Introduction to Cognitive Science



Dr. Mark Miller

MONASH University · UNIVERSITY OF TORONTO · CHAIN

Taxonomy of positions on the mind-body relation

Monism: There is only one fundamental type of stuff in the world, and both 'minds' and 'bodies' are versions of this stuff.

Idealism: Everything is mental, and matter exists only as perceptions in different minds.

'Berkeleyan' idealism: The only minds are human minds (and perhaps God's mind); nonhuman things exist only as perceived by humans (and God).

Panpsychist idealism: *Everything* has a mind of some sort, both human and nonhuman.

Physicalism (*a.k.a.,* materialism): Everything is physical, and minds are nothing over and above a particular type of physical thing.

Behaviourism: Each mental state is a disposition to behave in certain ways under certain circumstances.

Identity theory (*a.k.a.*, brain state theory): Each mental state is identical to a certain brain state.

Functionalism: Each mental state is a particular structure of causes and effects.

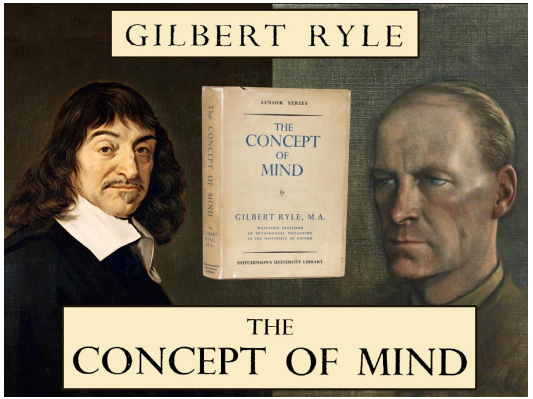
Computational (or 'machine') functionalism

Neutral monism: The basic stuff of reality is somehow 'in between' mental and physical, and can appear as one or the other in different contexts.









Ryle argues that this entire project is premised on a fundamental error – namely, **the assumption that 'mind' refers to a certain kind of *entity***, when actually it belongs to a different logical category. Ryle calls this a *category mistake:*

"It represents the facts of mental life as if they belonged to one logical type or category (or range of types or categories) when they actually belong to another".





What is Ryle's alternative?

Ryle thought of mental states as falling into two categories: dispositional and occurrent states. He arguably had better worked-out his views concerning dispositional states than he had about occurrent states. About the former, Ryle writes:

"I have already had occasion to argue that a number of words which we commonly use to describe and explain people's behavior signify dispositions and not episodes. To say that a person knows something, or aspires to be something, is not to say that he is at a particular moment in a process of doing or undergoing anything, but that he is prone to do certain things, when the need arises, or that he is prone to do and feel certain things in situations of certain sorts. This is, in itself, hardly more than a dull fact (almost) of ordinary grammar. The verbs 'know', 'possess' and 'aspire' do not behave like the verbs 'run', 'wake up', or 'tingle'; we cannot say 'he knew so and so for two minutes, then stopped and started again after a breather', 'he gradually aspired to be a bishop', or 'he is now engaged in possessing a bicycle.'" (Concept of Mind, 116)

What is Ryle's alternative?

If describing the gradual development or evolution of a mental state sounds odd in many cases, then the state is likely to be a standing mental state and, so, dispositional rather than occurrent. This seems to be the case with propositional attitudes like belief. According to Ryle to be in a dispositional mental state like belief is to be disposed, or (as Ryle says) 'prone', to behave in certain ways under certain circumstances. Given that beliefs and other attitudes are dispositional rather than occurrent, we should be able to analyze them in terms of complex behavioural dispositions. This offers an analysis of a large class of mental properties.



Behaviourism

As an analogy, consider the property of *fragility*. This is a dispositional property of an object that tells us about how it is prone, or disposed, to behave under various counterfactual conditions (e.g., that it is prone to shatter if dropped on a hard floor).

Similarly, to attribute a certain attitude to someone (e.g., the belief that it's raining today) is to say that they disposed, among other things, to wear a raincoat if they go outside, to take an umbrella with them, to reschedule their plan to go sit on the beach with friends, etc.

Behaviourism is the project of analyzing mental states exhaustively in terms of actual and possible behaviour. In a slogan, **mental properties are nothing but *behavioural properties*.**



Objections to behaviourism (1)

Take the mental state of believing that it's raining outside. Earlier, we suggested that we could analyze this as the disposition to, among other things, take an umbrella with one when one leaves the house, to cancel one's plans at the beach, etc.

However, someone who believes that it's raining outside will be disposed to behave in these ways only if they *want* to stay dry and *believe* that taking the umbrella with them (e.g.) provides a way to stay dry, etc.



Objections to behaviourism (1)

This suggests that which behavioral dispositions a belief is associate with depends on what other mental states the believer possesses.

Parallel points apply to behaviourist analyses of desire and other dispositional states. If this is right, then it seems we cannot define the conditions under which a given mental state will manifest in certain behaviours without *presupposing* many other mental states. In that case, we have analyzed mental states in terms of … mental states!

Objections to behaviourism (2)



Another class of objections comes from Hilary Putnam's thought experiment of 'super Spartans' (or 'super stoics') and 'super actors'. The former offers a counterexample to the behaviourist's claim that a certain behavioural disposition is *necessary* for the instantiation of pain; the latter provides a counterexample to the claim that a certain behavioural disposition is sufficient for the instantiation of pain.

Note: arguably, the vividness of Putnam's super Spartan/super actor cases comes from his choice to focus on the mental state of *pain* – an episodic mental state for which a purely dispositional analysis seems especially unlikely to succeed. However, we can ask whether the same argumentative strategy can be extended to other types of mental states, including beliefs, desires, etc.

Note: Putnam's cases might be viewed as merely a special case of Chisholm's observation that a person's behavioural dispositions depend on an entire *group* of mental states.

Objections to behaviourism (3)



Ryle objected to the official doctrine (i.e., Descartes') partly on the grounds that it failed to accommodate mental-physical causal interaction. But does Ryle's behaviourism fare any better at capturing this intuitive datum? We ordinarily think of mental states as *causing* our actions – as when one reaches for a glass because one wants to take a sip of water (and believes there is a glass of water there, etc.). But do we capture that intuitive datum if we regard mental states as merely clusters of behaviour, or as dispositions to behave in various ways? Compare: the glass of water shattered when it hit the ground because it is prone to shatter when its hits a hard surface with a certain force. This explanation of the glass-breaking does provide *some* information; but does it seem like the *same kind* of explanation of behaviour that we get when we cite the mental states that motivate and rationalize a person's behaviour?



Wrapping up

For these and other reasons, many philosophers have doubted whether logical behaviourism can provide a fully adequate theory of mental states (even of ‘standing’ mental states like belief, intention, etc.).

Nevertheless, many have wanted to preserve Ryle’s idea that behavioural dispositions are (at least partly) constitutive of mental states, even if behaviourism, by itself, is not fully adequate to capture that insight.


ANY
QUESTIONS?

Identity Theory

a.k.a.
Brain State Theory





Identity theory (a.k.a. brain state theory)

"Let me first try to state more accurately the thesis that sensations are brain processes. It is not the thesis that, for example, 'after-image' or 'ache' means the same as 'brain process of sort X' . . . It is that, in so far as 'after-image' or 'ache' is a report of a process, it is a report of a process which *happens to be* a brain process . . . All it claims is that in so far as a sensation statement is a report of something, that something is in fact a brain process. Sensations are nothing over and above brain processes . . . When I say that a sensation is a brain process … I am using 'is' in the sense of strict identity. . . . I do not mean just that the sensation is somehow spatially or temporally continuous with the brain process." (Smart, pp. 144-5)



Identity theory (a.k.a. brain state theory)

Basic idea: every mental state *is identical to* a physical state of the brain.

Some clarifications:

i) The identity theory concerns *mental states,* where a 'state' is an object's possessing a property or feature. Focusing on mental states lets us side-step the question of what a 'mind' is (though one natural proposal is that the mind is the brain, or at least some part of it).

ii) The identity theory identifies *types* of mental state (e.g., anger, pain, belief) with *types* of brain state (e.g. activation of some neural structure or other).

Stock example: pain = C-fiber activation

Argument from the unification with science

Smart believes that the identity theory, as contrasted with its main rival (dualism), best comports with the established practice of empirical science.

In particular, identity theory follows an established theoretical practice of inferring the identity of regularly observed, correlated phenomena …



Argument from unification with science

Consider how scientific theory frequently moves from an observed *correlation* between two phenomena to the *identification* of those two phenomena:

Observation: what we call ‘water’ *correlate* with a chemical compound, H_2O.

Inference: water *just is* the chemical compound, H_2O.

Note: the posited identity here is strict *numerical* identity between the type ‘water’ and the type ‘H_2O’.

Argument from unification with science

Consider how scientific theory frequently moves from an observed correlation between two phenomena to the identification of those two phenomena:

Observation: instances of what we call ‘lightning’ *correlate* with electrical discharge between clouds or clouds and the ground.

