

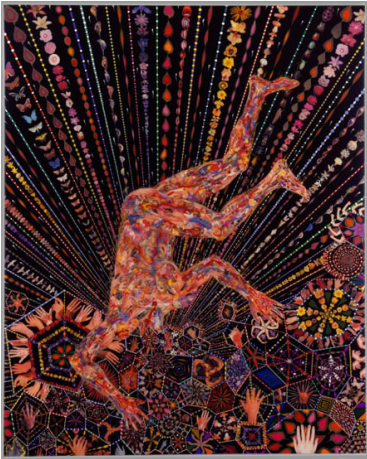
# Mind

Christian Wüthrich

<http://philosophy.ucsd.edu/faculty/wuthrich/>

**14 The Nature of Reality**

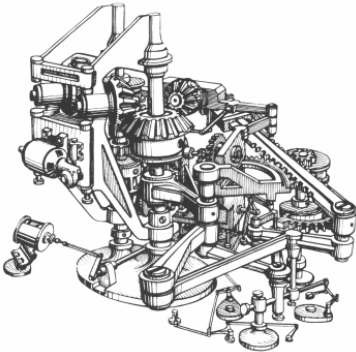
# The 'Organic' World



The 'organic' picture of the world:

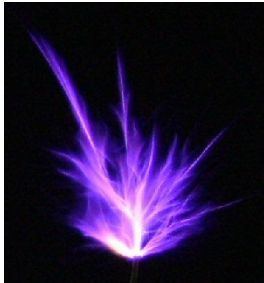
- Aristotle
- teleology
- vitalism

# The Mechanical Picture of the World



- Bacon's *Novum Organum*, Descartes' mechanical philosophy, Boyle, Leibniz, Newton
- Causes, laws of nature
- Non-teleological

# The Mechanical World



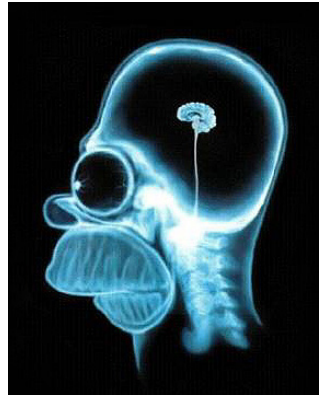
↑  
vital spark...

- Where does Life fit in?
  - vitalism: holds that there exists in all living things an intrinsic factor - elusive, inestimable, and unmeasurable - that activates life.
  - By 1940s or so, vitalism is dead
  - Viruses: like us, they have features of heredity, evolution, reproduction, but unlike us, they have no metabolism, cell structure, or homeostasis. Alive or dead?
- Where does Mind fit in?

# The Trouble with Minds 1

Minds seem to have some properties that are not easily seen as the result of atoms (or fields) in motion (Searle 1984):

- 1 **Consciousness:** We're aware of the things around us and experience all sorts of sensations. What exactly is consciousness? That is a tough question. Mmm... Consciousness does have a sort of 'know-it-when-you-experience-it' quality to it. Anyway, we're undeniably conscious and electrons plausibly are not. How can a swarm of unconscious particles become conscious?



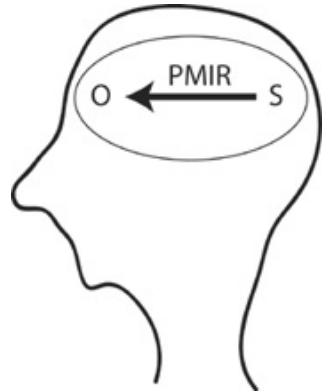
## The Trouble with Minds 2

- 2 Subjectivity:** Our mental lives are private. You may look like you're paying attention during lecture, thinking about philosophy, but for all I know you may be thinking about the Lakers. Arguably, no amount of digging around in your skull will ever let me know what you're thinking.



## The Trouble with Minds 3

- ③ **Intentionality:** Our thoughts represent things in the world. They are **about** the world. My thought 'The cat is on the mat' is about a cat, mat and a relation between the two. Yet how can thoughts (or anything) be about anything? What is 'aboutness' anyway? This category is more general than that of conscious thought. How can matter in motion represent things?



**Figure:** Metzinger's 'phenomenal model of the intentionality relation' (PMIR). This figure has been adapted from (Metzinger 2003).

