Lecture 22-23
THE LIMITS OF ARTIFICIAL INTELLIGENCE (AI)

Overview

The critics of AI show the limits of artificial intelligence. The computer scientists working for artificial intelligence design the appropriate hardware and programs, which simulate the human mind. For them, mind is the software and the brain is the hardware in which the mind works. Thus, they explain the human mind on the model of a computer. The artificially designed computing machines constitute the bulk of the field as cognitive science called artificial intelligence (AI). These machines do not purport to replace human mind but simulate it by various methods of cognitive modeling.

What will be attempted in this chapter is a critical evaluation of the arguments against AI put forward by Gödel, Searle, Putnam, Penrose, and Dreyfus. We will also critically examine arguments against Fodor’s computational representational theory of mind, in short, CRTM. The philosophers mentioned above propose to argue that there are limits of artificial intelligence which can be philosophically studied and laid down.

Keywords: Artificial Intelligence, Chinese Room Argument, Putnam, Penrose, and Dreyfus

THE LIMITS OF ARTIFICIAL INTELLIGENCE (AI)

I. The General Argument Against AI.

Over the past decades, electronics and computer technology has made great strides in the sphere of knowledge and has helped us in our dealing with the world. The computers of today are much more developed and sophisticated than the mechanical calculators of
yesterday. Already computers are able to perform numerous tasks that had previously been the exclusive province of human beings, with a speed and accuracy that far outstrip anything that a human being can achieve.

Moreover, the advent of computer technology has given a new direction to our understanding of intelligence, thought and other mental activities. We are inclined to raise such questions like: What does it mean to think or to feel? What is mind? Does mind really exist? Besides, we may raise the questions: To what extent are minds functionally dependent upon the physical structures with which they are associated? Are minds subject to the law of physics? If so, what are the laws of physics? Of course, to ask for definite answers to such questions would be a tall order. These questions are eminently philosophical in nature. In philosophy of mind we are interested in understanding the nature of mind, thought, intelligence, etc. as it enables us to appreciate the notions of machine-mind and machine-intelligence.

However, the idea of machine-intelligence has been challenged by philosophers and logicians in recent times. There are a number of results of mathematical logic, which can be used to show that there are limitations to the powers of discrete state machines. The best known of these results is known as Gödel’s theorem, which shows that in any sufficiently powerful logical system statements can be formulated which can neither be proved nor disproved within the system, unless possibly the system itself is inconsistent. As Lucas says, “Gödel’s theorem must apply to cybernetical machines, because it is of the essence of being a machine, that it should be a concrete instantiation of a formal system. It follows that given any machine which is consistent and capable of doing simple arithmetic, there is a formula which is incapable of producing as being true - i.e., the formula is unprovable-in-the-system -but which we can see to be true. It follows that no machine can be a complete or adequate model of the mind, that minds are essentially different from machines.”¹ According to Gödel, no logical system can be held to be self-

complete, so it always needs another system to prove its completeness. Now, in the light of this, it can be shown that no machine, like a logical system, can be self-complete, that is, cannot do everything. That is to say, there will be some questions to which it will either give a wrong answer, or fail to give an answer at all, however, much time is allowed for a reply. There may be many such questions, that cannot be answered by one machine, but may be satisfactorily answered by another. Thus machines have limitations in their functions. Therefore, Turing says, “I grant you that you can make machines do all the things you have mentioned but you will never be able to make them do X.” He mentioned numerous features of X in this connection, which could not be performed by a machine. These features are mentioned by him in the following passage:

“Be kind, resourceful, beautiful, friendly, have initiative, have a sense of humor, tell right from wrong, make mistakes, enjoy strawberries and cream, make someone fall in love with it, learn from experience, use words properly, be the subject of its own thought, have as much diversity of behaviour as a man do something really new.” These are the various disability arguments

A man uses a number of machines in his lifetime. The general belief that machines cannot commit mistakes is not always true. Machines do go wrong. Turing’s notion of imitation game may be taken into account. In this game machines would be given a set of problems to solve. One would deliberately introduce mistakes in a manner calculated to confuse the machine. As a result, a mechanical fault would probably show itself through introduction of a wrong move in the calculation. Thus machines are vulnerable to all kinds of mistakes.

