory
- **Optogenetics** (think installing a controllable switch-off or switch-on inside the very small brain part) might be the way forward here

Schalk et al. 2017 study

<https://www.youtube.com/watch?v=nHigKvuLAAy>

Beware the hype (but get excited anyway)

- **Neuro-nonsense:** rampant in coaching, marketing, business ventures.
- There are reasons for that - real-world incentives to overblow it. From vanity to stock prices.
- Media cycle is also at fault. Sensationalization, clickbaity headlining, sci-fi fantasizing and lazy futurism are all common when reporting on breakthrough neurostudies.
- But there is also justified hunger for **unificatory explanations** - and cognitive science (and AI) looks at 'information' as this central bridge concept.
- And for good reasons - CTM **is and has been** a prolific and fruitful *research tradition*. Will it be the last? Probably not, if you believe in (pessimistic) induction from history of science (see: Larry Laudan)



The Gartner hype curve



Nature

<https://www.nature.com> › articles › Tłumaczenie strony

Mind-reading devices are revealing the brain's secrets

ORK TIME · 2024 — **Mind-reading devices** are revealing the brain's secrets. Im technologies that decode neural activity can restore people's ...



Vox

<https://www.vox.com> › neurotec... › Tłumaczenie strony

Mind-reading technology has arrived

4 maj 2023 — **Mind-reading** technology has arrived. An AI-powered “brain decod thoughts with surprising accuracy.



New Statesman

<https://www.newstatesman.com> › ... › Tłumaczenie strony

Are you ready for Elon Musk to read your mind?

30 sty 2024 — In September, Elon Musk's **Neuralink** announced it was recruiting for its firs to implant a BCI into the brains of people who are ...



New Scientist

<https://www.newscientist.com> › 2... › Tłumaczenie strony

Mind-reading AI can translate brainwaves into written text

12 gru 2023 — A system that records the **brain's** electrical activity through the sca into words with help from a large language model ...



Futurism

<https://futurism.com> › neoscope › Tłumaczenie strony

First Neuralink Patient Using It to Learn New Languages

29 sie 2024 — In a tweet, **Neuralink** patient Noland Arbaugh revealed that he's spending roughly three hours a day **learning** French and Japanese.

Więcej pytań

Can Neuralink make you learn languages?



Reddit

<https://www.reddit.com> › Neuralink › comments › will_...

Will Neuralink make language learning obsolete?

15 lip 2020 — There won't be a need for **languages** as we know them anymore, and we'll be able to communicate ideas, senses, and imagery in a deeper, more meaningful way.

elon musk says **neuralink** brain chip will make **language** ... 9 cze 2021

Will **Language Learning** become obsolete? : r/languagelearning 13 paź 2022

What is the future of **language learning**? - Reddit 7 wrz 2023

Will **neuralink** help with **learning** concepts and memory like ... 26 paź 2019

Więcej wyników ze strony www.reddit.com



thetesspace.com

<https://www.thetesspace.com> › la... › Tłumaczenie strony

How Neuralink Will Make Language Obsolete..

20 maj 2024 — Elon Musk's **Neuralink** brain implant could very well become the future of communications. And to understand how exactly that might take place ...

Mindreading

- Do not confuse with mindreading from anthropology (as inferring mental states from behaviour as mentalists and... monkeys routinely do)
- What we have in mind - is **non-mediated, direct inference of mental content on the basis of some neuroregistration. We circumvent the ‘regular’ channels of communication (e.g. non-verbal cues, text, signs), but also we decode in line with the ‘actual content’** (but not according to public conventions like language).
- *Typically, the term “mind reading” is used to denote a certain achievement that has not yet been made, but which is likely to be made in the near future, given the pace of innovation in imaging technology (Peckham, 2011). This usage raises an obvious pair of questions: (i) what exactly would the predicted achievement look like? And (ii) how will we recognize it when it comes?*

Four-ingredient recipe to read a mind

- ❖ ‘On the view we have [R&H&H] developed, neuroscientific mind reading is
 - (i) discerning mental content [**decodability - labels are mine; WM**]
 - (ii) from a prediction about some property of an experimental condition, where the prediction [**forwardness**]
 - (iii) does not, during testing, capitalize on the intentional production of conventional symbols by the subject [**non-conventionality**] and
 - (iv) is computed by a prediction algorithm that takes exclusively neural data as input. [**directedness**]
- ❖ One implication of this view is that neuroscientific mind reading has, in fact, already been accomplished.’
- ❖ However, these results are still of limited scope and it is irresponsible to generalize it into a full-blown theory of the mind. We are very far off still.

Conclusions & Lessons

- In science, we search for **explanations** and these are more than *mere descriptions*.
- Also pragmatic success (even when super impressive) might not be always evidence enough that we have sufficient theoretical understanding. So mindreading needs CTM.
- **And neural decoding techniques do not give direct access to ‘mental currency’ and they are heavily assisted. Always be wary of that and avoid too simplistic interpretations.**
- **But in *limited* scenarios, we already succeeded in mindreading. And this is a huge win for CTM, but not the completed job. It is not even close to exhausting the totality of experience.**
- **Corollary: there is and will be important work for philosophers in this field.** Checking the validity of statements. Providing conceptual analysis and tests (sets of criteria) for theories, hypotheses and tools. Problems in the field are not just definitional but there is still pressing need to keep defining things properly (descriptively and normatively). Cited Rathkopf’s work is exemplary here.
- There are still major obstacles for CTM:
 - Embodiment
 - Homuncularism + externalism (problem with the encoding/decoding picture)
 - The unfinished nature of computer metaphor
 - Proof is in the pudding, but:
 - **Still, it’s the only game in town (Fodor’s catchphrase)!**

It’s a mature theory that checks a lot of boxes + the empirical success is astounding!

Warszawa. Pomnik Kopernika
z widokiem na b. pałac Staszica.

Thank you for coming and for your attention.

Any questions welcome! If any of the references
grabbed your interest, please do ask, I'll be glad to
provide the bibliography!