Inference: lightning *just is* electrical discharge between clouds and the ground

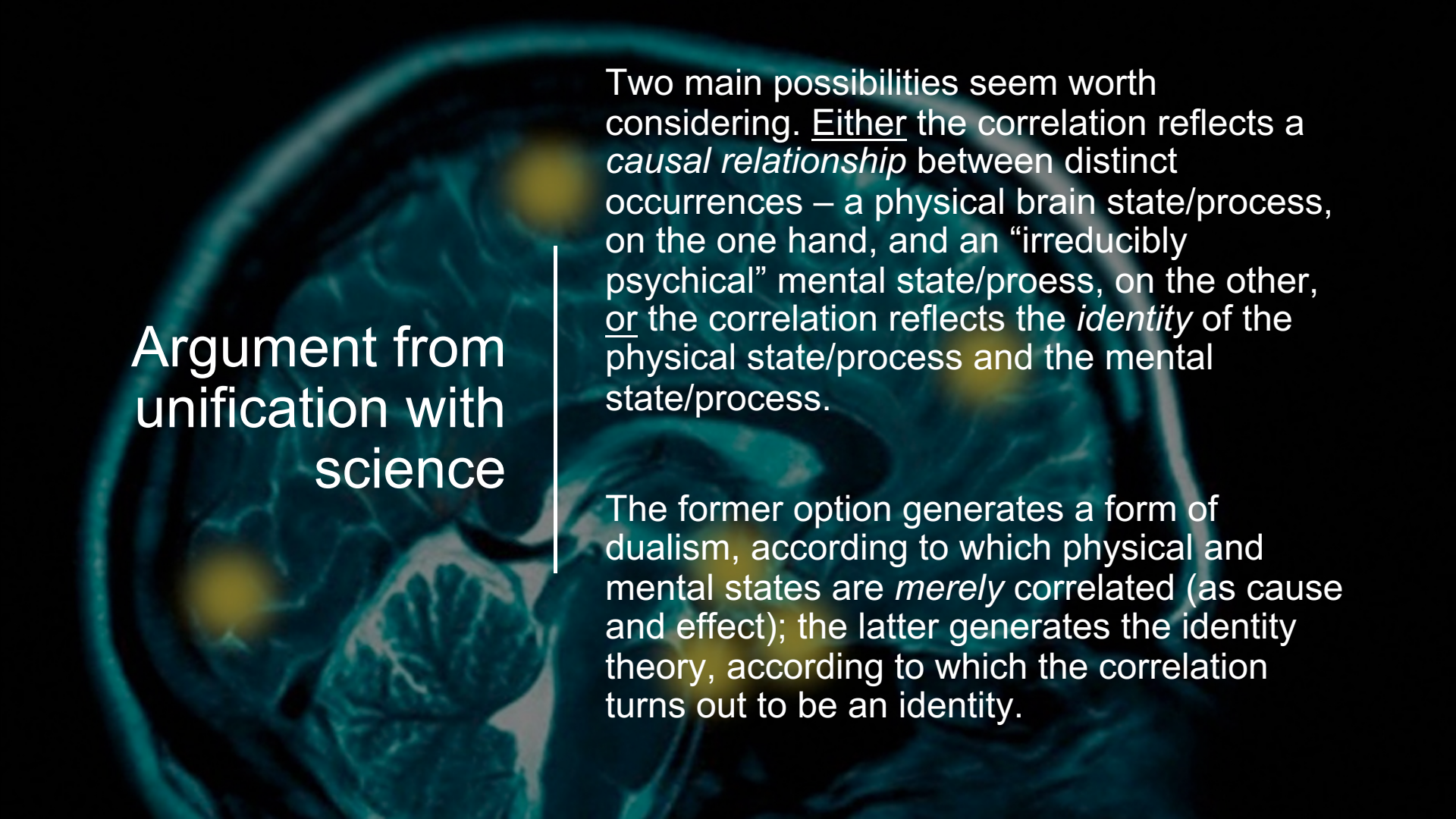
Note: the posited identity here is strict *numerical* identity between the type ‘lighting’ and the type ‘electrical discharge’.

Argument from unification with science

Now consider mind and brain: what sorts of observational evidence do we have?

- *Behavioral observation plus knowledge of physiology*: At least to date, any creature to whom we attribute mental life has a complex neural structure: a brain.
- *Scientifically observed mind-brain correlations*: We are learning more and more about the nervous system, and also discovering greater and more systematic *correlations* between types of mental state and types of brain state.

Question: what should we make of these mental-physical correlations?



Argument from unification with science

Two main possibilities seem worth considering. Either the correlation reflects a *causal relationship* between distinct occurrences – a physical brain state/process, on the one hand, and an "irreducibly psychical" mental state/proess, on the other, or the correlation reflects the *identity* of the physical state/process and the mental state/process.

The former option generates a form of dualism, according to which physical and mental states are *merely* correlated (as cause and effect); the latter generates the identity theory, according to which the correlation turns out to be an identity.

Argument from unification with science

Smart thinks that we should draw the same inference from observed mental-physical correlations that scientists drew in the cases of water-H_2O and lightning-electrical discharge. In all three cases, Smart claims, we should infer that what might initially have seemed to be two distinct (but correlated) phenomena turns out to be one and the same.

But now the question arises of *why* we should prefer Smart's proposed explanation of the correlation against the dualist's explanation.

This takes us to Smart's main argument …

Argument from parsimony

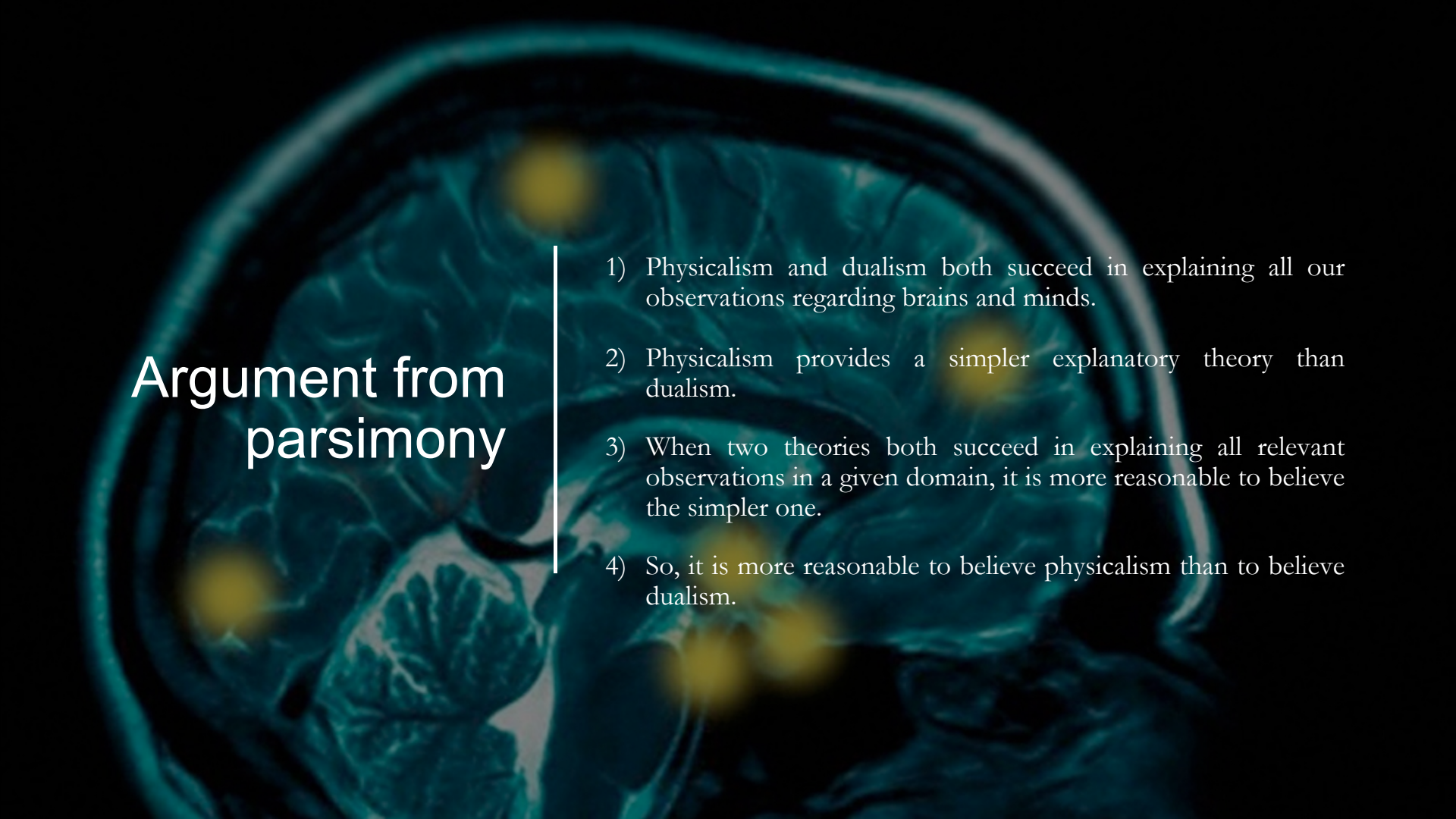
"Why do I resist this [the dualist's] suggestion? Mainly because of Occam's razor. It seems to me that science is increasingly giving us a viewpoint whereby organisms are able to be seen as physico-chemical mechanisms: it seems that even behavior of man himself will one day be explicable in mechanistic terms … All except for one place: in consciousness … for a full description of what is going on in a man you would have to mention not only the physical processes in his tissue, glands, nervous system, and so forth, but also his states of consciousness: his visual, auditory, and tactual sensations, his aches and pains. That these should be *correlated* with brain processes does not help, for to say that they are *correlated* is to say that they are something 'over and above'" … That everything should be explicable in terms of physics except the occurrence of sensations seems to me frankly unbelievable Such sensations would be 'nomological danglers'…" (Smart 142).

Argument from parsimony

In this passage, Smart contrasts two pictures of the world:

- **The Physicalist Hypothesis:** There is only one type of stuff, that described by physics, and this stuff has only one set of fundamental properties, those described by physics.
- **The Dualist Hypothesis:** Physical stuff with physical properties accounts for everything in the world except mental states like sensations. Sensations are correlated with certain (highly complex!) brain states, but the two are not the same thing.

Smart "wishes to resist" the dualist hypothesis, because of "Occam's razor" that is, because the physicalist picture provides the "simpler", more parsimonious explanation, and we should prefer theories that do that. We can put the argument as follows:



Argument from parsimony

1) Physicalism and dualism both succeed in explaining all our observations regarding brains and minds.

2) Physicalism provides a simpler explanatory theory than dualism.

