Can machines think?

Cognitive Science (the science of how the mind works) assumes that the mind is computation. At least we assume that, whatever it is that the brain does, is computable.

The idea, that thinking is some sort of computation goes back a long time...

Are we machines? If so, then machines can think too.

Given that we are physical systems (machines of sort), and given that our brains compute, and given that we are intelligent, then we will have to accept that other machines can also think, at least in principle.

We compute since 1651.

“For reason, in this sense, is nothing but reckoning (that is, adding and subtracting) of the consequences of general names agreed upon for the marking and signifying of our thoughts.”

Thomas Hobbes,
LEVIATHAN,
1651

Machines, who think.

Comp 2920
Professional Issues
& Ethics
in Computer Science
S2 · 2004
How do we get them to think?

GOFAI: Take a lot of facts and put them into a large database, then take a lot of rules about how these facts relate to each other. Add a few sort algorithms and heuristics, and there you have it.

- ELIZA (Joseph Weizenbaum, 1966)
- SHRDLU (Terry Winograd, 1968-70)
- CYC (Doug Lenat, 1990s)

Connectionism: Take many (very many) simple artificial neurons and connect them to each other. Then all that is left is to train the network.

- Neurons are logic devices (McCulloch & Pitts, 1943)
- Perceptrons - two layer neural nets (Rosenblatt, 1957)
- all is well, until ...
- 1969 Perceptrons is published (Minsky & Papert)
- 1980 an new start (McClelland, Rumelhart, et al.)
- NetTalk
- ALVINN (steers a car), NetTalk (learns to speak)
- Models of cognitive functions (Elman et al.)

Hang on a bit ... this won't do.

The 'facts and rules' approach will never work, because human life cannot be expressed in these terms. Humans know the world only how they perceive it, and not how the world actually is. Humans are embedded in their environment and it is impossible to express everything in terms of algorithms.

Hang on a bit ... this won't do.

Computers and their programs are formal systems. Thanks to Kurt Gödel and the Incompleteness Theorem, we know that in any formal system some things remain unprovable. But, we human can figure these unprovable things in formal systems out - therefore human are better and humans are not machine, and Mechanism is false.

Hubert Dreyfus,

John R. Lucas,
*Minds, Machines and Gödel*, 1961
John Searle's thought experiment claims that strong AI (conscious machines) is impossible. Syntax, so the claim, is not sufficient for intentionality (semantics). His (in)-famous argument goes like this...


---

**The Chinese Room Argument**

Does he understand Chinese?

Not a word, says John Searle

---

**Are the arguments against AI convincing?**

J. R. Lucas's (Gödelian argument) holds only for formal systems. Aaron Sloman, Jack Copeland, and others, have argued that many types of computation are not Turing equivalent. The Gödelian argument does not apply in such cases.

The jury is still out on J. Searle's 'Chinese Room'. Many people agree that the argument is 'not quite right'... but no-one has refuted the argument convincingly.

---

**How do we know if it thinks?**

That is easy! We let people talk to a machine or another human using a terminal and ask them to tell us who is which, or which is who. If they can't tell the machine from the human, then the machine is surely intelligent.

Alan M. Turing, *Computing Machinery and Intelligence*, 1950
The TT as a measure of intelligence.

The TT measures only performance not ability. The machine is not intelligent, it only looks that way. A chess computer does not know anything about chess, it just executes algorithms and does not make 'mistakes'.

So how can we recognise intelligence in a machine? How do we recognise intelligence (or consciousness) in us?

Zombies

- **“real”**
  - Completely deprived of conscious experience or subjective consciousness (qualia)

- **philosophical**

**Conclusion:**

*It is all a waste of time ...*

The premise that human thoughts are computations is wrong. The very idea that we are machines is wrong, for various philosophical, religious, political or other reasons.

There is no real success story. AI as a discipline, despite all the modern machinery, has not come up with an intelligent machine.

... oh no, it isn't a waste of time.

Some progress has been made, but machines might have to be much more complex than the machines we have today.

Others argue that we have to change our concept of computation.

It all depends on some fundamental philosophical questions ...
Thank you.