## The Trouble with Minds 4

- ④ **Incorrigibility:** Arguably, there are some mental states that you can't be wrong about. How could you be wrong about whether or not you're in pain, for instance? Feeling the sensation of pain just is being in pain. Yet if mental states are merely a product of matter in motion, why should our knowledge of them be any different from our knowledge of other things?





# Mind-Body Problem

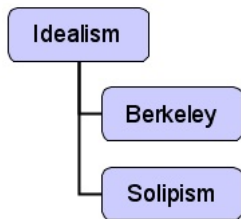
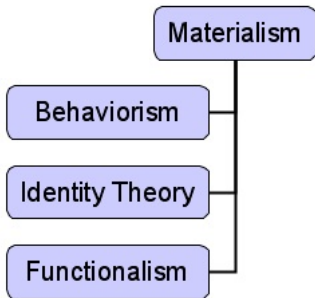


The mind-body problem is simply the question of how the mind and body are connected to each other. We believe the two are connected somehow: whenever we drink too much or injure our bodies physical events are affecting our mental lives; typically when we think 'raise arm' our arm rises, and in this case, mental events are affecting physical events. But how exactly are they connected? Are they one and the same thing, and if so, in what sense? Or are mental events radically different from physical events, and if so, do the two really interact?

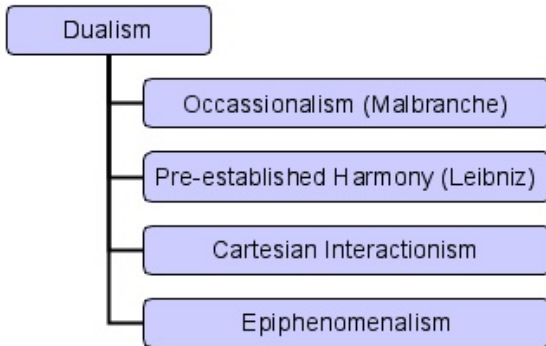
## Mind-Body Problem 2

Philosophers have offered a multitude of different answers to these questions. **Monism** is the view that there is only one kind of elemental substance, usually judged to be matter (materialism or physicalism). According to **dualism**, there are two types of elemental substance. Besides this division of theories there are many more distinctions that mark off different theories.

# Monism



# Dualism



# Descartes and Dualism

Method: find proposition  $p$  that is absolutely certain....

- 1 If I have a clear and distinct idea of myself without my body, then I am not my body.
- 2 I do.
- 3 Therefore, I am not my body.



# Defense

- **Defense of (1):** What I can perceive clearly and distinctly can be made so by God. (God wouldn't trick me when it comes to clear and distinct ideas.)
- **Defense of (2):** I can think of myself (clearly and distinctly) as only a thinking thing.

# Existence of Material Things

- 1 Sensory ideas are caused independently of my thought.
- 2 I am not the cause of these sensory ideas.
- 3 If so, then the cause is either (a) physical objects, (b) God, or (c) created non-physical thing.
- 4 Not (b) or (c).
- 5 Therefore, (a), physical objects exist.

## Defense of (4):

- If God, then He is a deceiver.
- If nonphysical created thing, then God is the author of a deceiver.

## Another Type of Argument

(see Churchland, *Matter and Consciousness*)

- 1 My mental states are introspectively knowable to me as states of my conscious self.
  - 2 My brain states are not.
  - 3  $X$  and  $Y$  are identical iff they have all and only the same properties.
- 
- 4 Therefore, my mental states are not identical with my brain states.



But...



Superman = Clark Kent

# Knowledge Argument for Dualism

Originally in Frank Jackson, 'Epiphenomenal qualia', *Philosophical Quarterly* 32 (1982): 127-136.



# Knowledge Argument: The Set-up

Robert Stalnaker, *Our Knowledge of the Internal World*, Oxford UP (2008).