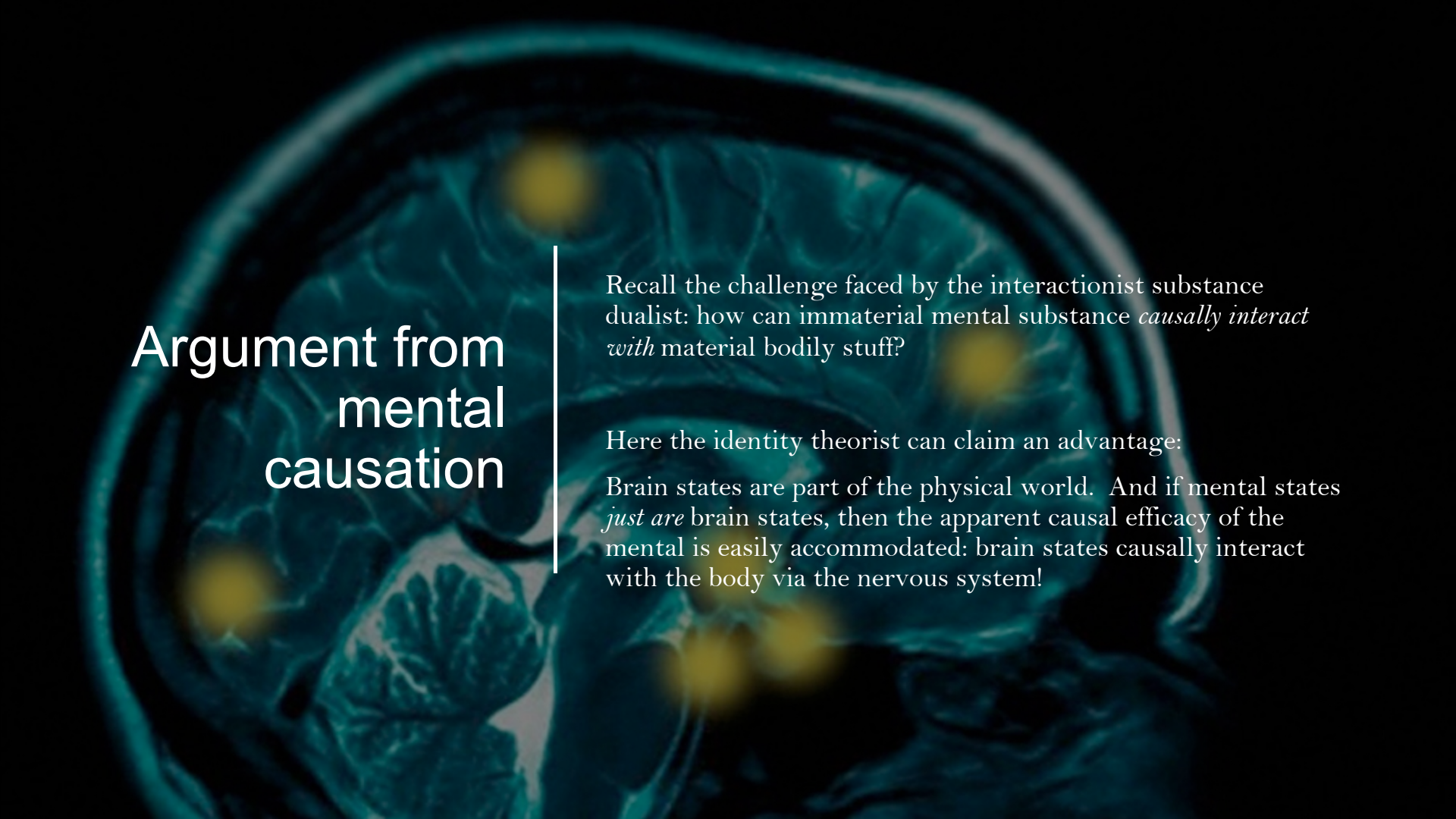
3) When two theories both succeed in explaining all relevant observations in a given domain, it is more reasonable to believe the simpler one.

4) So, it is more reasonable to believe physicalism than to believe dualism.

Argument from parsimony

At the end of his paper, Smart offers what he thinks is an analogous theoretical decision that confronted geologists in the 19th century. We observe that there are lots of strange fossils in very deep rocks, and also that minerals can be deposited in objects by water. This is compatible with the theory that there were strange animals a long time ago, whose bones fossilised; but it's also compatible with the theory that these fossils came into existence all at once in 4004 BCE at the creation of the universe (according to the timeline described in *Genesis*). As Smart writes:

"No scientist would ever entertain this as a serious hypothesis, consistent though it is with all possible evidence. The hypothesis offends against the principles of parsimony and simplicity. There would be far too many brute and inexplicable facts. Why are pterodactyl bones just as they are? No explanation in terms of the evolution of pterodactyls from earlier forms of life would any longer be possible. We would have millions of facts about the world as it was in 4004 B.C. that just have to be accepted." (pp. 155-6).



Argument from mental causation

Recall the challenge faced by the interactionist substance dualist: how can immaterial mental substance *causally interact with* material bodily stuff?

Here the identity theorist can claim an advantage:

Brain states are part of the physical world. And if mental states *just are* brain states, then the apparent causal efficacy of the mental is easily accommodated: brain states causally interact with the body via the nervous system!


ANY
QUESTIONS?

Multiple realizability: A (genuine!) challenge for identity theory



Multiple realizability: A challenge for identity theory

We can put this argument in the form of a ‘reductio ad absurdum’:

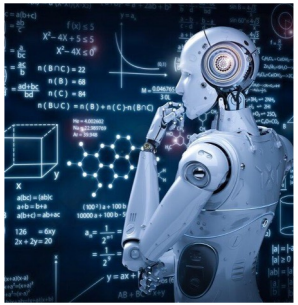
Premise 1: There could be alien creatures with mental lives similar to ours.

Premise 2: These creatures might be so different from us physiologically that no brain-state identifiable in us can occur in them.

Assumption for reductio: The brain-state theory is correct, and each mental state is a certain identifiable brain state.

Conclusion 1: These alien creatures do not have any mental state that we also have. (from **Premise 2** and **Assumption**)

Conclusion 2: The brain-state theory is false. (from the contradiction between **Premise 1** and **Conclusion 1**).

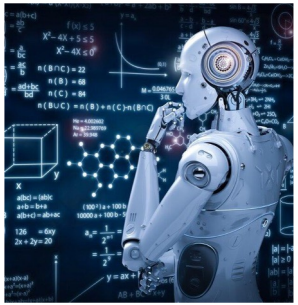






Multiple realizability: A challenge for identity theory

"Consider what the brain-state theorist has to do to make good on his claims. He has to specify a physical-chemical state such that *any* organism (not just a mammal) is in pain if and only if (a) it possesses a brain of suitable physical-chemical structure; and (b) its brain is in that physical-chemical state. This means that the physical-chemical state in question must be a possible physical state of a mammalian brain, a reptilian brain, a mollusc's brain…etc. At the same time, it must *not* be a possible (physically possible) state of the brain of any physically possible creature that cannot feel pain. Even if such a state can be found, it must be nomologically certain that it will also be a state of the brain of any extra-terrrestrial life that may be found that will be capable of feeling pain before we can even entertain the supposition that it may *be* pain" (Putnam, 'The Nature of Mental States', 77).

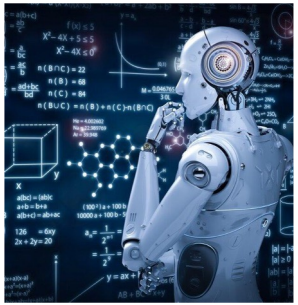






Multiple realizability: A challenge for identity theory

"…the hypothesis becomes even more ambitious when we realize that the brain-state theorist is not just saying that *pain* is a brain state; he is, of course, concerned to maintain that *every* psychological state is a brain state. Thus if we can find even one psychological predicate which can clearly be applied to both a mammal and an octopus…, but whose physical-chemical 'correlate' is different in the two cases, the brain-state theory has collapsed" (Putnam 77).



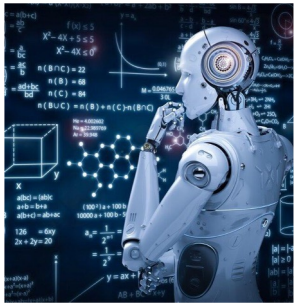




Multiple realizability: A challenge for identity theory

Multiple realizability: Mental states can be *realized* by different physical processes or structures.

This is what the thought experiments teach us: it is possible for a human, a Martian, and an octopus to all experience the same *type* of mental state (e.g. PAIN) in spite of significant physical differences (such that they could not be in the same *type* of brain state).



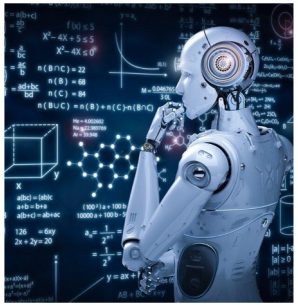




Multiple realizability: A challenge for identity theory

The possibility of multiple realizability sheds significant doubt upon the ultimate success of identity theory. This worry can be expressed as a:

> *Conceptual worry*: Agents with the same types of mental states but differing brains are easily conceived.
>
> *Empirical worry*: It seems "overwhelmingly likely" that we will discover *one* creature with the same mental state type but distinct neural structures. So identity theory looks doomed as a science of the mind.

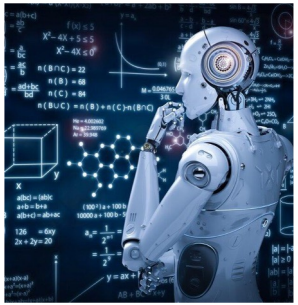






Multiple realizability: A challenge for identity theory

Question: do you understand why the identity theory is unable to accommodate multiple realizability? Do you find this argument against identity theory convincing?

