

CHAPTER 1

INTRODUCTION

*One of these days,
not very far away in the future
a machine may, after a thorough reflection,
reach the conclusion: "I think; therefore I am... immaterial".*

Hardly ever has a technological pursuit had such philosophical importance as that of the quest for conscious machines. Our own consciousness and even our self-existence are still poorly understood. What should we think then about machine consciousness if we were ever able to create it?

Will machine consciousness be possible at all, will it be the greatest invention of mankind or will it only remain a science fiction dream? Is our consciousness unique, excluding by its special nature all reproduction by technical means known to man today? Or, could artificial consciousness nevertheless be possible? What would it take to design a conscious machine and how should we approach this task? These are the questions that are addressed in this book.

A prominent hallmark of consciousness is the awareness of one's thoughts. Thus a prerequisite for consciousness would seem to be the ability to think, to have mental content to be aware of. A system cannot be aware of its thoughts if it does not have any. Therefore in order to create machine consciousness we should first seek to create thinking machines.

Our thoughts appear to us as immaterial. This gives rise to the problem of mind-body interaction. Would a thinking machine perceive its thoughts as immaterial, too? Will we create a ghost in the machine if we succeed in the construction of thinking machinery?

Everybody knows how to think, but what is thinking actually? Is it the silent “inner speech” that keeps going on in our heads whenever we are awake? Is there a language of thought with its own grammatical rules? If so, then the brain would assemble thoughts according to the rules of that grammar. But then, what would be the difference between the brain and a computer that also operates by rules? Could we build a cognitive machine by reproducing the rules of the brain in a computer? Could we thus create a thinking robot with a mind of its own and would it also be conscious? The answer to these questions may not be the first thought that comes to mind.

What exactly are the rules of the brain? So far we have had only vague ideas that have not offered a straightforward basis for computational implementation. But, on the other hand, do we actually need to know the rules of the brain? After all what we want are the products of intelligence, the products of thinking? If the answers given by a computer and the brain to a question are the same then aren't the computer and the brain also equal? Thus we would only need to find out computational rules that produce the results of intelligence. This is the “same results” hypothesis that is behind much of the present-day discipline of Artificial Intelligence (AI).

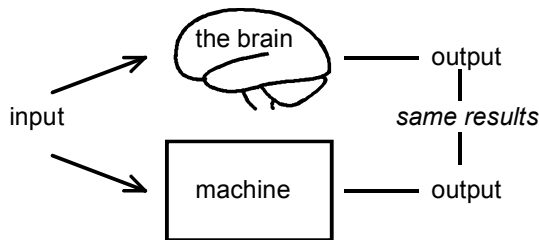


Fig. 1.1. A thinking machine via the “same results” principle?

Computers are numerical machines. Sound and music are represented by strings of numbers describing the intensity of the sound at each moment. Likewise visual information is represented by the brightness values of picture elements. Consequently whatever is to be deduced from the auditory or visual information is achieved via numerical computation. Thus Artificial Intelligence calls for computational rules for each problem in the application. These rule sets, programmed strings of instructions are also called algorithms.

AI has imitated the products of intelligence by dedicated case-by-case algorithms. So far however, these algorithms have not constituted real intelligence outside their limited application, perhaps not even there.

The shortcomings of classical AI can also be attributed to the difficulty or implausibility of foreseeing and preprogramming each and every relevant rule of inference. Self-learning parallel networks, more or less modeled after biological neurons and synapses, have been proposed as a solution as with these no preprogrammed rules are needed. These artificial neural networks, also known as connectionist networks, adjust themselves to large sets of incoming data and produce output according to the implicit rules that they have acquired statistically. However, usually these rules cannot be easily recovered as formulas and the actual workings of a neural network may remain obscure to a human observer.

Traditionally neural computing has been applied to strictly defined problems such as pattern recognition, classification, categorization, function approximation, signal prediction, etc. More often than not neural computing is nowadays performed by a conventional computer so that the speed advantage of the inherent parallelism is lost and the actual process is reduced to a calculation style.

Artificial Intelligence and Neural Networks have produced some remarkable results, but these approaches do not excel in applications where true understanding is needed. Speech, text, story understanding, text summarization or image, scene, episode and movie understanding are especially difficult areas. The performance of programs in these fields has varied from barely acceptable to outright ridiculous. As for general intelligence and true creativity, there has been definitely none. At the end of the day no thinking machine has been built and the human brain still remains far superior to any computer.

So what is wrong here, why has the computational production of “same results” been so elusive?

Is the brain really a rule-based self-programmed computer? Does the brain execute programmed commands? Has Nature devised a program for everything we do, every motion, every emotion, every thought?