“[Mary] is a brilliant scientist who has been confined since birth to a black and white room. She knows, from reading the black and white books that line the shelves of her room, all there is to know about the physics of color, and the neurophysiology of color vision, but she has never had the opportunity to see colors. Even though she knows all the relevant physical and biological science, there is still something she does not know, something that she will learn only when she first emerges from her room, and sees colored things: she doesn’t know what it is like to see colors.” (24)

# Knowledge Argument: The Argument



- 1 “The story seems to imply that a person might know all the relevant physical facts while remaining ignorant of certain further facts—facts about the qualitative character of our visual experience.
- 2 “So there must be facts to be known that are not physical facts.
- 3 “But if there are facts that go beyond the physical facts, then materialism—the thesis that all the facts are physical facts—is false.” (24f)

## Arguments Against Dualism (Churchland 1984)

- Simplicity
- Explanatory Impotence: “Compare now what the neuroscientist can tell us about the brain, and what he can do with that knowledge, with what the dualist can tell us about spiritual substance, and what he can do with those assumptions?”
- Interaction?
- Neural dependence of all known phenomena
- Evolutionary history

# Philosophical Behaviorism



What is solubility, really? One answer is that there is not really any sort of thing that is solubility; rather claims about solubility are claims about the dispositions to engage in various types of observable behavior.

## Inputs

Placing aspirin in water

Placing aspirin in soda

⋮

## Outputs

dissolve

dissolve

⋮

What is the mental state pain, really? One answer is that there is not really any sort of thing that is a mental state; rather claims about pain, etc., are claims about the dispositions to engage in various types of observable behavior.

### Inputs

Testing blood pressure

Ask: "Do you want to play chess?"

Ask: "Does THIS hurt?"

Say: "Really, there is no such thing as pain"

### Outputs

higher than normal

"No!"

"Arghh!"

Punch to the nose of speaker

### Thesis (Philosophical behaviorism)

*Philosophical behaviorism is the thesis that all mental states are to analyzed as dispositions to engage in certain types of observable behavior. Psychological terms can just be translated into non-psychological behavioral terms.*





# Problems

- Perfect actor example?
  - Observable?
  - Finite list of events?
  - Qualia?
  - Main one: need mental states to explain behavior
- ⇒ marks of a degenerate research programme



# Identity theory

<http://plato.stanford.edu/entries/mind-identity/>

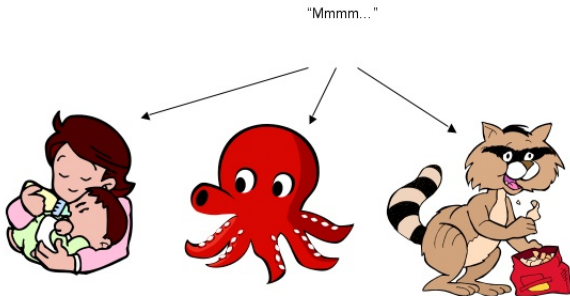
My mental states just are my brain states. This implies that there is a 1:1 mapping between my brain states and my mental states.

Water is H<sub>2</sub>O  
Temperature is mean kinetic energy



# Types and Tokens

- **Types** versus **tokens**: distinction between a general sort of thing and its particular concrete instances (“Rose is a rose is a rose is a rose.” Gertrude Stein)
- **Type-Type Identity Theory**: Every mental type is equivalent to some physical type, e.g., sharp pain = some *c*-fiber firing at a certain frequency
- **Objection: Multiple Realization**



# Functionalism

<http://plato.stanford.edu/entries/functionalism/>

**'Software'**

Duke Nukem



**'Hardware'**

IBM, Mac

**'Mental state'**

Pain



**'Physical state'**

Man, woman, octopus

## Coke Machine (Coke costs 10¢)

Current Internal State	Coin Input	Coke Output	Coin Output	New Internal State
S1	0.05	No	No	S2
S1	0.10	Yes	No	S1
S2	0.05	Yes	No	S1
S2	0.10	Yes	Yes	S1

# Functionalism



“My argument is this: The definitive characteristic of any (sort of) experience as such is its causal role, its syndrome of most typical causes and effects. But we materialists believe that these causal roles which belong by analytic necessity to experiences belong in fact to certain physical states. Since these physical states possess the definitive character of experiences, they must be experiences.”

(David Lewis, *Journal of Philosophy* **63** (1966): 17-25)

## Characterization (Functionalism)

*Mental states and processes are defined **in terms of their causal roles** (their functions). And it is then noted that neurons instantiate these roles.*

# Against functionalism: Inverted spectrum

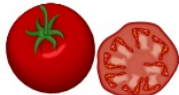
INPUTS



RED



OUTPUTS



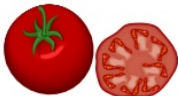
“Look at that red tomato!”