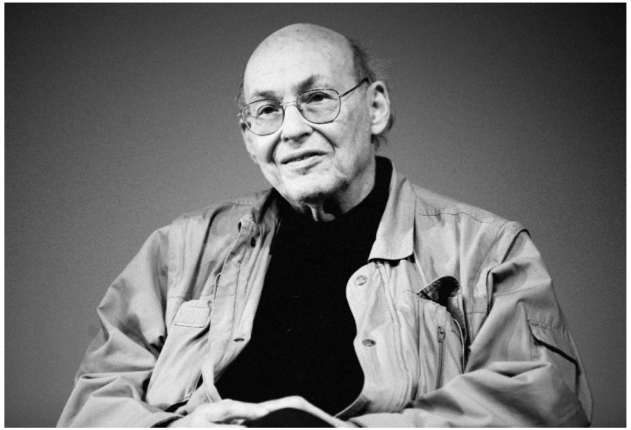


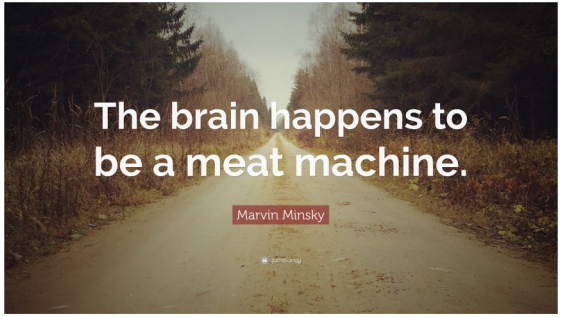


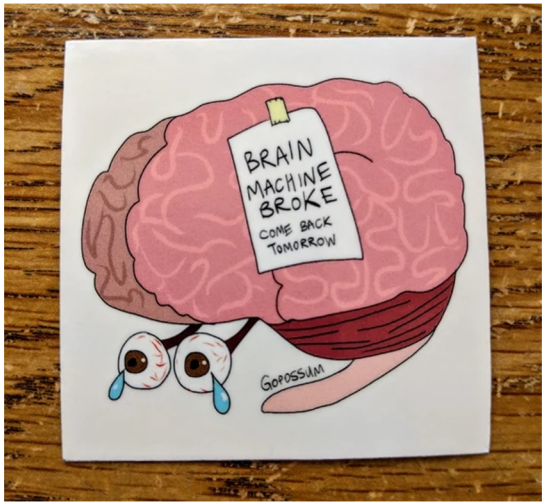


More?



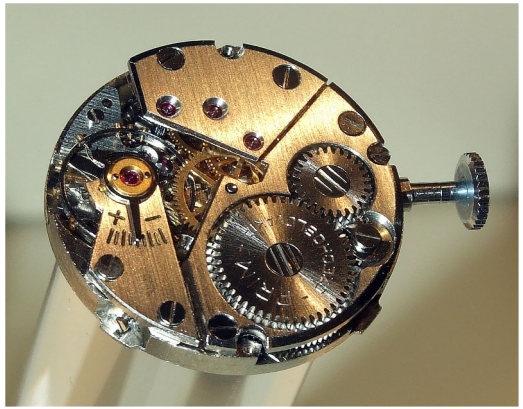


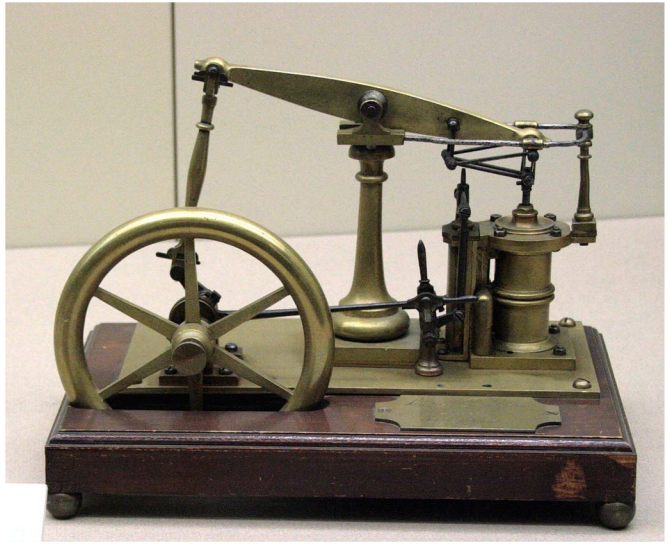


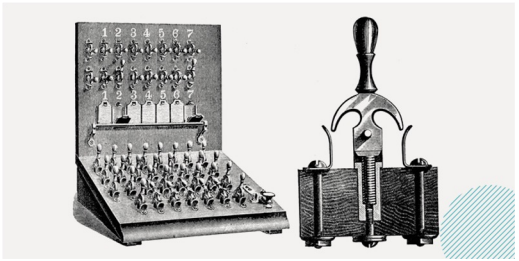


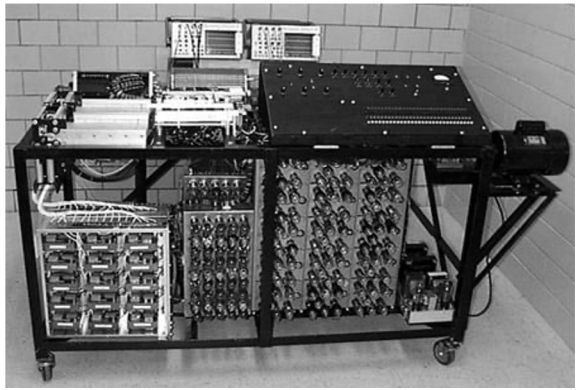
Minsky's description emphasizes that what matters isn't only (or primarily) the material (the "meat"), but how it's organized to support thinking - the machine part - the information processing itself.

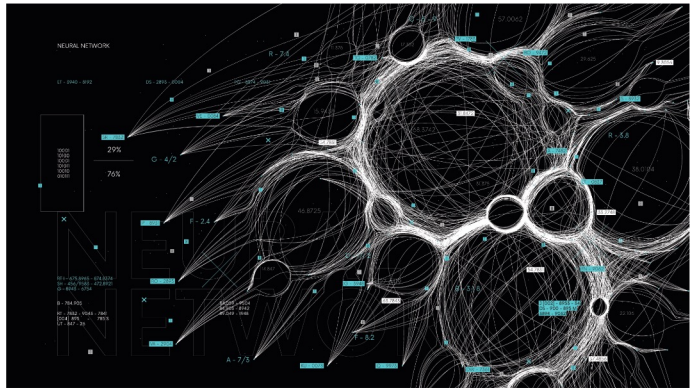






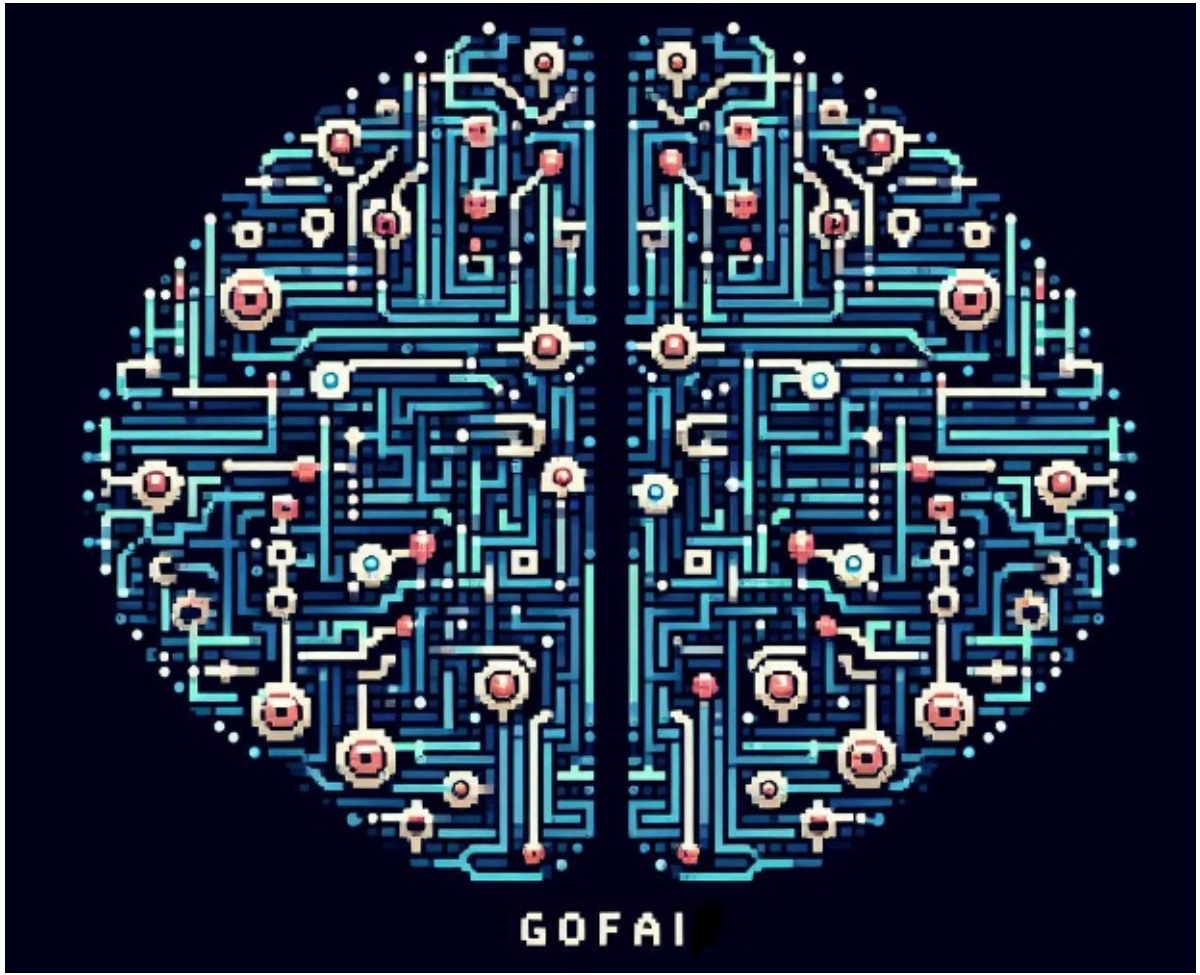


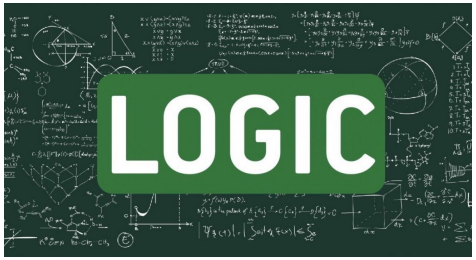



NEURAL NETWORK

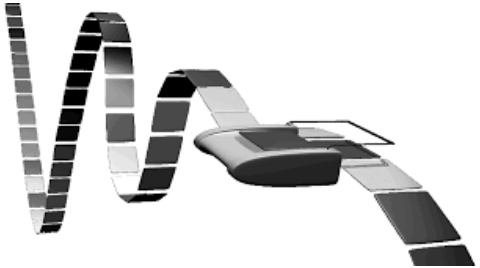
First major paradigm in cognitive science and artificial intelligence.