No, the brain is definitely not a computer. Thinking is not an execution of programmed strings of commands. The brain is not a numerical calculator either. We do not think by numbers. The brain does not represent its mental content by numbers. Therefore it should not be any wonder that apart some trivial cases “same results” cannot be easily achieved by computational algorithms.

The real sources of mind and intelligence are the non-numeric cognitive processes of the brain, and we should study and seek to reproduce these in order to achieve real machine cognition, comparable

to that of humans. So, instead of the “same results” principle I am proposing here a “same processes” principle.

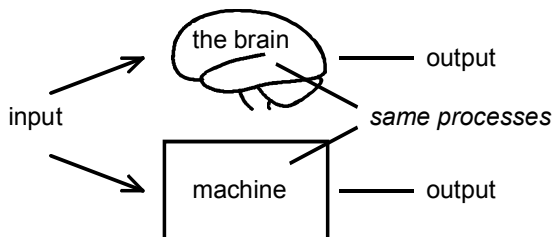


Fig.1.2. A thinking machine via the “same processes” principle

What is involved in the cognitive processing style of the brain? The human mind seems to operate with the flow of inner imagery, inner speech and sensations, with their meanings and significance. Cognition is also accompanied by feelings, emotions and consciousness. We can observe these, but with our natural means we cannot observe any possible material processes behind these. We are not able to introspect the actual neural firings inside the brain. Thus these processes are seemingly immaterial and therefore some people consider their implementation by material means impossible.

However, the seat of thinking and consciousness, the brain, is very much a material entity. Therefore it should be safe to maintain that a material thinking machine can be built because we already have one, the brain.

Thus we should first set out to study the processes of the brain. Neuroscience has found out that the brain uses stereotyped electrical signals for all processing and that the basic signal processing element is a biological cell, the neuron. Now we could start with the operation of the neuron, which is fairly easy to study, and proceed to investigate the combined operation of ever-larger neural assemblies until the whole brain is covered. This however, is a very tedious approach, it would be the same as to figure out how a computer works without any system-level knowledge, only by tracing the signal paths between the millions of transistors. This is not practical and therefore some information about the higher-level cognitive processes would be needed.

This kind of information is provided by cognitive psychology and cognitive neuroscience. The operation of our senses, eyes, ears etc. is quite well known up to the initial processing in the brain. There are general theories for perception, attention, learning, memory, emotions, etc. The search for thinking machines is now transformed into the study

of these cognitive processes and their possible implementation by electronic means.

In principle there are two different approaches available. We could try to create programs that reproduce cognitive processes directly at higher representational level or we could go down to signal level representations and create neural network systems on silicon. The first approach might seem to be more attractive as we could use existing computers for the job. However, there may be some fundamental problems here. We may ask if programmed cognitive processes really equal those of the brain: do we really get the real thing. A part of human cognition deals with representations, these we can easily implement in a computer. Human cognition involves also system reactions that are characterized by a specific feeling such as pain. These can also be represented in a computer but not truly realized because, as I see it, the computer does not have the respective system reactions, therefore it will not feel anything. And if we eventually could get it working it would still basically be a computer, subject to system errors and crashes.

I am proposing here the hard way, realization with novel artificial neurons and non-numeric signal-level representations that carry dedicated meanings. I will also propose a special cognitive architecture to reproduce the processes of perception, inner imagery, inner speech, pain, pleasure, emotions and the cognitive functions behind these. This machine would produce higher-level functions by the power of the elementary processing units, the artificial neurons, without algorithms or programs. However, there is a penalty; this machine cannot be implemented by existing digital microcircuits, therefore dedicated chips would have to be designed.

How about consciousness then? The proposed machine is a thinking machine that will be able to sustain the flow of inner imagery and inner speech, “mental content”. The machine has also the faculties of perception and introspection that allow the awareness of environment and the machine’s mental content. The real challenge of consciousness is its apparent immaterial nature. Does a “conscious” machine really perceive the flow of inner speech and imagery as immaterial? Does it perceive the possession of an immaterial self? Is it aware of its own existence? Answers will be proposed to these questions.

Part I of this book begins with a little bit of history and a review on computing principles. Arguments against the concept of thinking as the execution of strings of program commands are presented. Artificial neural networks have been proposed as a better way to realize thinking machines. Therefore the classical artificial neural network approach is also summarized. Again, arguments against neural network-based thinking machines are presented.

In part II I discuss the fundamentals of cognition and consciousness, the stuff that should be implemented in any real cognitive machine worth its while.

In part III I present my design philosophy and model for machine cognition and consciousness and discuss the general nature of consciousness in the light of this model.