---


3 Ibid.
The above criticism of machines shows two kinds of mistakes that the machines can commit. We may call them ‘errors of functioning’ and ‘errors of conclusion.’ Errors of functioning are due to some mechanical or electrical faults, which cause the machine to behave otherwise than it is designed to do. In philosophical discussions, one likes to ignore the possibility of such errors, because we are discussing ‘abstract machines.’ These abstract machines are mathematical fictions rather than physical objects. By definition they are incapable of errors of functioning. In this sense, we can say that, machines can never make mistakes. However, the machines can commit errors of conclusion because they can make mistake moves in this function. These mistaken are the errors of argument.

According to Lovelace, “The Analytical Engine has no pretensions to originate anything. It can do whatever we know how to order it to perform.” Here, the analytical Engine referred to is a universal digital computer, which can be made to perform many task but cannot originate anything on its own. That is to say, as a machine it fails in the matter of creativity. It is for this reason we can say that a machine can ‘never do anything really new.’ In comparison to the machine, we can argue, the human mind is not a machine at all, since it originates many new things. The human system as a whole is a creative system which cannot be mimicked by any machine.

Again, we may argue that human beings have some psychological qualities such as intelligence, consciousness or originality, etc. which are said to be necessarily lacking in machine. That is why, a machine is normally treated as an artefact and a mere mechanical contrivance manufactured for a definite purpose.

---

4 *Ibid,* p.18
However, when it is said that it is impossible for a machine to be conscious, it is not always clear to what extent this is intended to be a logical objection, and to what extent empirical. Empirically, machines are not conscious, but this cannot be proved logically. Robots are well known for duplicating human behaviour. For a robot, ex-hypothesis, is capable of behaving like a human being. We have no doubt that a human being is conscious, when he or she is doing work. Though a machine might do the same work, we are not inclined call the latter conscious. Thus it is taken for granted that humans are conscious, whereas of machines we enquire whether they are capable of consciousness or not. We know that the question of consciousness is appropriate in the context of human beings, but not so in the case of machines. A machine is essentially distinct from a man so far as consciousness is concerned. The machine-intelligence and machine-behaviour are not indicative of consciousness at all.

Michal Scriven has raised the question as to whether a robot can be considered as conscious. He is of the view that a robot is simply a machine which is indistinguishable from humans in behavioural aspects. In spite of the close similarity in behaviour, however, human and robots are essentially different, belonging to two different types of entities. Campbell, therefore, has argued, following Descartes, that human beings alone and not machines are conscious. Machines are unconscious material bodies while human beings are conscious entities.

Here, a question may arise: Is it a blind prejudice which prevents us from extending the attribute of consciousness to robots, when robots can calculate more quickly, react more swiftly, see more clearly, and remember more accurately than men? What is it that they lack when they can do so many things? They do what humans do, yet they cannot be treated at par with human beings. The obvious answer to this is that the robots have no consciousness. They are only machines imitating human beings.
As Michael Scriven points out, the sense of ‘conscious’ is contrasted with ‘incapable of being conscious’ and we can ascribe consciousness to creatures only if we can also withhold such ascription to them. He says that it is absurd to ask of a stone or a stopwatch ‘Is it conscious?’ because it is absurd to talk of its being dead, asleep, drugged or stunned, i.e., unconscious. There are cases where it is very difficult to decide the question of consciousness. For example, let us take a man who has a completely anaesthetized cortex, and who has an external operator controlling his or her outer behaviour. In such cases, the man is unconscious, though outwardly conscious. But such difficulties arise only about conscious beings. In this case, the man was conscious but now has no consciousness, his brain beings out of action temporarily.

It is now certain that under no circumstances can we prove that the robot is conscious. We have a complete causal explanation of all its behaviour, and this explanation does not at any stage depend on its consciousness. So its behaviour cannot be a proof of the possession of consciousness. Following the above argument, Scriven holds that consciousness is not a property, which can be detected in a machine by any physical examination, because it cannot be identified with any physical characteristic of a machine.