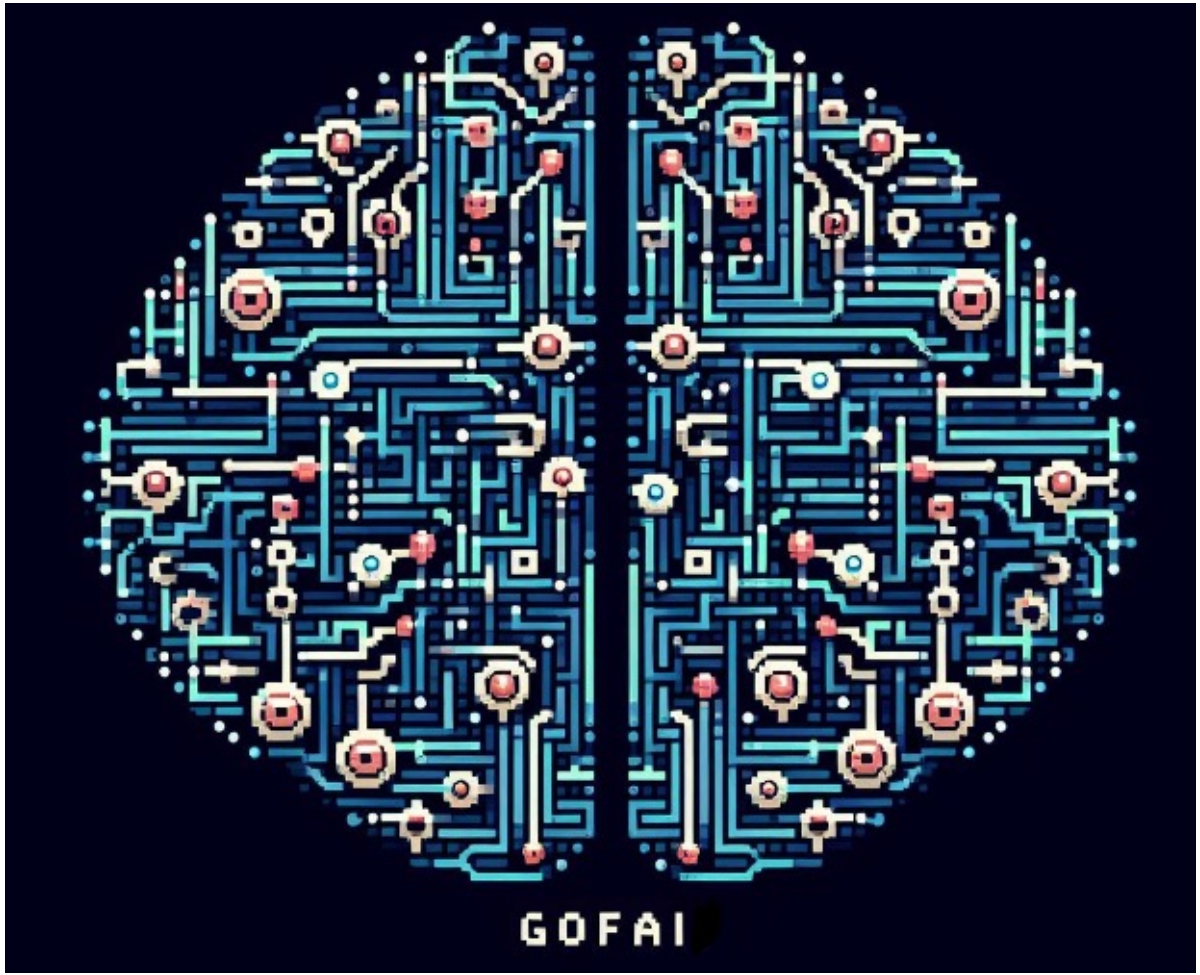
This approach is built on the central idea that **cognition is fundamentally about rule-governed symbol manipulation**, the very same process that **computers perform**.

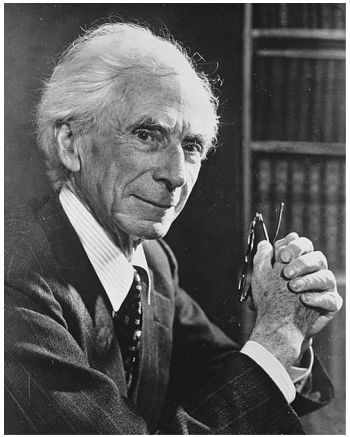



LOGIC





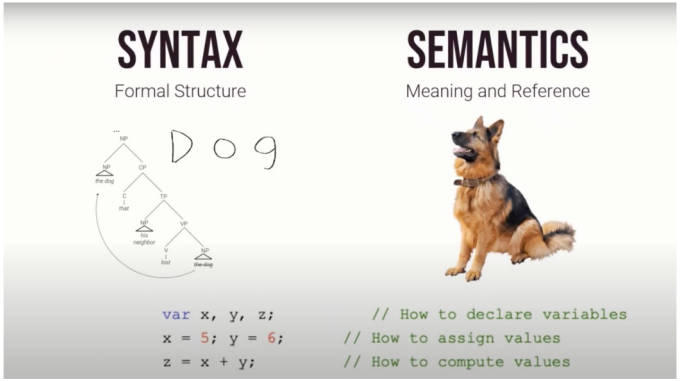

GOFAI

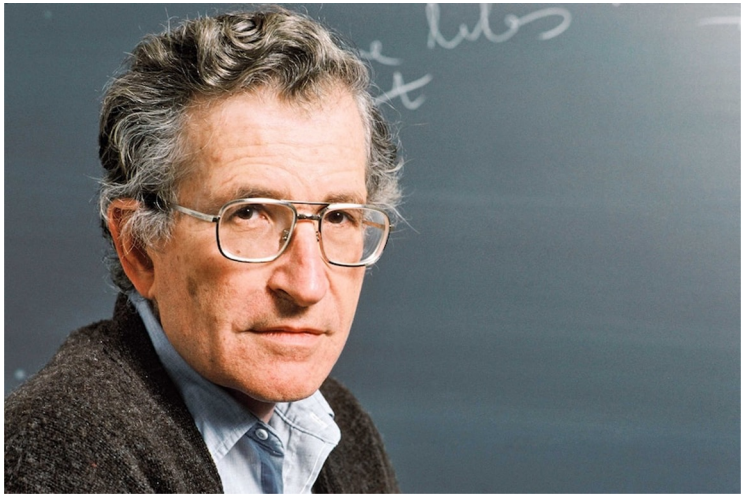






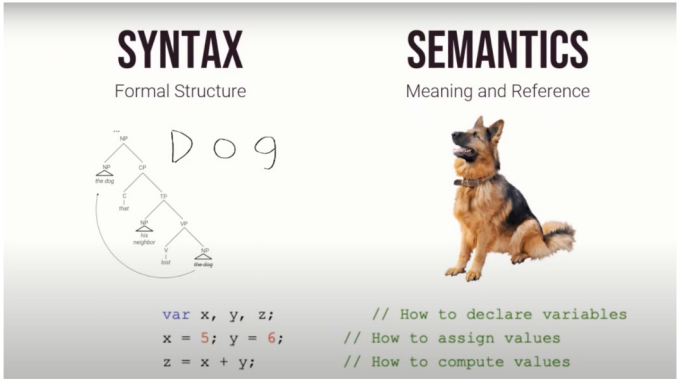
Formal logic changed how we approached symbols, moving away from their meaning (**semantics**) and focusing more on their form and structure (**syntax**). In this context, logic became about **manipulating abstract symbols according to rules**, regardless of what the symbols actually represent.