“That tomato is the same color as a stop sign”

INPUTS



BLUE



“Look at that red tomato!”

“That tomato is the same color as a stop sign”



## Responses by functionalists

- Not possible: Inverted goggle experiment
- Functionalism incomplete (Searle)

# Can computers think?



# Alan Turing (1912-1954)



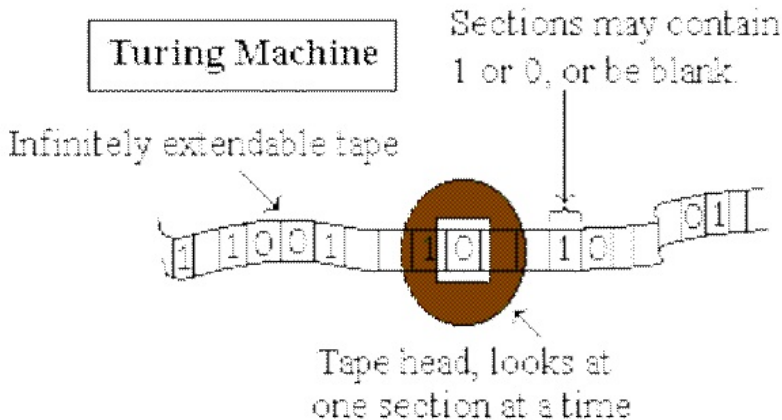
- 'father' of modern computer science
- Turing machine, Turing test
- contributions to artificial intelligence
- during WWII: worked at Bletchley Park (UK's codebreaking centre), where he headed a department that helped breaking the Enigma machine

# Turing Machines

From <http://plato.stanford.edu/entries/turing-machine/>:

“A Turing machine is an abstract representation of a computing device. It consists of a read/write head that scans a (possibly infinite) one-dimensional (bi-directional) tape divided into squares, each of which is inscribed with a 0 or 1. Computation begins with the machine, in a given “state”, scanning a square. It erases what it finds there, prints a 0 or 1, moves to an adjacent square, and goes into a new state. This behavior is completely determined by three parameters: (1) the state the machine is in, (2) the number on the square it is scanning, and (3) a table of instructions. The table of instructions specifies, for each state and binary input, what the machine should write, which direction it should move in, and which state it should go into. (E.g., “If in State 1 scanning a 0: print 1, move left, and go into State 3”.)

“The table can list only finitely many states, each of which becomes implicitly defined by the role it plays in the table of instructions. These states are often referred to as the “functional states” of the machine. A Turing machine, therefore, is more like a computer program (software) than a computer (hardware)... Computer scientists and logicians have shown that Turing machines—given enough time and tape—can compute any function that any conventional digital computers can compute.”



# What is a computer?

1 0 1 1 0 1 0 0 1 0



0 1 0 0 1 0 1 1 0 1

Current State	Current Symbol	New Symbol	Move	New State
A	space	space	L	A
A	0	1	L	B
A	1	0	L	B
B	0	1	L	B
B	1	0	L	B
B	space	halt	L	B

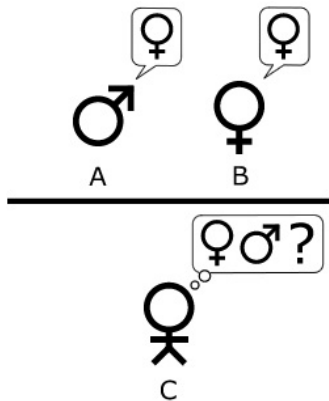
# Turing machines

- The Turing machine can compute any function that a regular computer can compute
- **Church-Turing thesis**: anything computable is Turing computable
- **Universal Turing Machine**: a Turing machine capable of simulating any other Turing Machine



# Imitation game

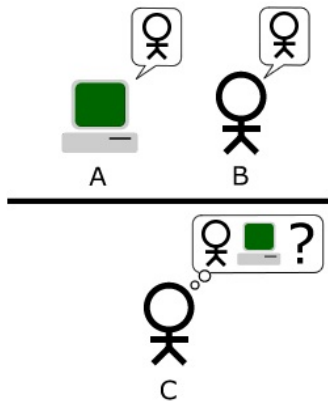
Alan Turing, 'Computing machinery and intelligence', *Mind* **59** (1950): 433-460.