Therefore, even if a computer does exactly what a human being does, it can never be ascribed consciousness. It never does anything creative or new or unpredictable. Its output is the result of its physical structure, its program, and the input it is given. A human being, on the other hand, initiates novel, creative and unpredictable actions. Thus a human being stands on a different footing from the computer. This argument can be laid against AI, since there is a wide logical gap between human beings and the computing machines. Computers not only lack creativity but they lack basic capacity to learn. Many people take unpredictability as an evidence for originality, and fear that if it is true that mentality bottoms out in straight forwardly mechanical processes, we eventually will be able to predict everything about people. And at that point human life will lose its joy and mystery. Hence we can argue that people, and not machines have creativity.
In defence of AI, however, one may say that the computer is an ideal model of mentality, because it is an ideal model of the brain. But the computer, in practice, does not simulate all the functions of the brain, and so remains an incomplete model. The Von Neumann devices also have little in common with brains. Even the claim that connectionist machines are biologically realistic requires a good deal of charity. They are more like brains than the Von Neumann devices are, but they are a lot less like brains than other brains are. Thus it can be said that computers will never come to possess genuine mentality because they do not have the human brain.

Proponents of weak psychological AI claim that we can write programs that test the relative plausibility of different psychological hypotheses. For example, one can write programs that purport to describe the cognitive mechanisms underlying language production. The correct programs will be the ones which pass the test of descriptive adequacy. However, the proponents of AI would not be able to claim that computers can accurately simulate human behaviour because they do not posses the competence of the human brains to produce conscious activity. Despite the differences between computers and brains, there is no reason to think that computers cannot represent any relevant information we desire about neural processes. The point is that a computational system can simulate a brain system without being just like the brain.

II. Gödel’s Argument.
According to J.R. Lucas, Gödel’s theorem states that in any consistent system, which is strong enough to produce simple arithmetic, there are formulas which cannot be proved-in-the-system, but which we can see to be true. Such a formula which is the, “This formula is unprovable-in-the-system.” If this formula were provable-in-the-system, then it will be unprovable-in-the-system. So there will be a contradiction. So the formula ‘This formula is unprovable-in-the-system’ is not provable-in-the-system, but unprovable-in-the-system. Further, if the formula is ‘This formula is unprovable-in-the-system’, then it is true that the formula is unprovable-in-the-system, that is, ‘This formula is unprovable-in-the-system’ is true.

The whole effort of Gödel’s theorem is to show that all formal systems which are (i) consistent, (ii) adequate for simple arithmetic, i.e., contain the natural numbers and the operations of addition and multiplication, and that (iii) they are incomplete, i.e., contain unprovable, though perfectly meaningful, formulae, which we can see to be true, standing outside the system.

According to J.R. Lucas, Gödel’s theorem must apply to cybernetical machines, because it is of the essence of being a machine that it should be a concrete instantiation of a formal system. It follows that given any machine that is consistent and capable of doing simple arithmetic, there is a formula, which, though true, is not provable in the formal system of the machine. Thus it follows that no machine can be a complete or adequate model of the mind; that is, the minds are essentially different from machines.

As we know, a cybernetical machine is a device, which performs a set of operations according to a definite set of rules. Normally we ‘program’ a machine, that is, we give it a set of instructions about its functioning and we feed in the initial ‘information’ on which the machine is to perform its calculations. When we consider the

5 Lucas, J. R., “Mind, Machines And Gödel” in Minds and Machines, A. R. Anderson (ed.), p.44.
mind on the model of cybernetical mechanism, we have a similar model in view. If human mind is such a model, mind is determined by the way it is made. Then, there is no possibility of its acting on its own, as it is governed by certain rules of construction and certain input of information. But this is not the characteristic of mind, as the mind does not act under ready-made rules.

In the machine, there are some formal rules of inference having been applied to some previous formula. According J.R. Lucas, we can construct a Gödelian formula in this formal system. This formula cannot be proved-in-the-system. Thus the machine cannot prove the corresponding formula as true. But one can see that the Gödelian formula is true. We can now see that any mechanical model of the mind must include a mechanism that can elucidate truths of arithmetic, because this is something, which minds can do. In fact, it is easy to produce mechanical models which will in many respects produce truths of arithmetic far better that what the human beings can do, but for every machine there is a truth which it can not prove, but which can be proved by the mind. Thus in the words of Lucas, “This is not to say that we cannot build a machine to simulate any desired piece of mind-like behaviour: it is only that we can not build a machine to simulate every piece of mind-like behaviour. We can (or shall be able to one day) build machines capable of reproducing bits of mind-like behaviour, and indeed of outdoing the performances of human minds: but however good the machine is, and however much better it can do in nearly all respects than a human mind can, it always has this one weakness, this one thing which it cannot do, where as a mind can.”