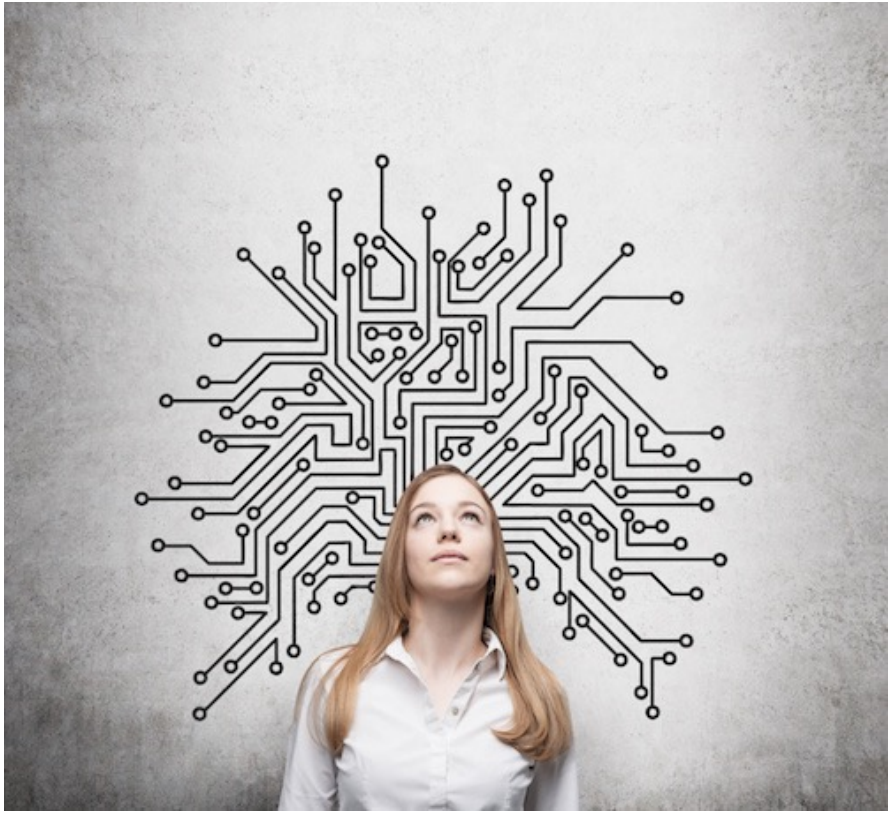






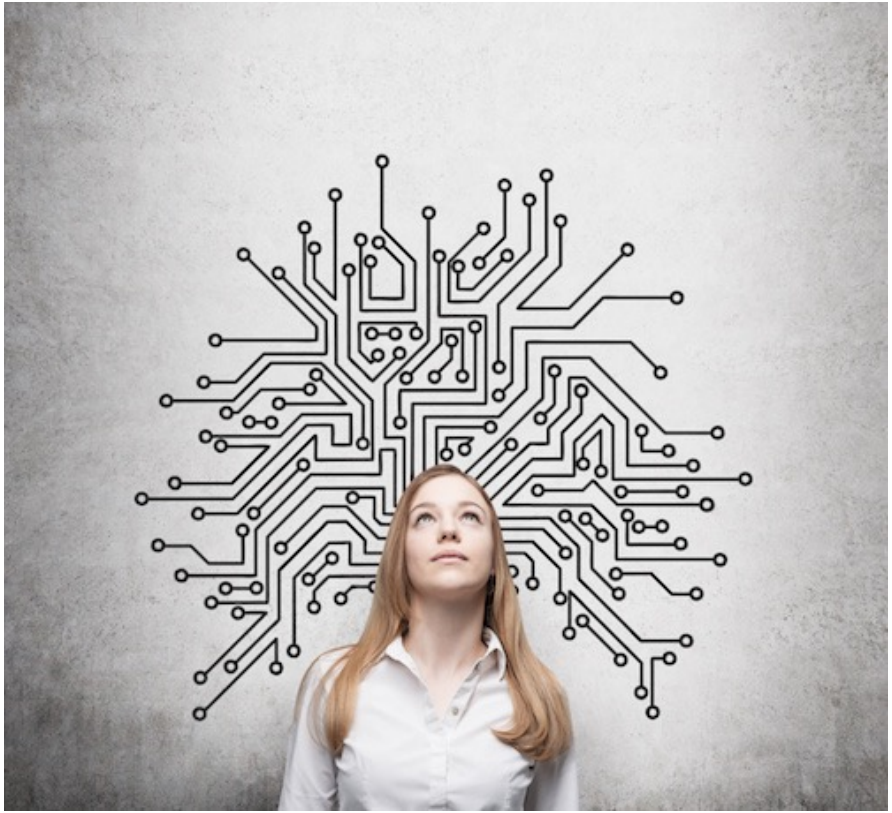
"A sentence was a string of symbols. A language was a set of such strings. A grammar was a set of recursive procedures for generating such sets. Language was syntacticized—placed mathematically within a Post system, with abstract symbols manipulated in algorithmic fashion by precise formal rules".





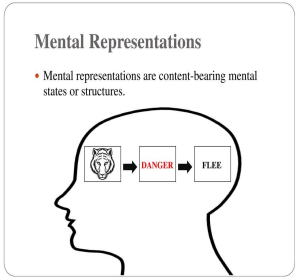
Functionalism is the idea that mental states are not defined by what they are made of, **but by what they do**

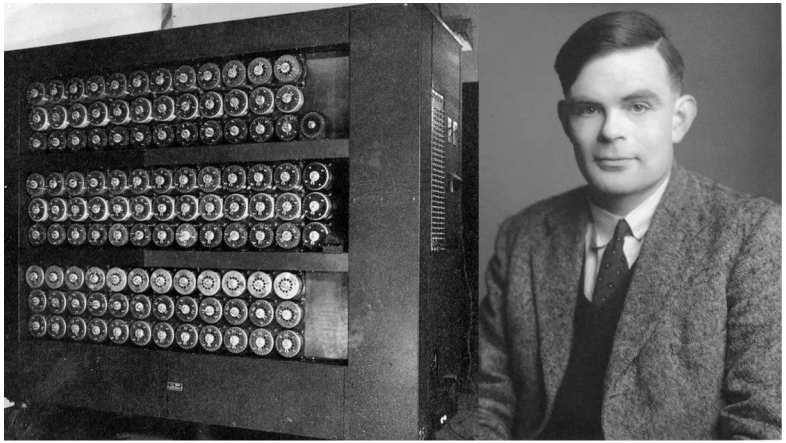


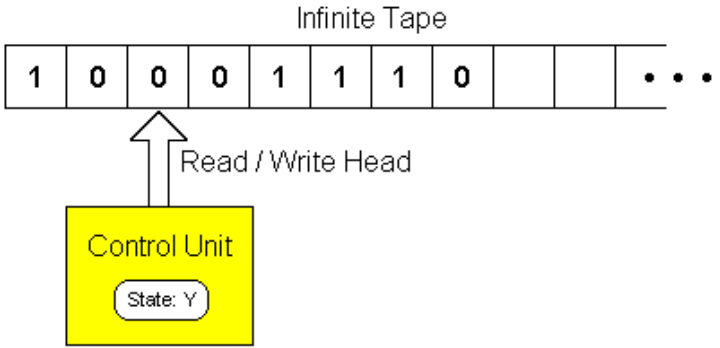


In this view, mental states are also **representational**. They represent the world and allow us to think about things that aren't immediately present

This ability to manipulate internal representations according to rules is precisely what we expect from both humans and computers. In this sense, the **Representational Theory of Mind** aligns with classical AI–human cognition and machine processes both operate by manipulating symbols according to rules.

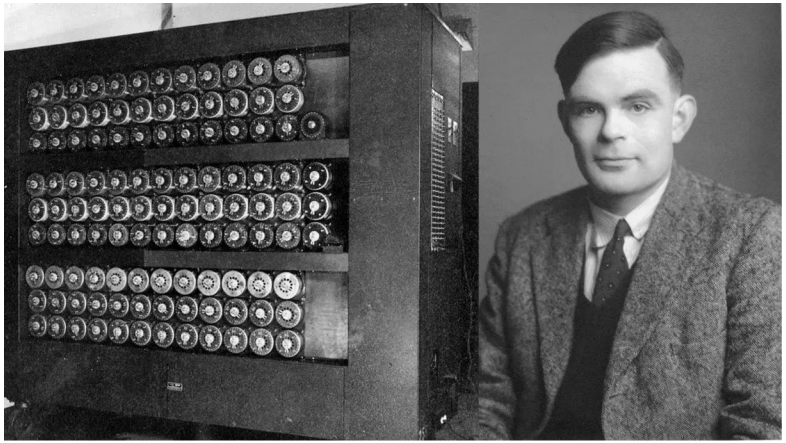


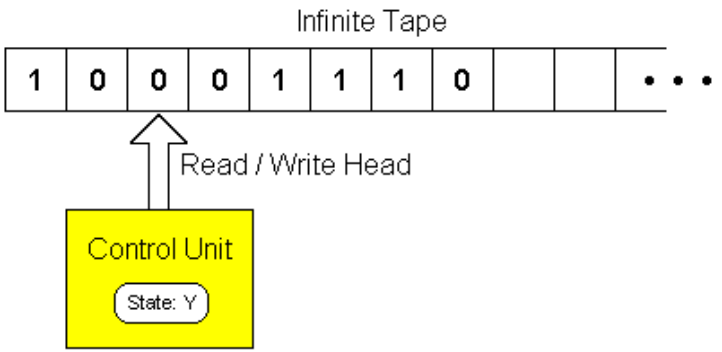




Alan Turing demonstrated in 1936 that a machine performing a sequence of simple computations could solve any well-defined problem.

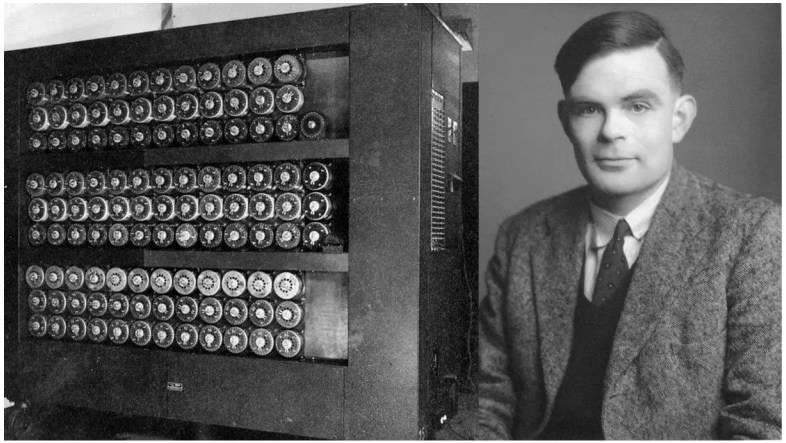
The machine reads symbols on the tape one at a time, follows predefined rules to either write a new symbol, move the tape left or right, or change its internal state. By following these simple rules, a Turing machine can simulate any algorithmic process, demonstrating how computers perform calculations.