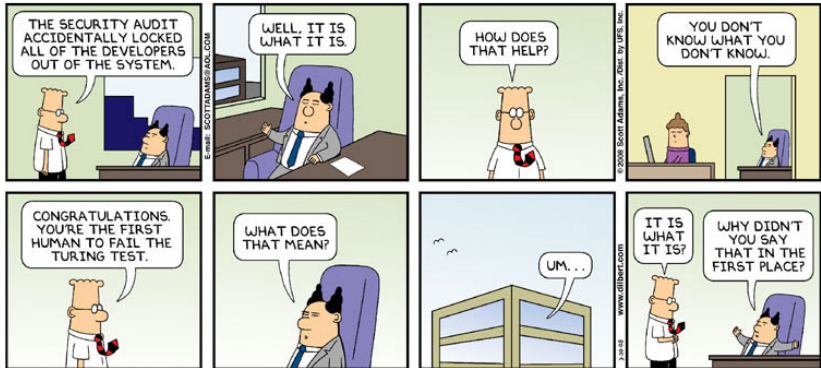
Player C, through a series of written questions, attempts to determine which of the other two players is a man, and which of the two is the woman. Player A, the man, tries to trick player C into making the wrong decision, while player B tries to help player C.

# Turing test

Alan Turing, 'Computing machinery and intelligence', *Mind* **59** (1950): 433-460.



Described by Alan Turing in the same paper, the Turing test proceeds as follows: a human judge has to find out by means of typed conversation with one human and one machine, each of which try to appear human, which one is the human. If the judge cannot reliably tell the machine from the human, the machine is said to have passed the test.



## Computer can't (??)

- be kind
- be resourceful
- be beautiful
- be friendly
- have initiative
- have a sense of humor
- tell right from wrong
- make mistakes
- fall in love
- enjoy strawberries and cream
- make someone fall in love with one
- learn from experience
- use words properly
- be the subject of one's own thoughts
- have as much diversity of behavior as a human
- do something really new

# Artificial intelligence

- Is passing the Turing Test necessary for intelligence?
- Is passing the Turing Test sufficient for intelligence?

# A couple of definitions

## Definition (Syntax)

*"The **syntax** of a language is its grammar, or the way its expressions may be put together to form sentences. A syntactic study is one that is not concerned with sentence-meaning, but with the purely formal aspects of word combination in a language."* (Simon Blackburn, Oxford Dictionary of Philosophy)

## Definition (Semantics)

*Semantics is "the study of meaning of words, and the relation of signs to the objects to which the signs are applicable. In formal studies, a semantics is provided for a formal language when an interpretation or model is specified."* (ibid.)

## A couple of theses

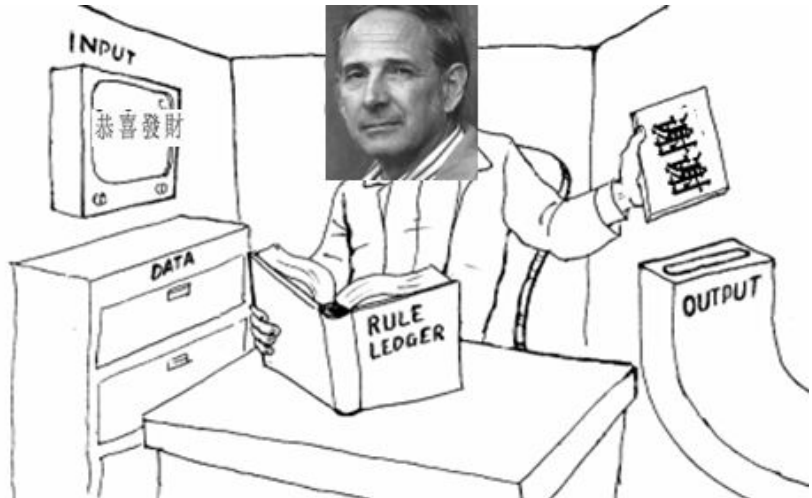
### Thesis (Strong AI)