Moreover, Gödels argument shows that the mechanical model of mind, because of its inherent limitations cannot simulate the functions of the mind which are infinite and

---

6 Ibid., p.47.
indefinite. Further, it shows that machines are finitely closed and hence cannot be compared with human minds.

III. Searle’s Argument Against AI.

Searle’s main intention is to criticize and overcome the dominant traditions in the study of minds, both ‘materialist’ and ‘dualist’. For him, consciousness is central to the mental phenomena. We think of ourselves as conscious, mindful, rational agents in the world, but science tells us that the world consists entirely of mindless physical particles. But, the question is: How can we match these two conceptions? According to Searle, can it be the case that the world contains nothing but unconscious physical particles, and yet that it also contains consciousness? Can an essentially meaningless world contain meanings? Searle writes, “I believe that the mind-body problem has a rather simple solution, one that is consistent both with what we know about neurophysiology and with our commonsense conception of the nature of mental states –pains, beliefs, desires and so on. But before presenting that solution, I want to ask why the mind-body problem seems so intractable. Why do we still have in philosophy and psychology, after all these centuries, a ‘mind-body problem’ in a way that we do not have, say, a ‘digestion-stomach problem’? Why does the mind seem more mysterious than other biological phenomena?”

Moreover, for Searle, all the above problems spill over into other contemporary materialistic interpretations of the issues of mind. Materialism asks the question: How should we interpret the recent work in computer science and artificial intelligence aimed at making intelligent machines? More particularly, does the digital computer give us the

right picture of the human mind? Thus the central issue is: What is the relation between the ordinary, commonsense explanations of people’s behaviour and its scientific modes of explanation? Searle seeks to answer this question in his attack on materialism in his philosophy of mind.

Searle offers a biological explanation of mind according to which mind is a biological offshoot of the brain. In order to distinguish this view from others in the field, Searle calls it “biological naturalism.” Mental events and processes are as much part of our biological natural history as digestion, mitosis, meiosis, or enzyme secretion, says Searle.

The biological naturalism raises many questions of its own. But one of the fundamental questions is: What about the great variety of our mental life—pains, desires, tickles, thoughts, visual experiences, beliefs, tastes, smell, anxiety, fear, love, hate, depression and elation? Again, some of the philosophical questions, which were raised by Searle, are: What exactly is consciousness and how exactly do conscious mental phenomena relate to the unconscious? What are the special features of the ‘mental’, phenomena such as consciousness, intentionality, subjectivity, and mental causation? And how exactly do they function? What are the causal relations between ‘mental’ phenomena and ‘physical’ phenomena? And can we characterize those causal relations in a way that avoids epiphenomenalism?

Searle’s biological naturalism provides an effective counter argument to the currently fashionable computational theory of mind according to which, the mind is a computer program. According to this theory, the mind is to the brain what the program is to the hardware. In short, minds are computer programs implemented in brains. In Searle’s words: “The brain is just a digital computer and the mind is just a computer

---

program. One could summarize this view— I call it ‘strong artificial intelligence’ or ‘strong-AI’—by saying that the mind is to the brain, as the program is to the computer hardware.”

The notion of strong AI is called by Dennett ‘computer functionalism’. Both the discipline of artificial intelligence and the philosophical theory of functionalism converge on the idea that the mind is just a computer program. For both the theories the human mind is a computational system that realizes programs, that is, it is a formal device. That produces functions of various kinds called the mental functions. It is a system which functions with the right inputs and outputs so that the resulting activities are treated as the mental activities. The supporters of strong AI say that there is a general agreement among them that it is only a matter of time until the computer scientists and the workers in artificial intelligence design the appropriate hardware and programs, which will be the equivalent of human brains and minds. These will be artificial brains and minds, which are in every way the equivalents of human brains and minds. As Herbert Simon says, “We already have machines that can literally think. There is no question of waiting for some future machine, because existing digital computers already have the same sense that you and I do.” That is, the idea of a thinking machine is no more a dream but a reality. Hence, the legitimacy of strong artificial intelligence.