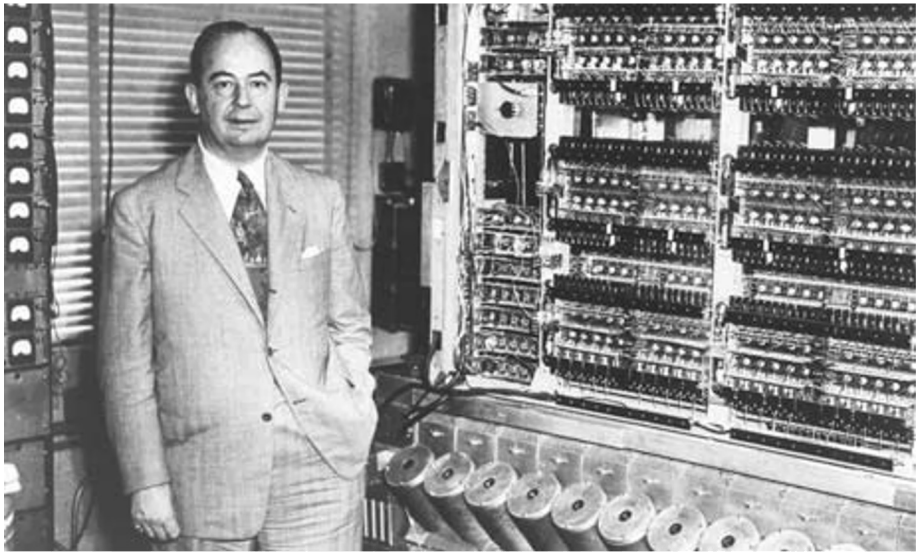
Alan Turing demonstrated in 1936 that a machine performing a sequence of simple computations could solve any well-defined problem.

Turing's thesis, which claims that if cognitive processes are **algorithmically calculable**, then they could be carried out by a Turing Machine



Alan Turing demonstrated in 1936 that a machine performing a sequence of simple computations could solve any well-defined problem.

The digital computer, designed by John von Neumann in the 1940s, became the model for this line of thinking. Just as software runs on hardware, cognition was thought to operate on the brain's physical substrate.




ANY
QUESTIONS?



GOFAI has had its share of **success stories**, most notably **Deep Blue**, the chess-playing AI that defeated world champion Garry Kasparov in 1997.

More recently, AI models like **ChatGPT** demonstrate how symbol manipulation on a massive scale can simulate human-like language





“In this view, "mindware" is like software—dependent on the physical brain but abstract enough to be understood in terms of computation rather than biology. The key idea is that cognition is a computational process, one that can be studied and understood through the lens of formal systems and algorithms, rather than the specific biological details of the brain”.



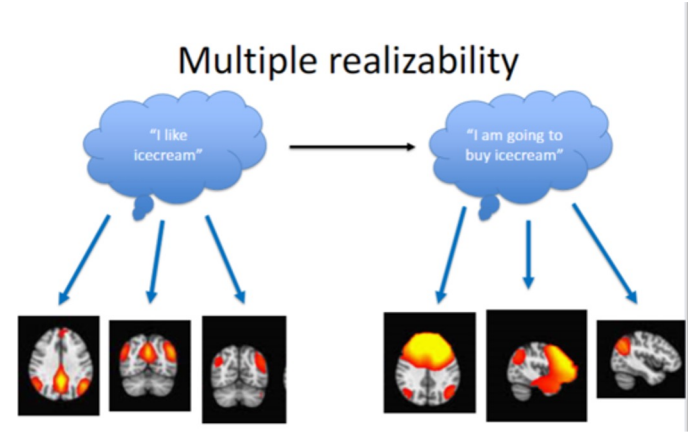


Churchland and Sejnowski (1992) suggest that something is considered a computer when its inputs and outputs can be understood as representing parts of a function that we care about. For example, a pocket calculator computes because its buttons and display can be interpreted as numbers, and its internal processes match meaningful changes in arithmetic



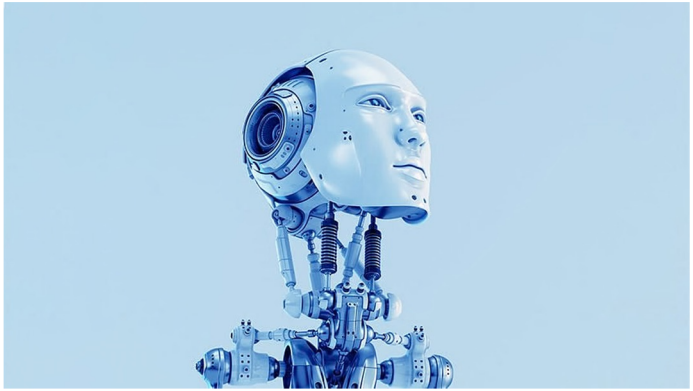


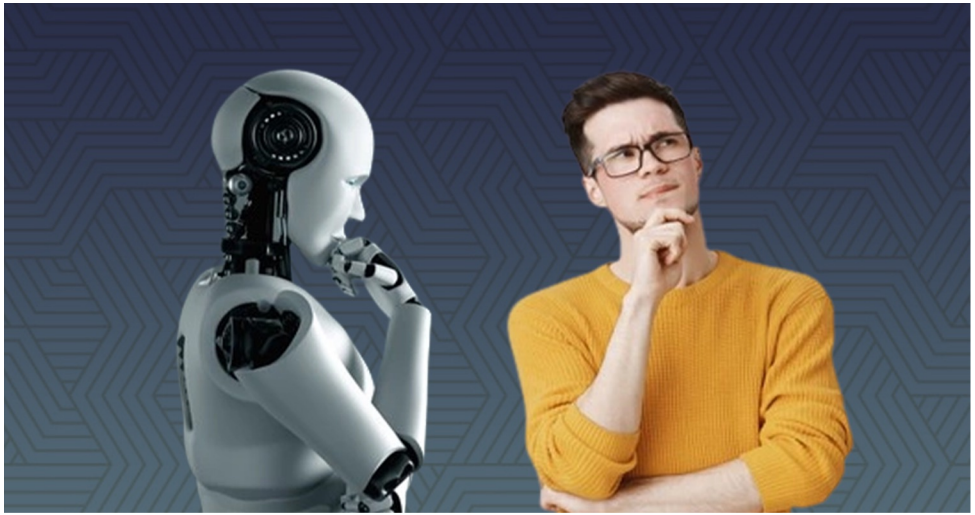
Dave Chalmers adds that a device performs computation if its physical state changes follow the steps of a formal set of rules. This idea can even apply to natural systems. For instance, if an animal's behavior changes based on how many predators it detects, nature has, in a way, created a basic computer..

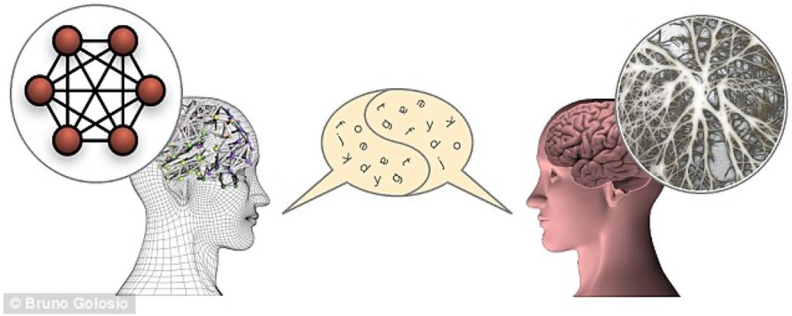

Multiple realizability
"I like icecream"
"I am going to buy icecream"


© BNPS.CO.UK





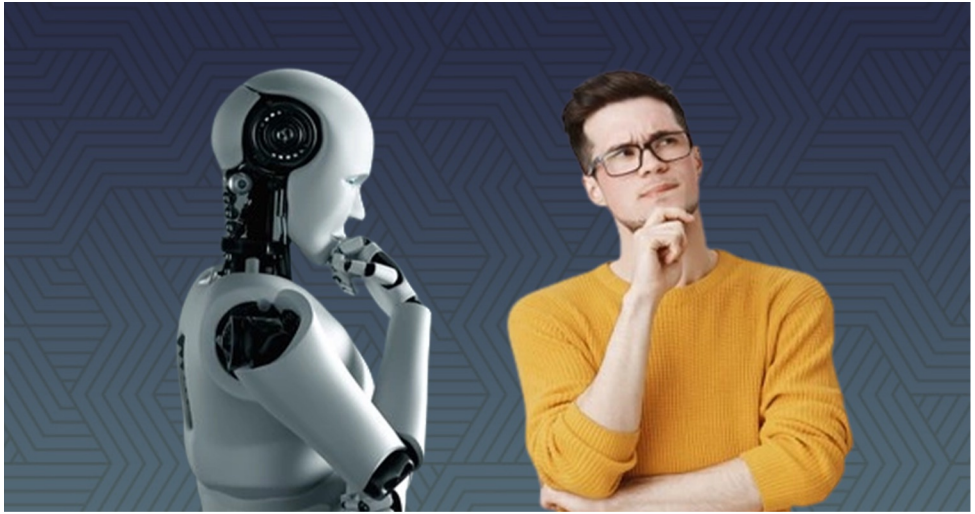


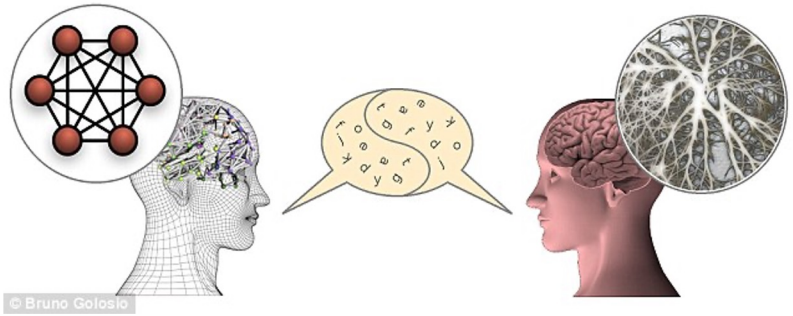


Are these systems **truly intelligent**?