*“Strong AI is the philosophical thesis that appropriately programmed computers have minds in exactly the same sense that we do.” (ibid.)*

### Thesis (Weak AI)

*“Weak AI is the methodological belief that the best way to explore the mind is to proceed as if [the thesis of strong AI] were true, without commenting on the legacies of dualism that lead to discomfort with the strong thesis.” (ibid.)*

# Searle's Chinese Room





“Searle requests that his reader imagine that, many years from now, people have constructed a computer that behaves as if it understands Chinese. It takes Chinese characters as input and, using a computer program, produces other Chinese characters, which it presents as output. Suppose, says Searle, that this computer performs its task so convincingly that it comfortably passes the Turing test: it convinces a human Chinese speaker that the program is itself a human Chinese speaker. All of the questions that the human asks it receive appropriate responses, such that the Chinese speaker is convinced that he or she is talking to another Chinese-speaking human being. Most proponents of artificial intelligence would draw the conclusion that the computer understands Chinese, just as the Chinese-speaking human does.

“Searle then asks the reader to suppose that he is in a room in which he receives Chinese characters, consults a book containing an English version of the aforementioned computer program and processes the Chinese characters according to its instructions. He does not understand a word of Chinese; he simply manipulates what, to him, are meaningless symbols, using the book and whatever other equipment, like paper, pencils, erasers and filing cabinets, is available to him. After manipulating the symbols, he responds to a given Chinese question in the same language. As the computer passed the Turing test this way, it is fair, says Searle, to deduce that he has done so, too, simply by running the program manually. “Nobody just looking at my answers can tell that I don’t speak a word of Chinese,” he writes.” (from Wikipedia)

## Searle's conclusion

- ⇒ computer no more understand Chinese than he does
- ⇒ computers don't have 'understanding' of what they say
- ⇒ computers don't have conscious mental states
- ⇒ passing the Turing test is no proof for the having of conscious mental states, i.e. for the having of a mind

# Logical Structure of Chinese Room Argument

P1: programs are formal (syntactical)

P2: minds have contents (semantics)

P3: syntax is not sufficient for semantics

---

C: programs are not minds

# Logical Structure of Chinese Room Argument

P1': processes are formal (syntactical)

P2: minds have contents (semantics)

P3: syntax is not sufficient for semantics

---

C': processes are not minds

# Systems Reply

- Compare with Turing machine
- Searle: it's just ridiculous to say "that while [the] person doesn't understand Chinese, somehow the conjunction of that person and bits of paper might" (1980a, 420).

# Robot Reply

- what prevents the person in the Chinese room from attaching meanings to (and thus prevents them from understanding) the Chinese symbols is the sensory-motoric disconnection of the symbols from the realities they are supposed to represent—a causal connection is required
- Searle: rerun story but with room in robot's head
- Searle: reply tacitly concedes the point, namely, that one needs a set of causal relations to the outside world

## Brain Simulator Reply

- Imagine that the program implemented by the person in the room “doesn't represent information that we have about the world... but simulates the actual sequence of neuron firings at the synapses of a Chinese speaker when he understands stories in Chinese and gives answers to them.”
- No difference between program of native Chinese speaker's brain and program of person in room.
- Searle: water pipes



## Combination Reply

- Systems, Robot and Brain simulator together
- Searle:  $3 \times 0 = 0$

## The Other Minds Reply

- “if the computer can pass the behavioral tests as well” as a person, then “if you are going to attribute cognition to other people you must in principle also attribute it to computers” (1980a, 421).
- Searle: it’s “not... how I know that other people have cognitive states, but rather what it is that I am attributing when I attribute cognitive states to them. The thrust of the argument is that it couldn’t be just computational processes and their output because the computational processes and their output can exist without the cognitive state” (1980a, 420-421)

- Syntax does not determine semantics. However, we are thinking things, and apparently our brain operates by manipulating representations. What could possibly give those representations the meaning they have?
- Hard to reconcile with Searle's 'biological naturalism'
- Is it the identity theory or dualism? Searle says neither...

# Functionalism again

- Bizarrely implemented programs (or perhaps Searle's mind is made of waterwheels, tin cups, etc.): is there a **right stuff**?
- Functionalism: programs might be implemented in bizarre stuff
- Searle: right programming does not suffice; rather, right stuff is needed