Alan Newell holds a similar view when he says, “we have now discovered that intelligence is just a matter of physical symbol manipulation; it has no essential connection with any specific kind of biological or physical wetware or hardware. Rather, any system whatever that is capable of manipulating physical symbols in the right way is capable of intelligence in the same literal sense as human intelligence of human

9 Searle, John R., Minds, Brains and Science, p.28.

10 This is quoted by John R. Searle, in his Minds, Brains and Science, p.29.
beings."\textsuperscript{11} Marvin Minisky holds that, “the next generation of computers will be so intelligent that we will be lucky if they are willing to keep us around the house as household pets."\textsuperscript{12} McCarthy says that even “machines as simple as thermostats can be said to have beliefs.”\textsuperscript{13} All these declarations by eminent scientists justify the idea of strong AI.

Searle, however, refutes the very idea of strong AI. His argument against AI has nothing to do with any particular stage of computer technology. It is important to emphasize this point because the temptation is always to think that the solution to our problems must wait on some as yet uncreated technological wonder. This refutation has to do with a definition of digital computers, and the idea of artificial intelligence underlying it.

As we know, the conception of a digital computer is that its operations can be specified purely formally; that is, we can specify the steps in the operation of the computer in terms of abstract symbols-sequences of zeros and ones printed on a tape. But the symbols have no meaning; they have no semantic content, they are not about anything. They have to be specified purely in terms of their formal or syntactical structure. By definition, our internal mental states have certain sorts of contents. Searle says, “In a word, the mind has more than a syntax, it has a semantics. The reason that no computer program can ever be a mind is simply that a computer program is only syntactical, and minds are more than syntactical. Minds are semantic in the sense that

\textsuperscript{11} This is quoted by John R. Searle, in his \textit{Minds, Brains and Science}, p.30.

\textsuperscript{12} This is quoted by John R. Searle, in his \textit{Minds, Brains and Science}, p.30.

\textsuperscript{13} This is quoted by John R. Searle, in his \textit{Minds, Brains and Science}, p.30.
they have more than a formal structure, they have content.”

Searle presents a thought experiment about a Chinese Room for refuting the possibility of AI. This is called the Chinese Room Argument. He asks us to imagine that the computer programmers have written a program that will enable a computer to simulate the understanding of Chinese. Thus, for example, if the computer is given a question in Chinese, it will match the question with its memory, or database, and produce appropriate answers to the questions in Chinese. Suppose that the computer’s answers are as good as those of a native Chinese speaker. Then the question is: does the computer literally understand Chinese in the way the Chinese speakers understand Chinese? Again, let us imagine that someone is locked in a room, with several baskets full of Chinese symbols. However, let’s imagine that he or she does not understand a word of Chinese and he or she is given a rulebook in English for manipulating these Chinese symbols. The rules specify the manipulations of the symbols purely formally, that is, in terms of their syntax, but not their semantics. So the rules might say, ‘take a squiggle-squiggle sign out of basket number one and put it next to a squiggle squiggle sign from basket number two.’ Now suppose that some other Chinese symbols are passed into room, and he is given further rules for passing back Chinese symbols out of the room. Supposing that unknown to him, the symbols passed into the room are called the ‘questions’ by the people outside the room, and the symbols he passes back out of the room are called ‘answers to the question’. Furthermore, the programmers are so good at designing the programs that the person in the Chinese Room can easily manipulate symbol so that very soon the answers are indistinguishable from those of a native Chinese speaker. In this case the man in the Chinese Room manipulates Chinese symbols mechanically without understanding what they mean. Yet his answers are indistinguishable from those of the native Chinese speakers.

---

14 Ibid., p.31.

15 Ibid., p.32.
The above situation shows that a computer has syntax, but not semantics. Indeed, understanding a language having mental states at all, involves more than just having a bunch of formal symbols. It involves having a meaning attached to those symbols. And a digital computer, as defined above, cannot have more than just formal symbols, because it operates, as Searle says, in terms of its ability to implement programs. As these programs are purely formal they cannot have semantic content.