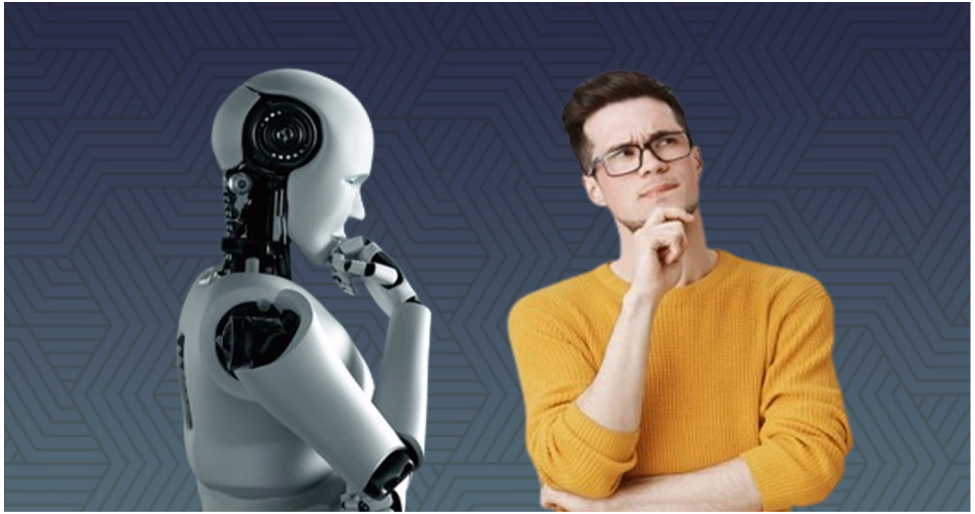
Is this intelligence fundamentally **different from ours**?

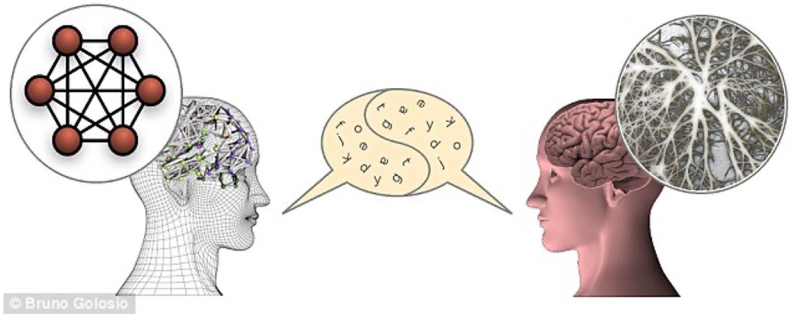
How do we tell?

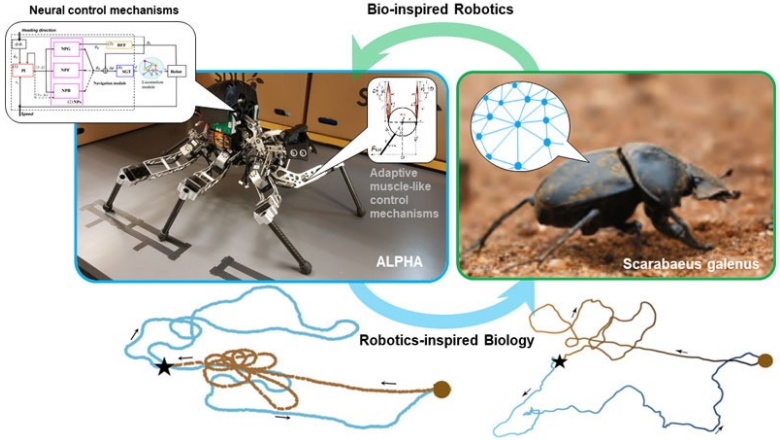



© Bruno Golosio






© Bruno Golosio


Neural control mechanisms
Bio-inspired Robotics
Adaptive muscle-like control mechanisms
ALPHA
Scarabaeus galenus
Robotics-inspired Biology




ANY
QUESTIONS?



Now ask yourself; does John **understand Chinese**?



“If thoughts are composed out of symbols, then from **where to those symbols get their meanings**?...

So the problem is this: once we’ve stripped out meanings from the foundations of our mechanical account of cognition, ***how do they get back in again***?”

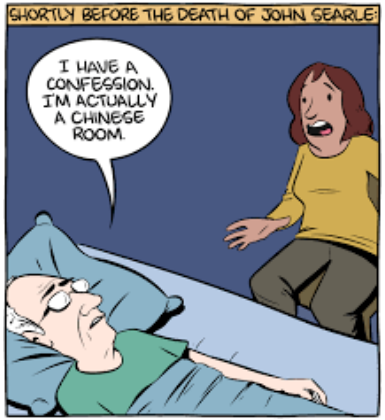


First, the thought experiment does not show that machines cannot think – Searle clarifies his view: “On the contrary, the brain is a machine and the brain can think” and in his 1980 paper, in response to the question ‘Could a machine think?’ he writes “The answer is, obviously, yes. We are precisely such machines.”

Second, the thought experiment does not show that *computers* can’t think. As we saw from Turing, since we human beings can carry out computations, we are computers, and we can think

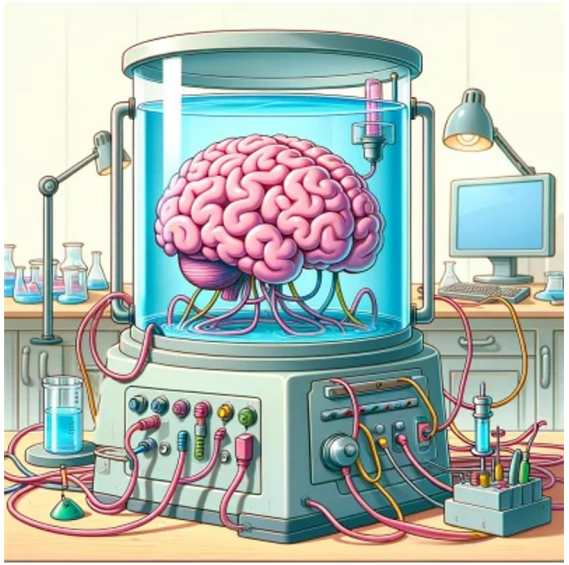
Finally, the Chinese Room does not show that only brains can think; it is intended to show that brains must be doing more than the “mere shuffling of formal symbols” and that any artificial thinking machine would have to duplicate the specific neurophysiological processes responsible for thought in the human case

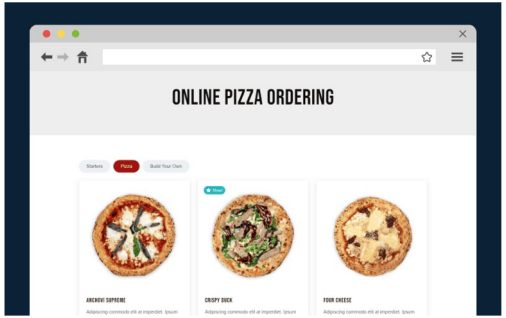

INPUT
牛对
DATA
RULE LEDGER
OUTPUT


SHORTLY BEFORE THE DEATH OF JOHN SEARLE:
I HAVE A CONFESSION. I'M ACTUALLY A CHINESE ROOM.

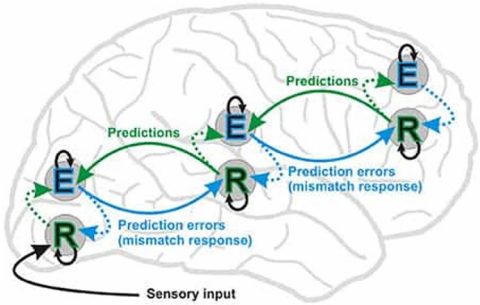

That whole system understands Chinese.

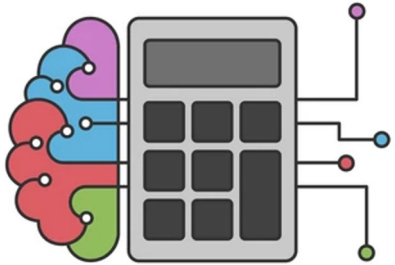

Consciousness

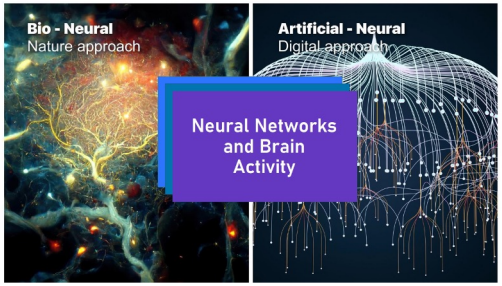



ONLINE PIZZA ORDERING

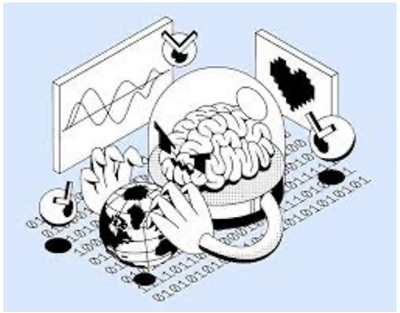


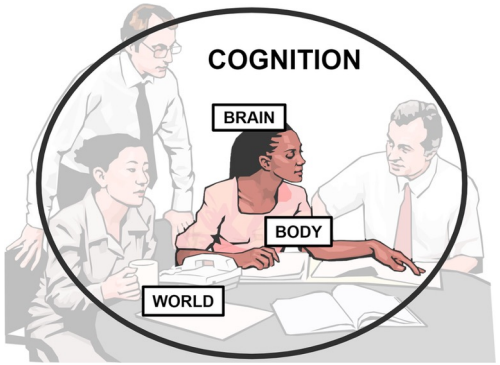


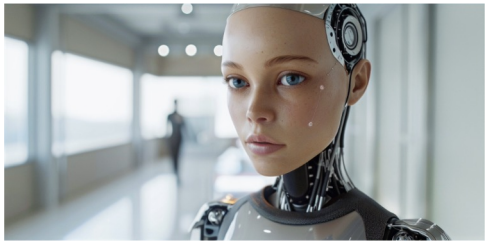




Flow of the Course:







Thanks