The supporters of AI argue that we can feed the understanding of Chinese into a robot. If the robot operates the Chinese symbols property, would not that be enough to guarantee that it understands Chinese? Searle replies that the robot lacks understanding. Even though it might behave exactly as if it understands Chinese, it would still have no way of getting from the syntax to the semantics of Chinese. Thus there is no way that the supporter of strong AI can argue that the mind consists of purely formal or syntactic operation, and the mind is nothing but a computing machine.

Searle’s Chinese Room Argument is concerned with the issue of understanding and the question whether an appropriately sophisticated computer action can be said to have mental properties. It is concerned with some programs that purport to simulate human understanding by providing replies to questions in Chinese by following purely formal rules. However, despite the appearance of understanding that is involved in the computational output when it performs the computations, no such understanding is actually experienced by the computer performing manipulations that enact these computations. Accordingly, Searle argues that the mental quality of understanding cannot be just a computational matter. It is because the computer is unable to duplicate human intelligence, though it has the ability to simulate the latter. Here, the key distinction is between duplication and simulation. And no simulation by itself ever constitutes duplication. At the end of the argument, he says, “for any artefact that we might build which had mental states equivalent to human mental states, the implementation of a
computer program would not by itself be sufficient. Rather the artefact would have to have powers equivalent to the power of the human brain.”16

In his paper “Computing Machinery and Intelligence,”17 Turing has suggested an operational test for the machine-intelligence in the form of an ‘imitation game’. Accordingly, if a computing machine can give responses to questions that make it impossible for us to distinguish this computer from a fellow human being, then we can test whether a machine can think or not.

Searle objects to the Turing Test on the ground that the normal criteria we apply in ascribing intelligence to persons are based on behavioural, biological, and phonological evidence. According to him, normal human beings have intentionality, consciousness, and free will, etc., which the machines like computers lack. In effect, the Turing test is a form of reverse discrimination against humans, as it shows the humans in a poorer light in comparison the machines which are made by the human beings.

Searle is certainly correct in saying that merely instantiating a computer program is not sufficient for the possession of our kind of mentality. Mere exhibition of a formally accurate operation does not suffice to make the operation intelligent in the human sense. The fact that human beings have intelligent operations of mind is biologically conditioned and cannot be transformed to non-human machines.

16 Searle, John R., Minds, Brains and Science, p.41.

Searle offers two different sets of criteria for applying the expression ‘intelligent behaviour.’ One of these sets consists of third-person or ‘objective’ criteria that are not necessarily of any subjective psychological interest whatever. But the other set of criteria are essentially subjective and involve the first-person points of view. According to him, ‘intelligent behaviour’ on the second set of criteria involves thinking, and thinking is essentially a subjective mental process. Now, if we adopt exclusively the third-person criteria to intelligent-behaviour, then computers, such as not to mention pocket calculators, cars, thermostats, and indeed just everything in the world, engage in intelligent behaviour. But this yields no specific result regarding intelligent behaviour of machines.

IV. Putnam’s Argument Against AI

In this section, I shall discuss the reasons that led Putnam to propose functionalism as a theory of mind supporting artificial intelligence and the reasons that subsequently led him to abandon it. I would like to discuss Putnam’s views as belonging to early Putnam and later Putnam. Early Putnam has argued for the possibility of robotic consciousness. As a functionalist, early Putnam shows that a human being is an automation: that is, the human mind is a computing machine. The later Putnam, however, has found that his earlier thesis was wrong as mind can never be reduced to a machine.

Functionalism is the view that mental states are defined by their causes and effects. It holds that what makes an inner state is not an intrinsic property of the state, but rather its relations to sensory stimulation (input), to other inner states, and to behaviour (output). And according to the functionalists, all these functional states are multiply realizable in different kinds of machines. The developments in computer science have given impetus to functionalism. Firstly, the distinction between software and hardware suggested the distinction between function and structure. Secondly, since computers are automated, they demonstrate how inner states can be causes of output in the absence of a homunculus. Thirdly, the Turing machine provided a model for functionalism. According
to Turing machine functionalism, each psychological state is identical to a Turing machine state. This Turing machine functionalism is largely developed by early Putnam. Thus, in short, ‘functionalism’ may be defined as the theory that explains mental phenomena in terms of the external input and the observable output. It explains the mind as a complicated machine.

According to Putnam, autonomy of our mental life has nothing to do with the old question about the soul-stuff. As Putnam puts it, “If it is built into one’s notion of the soul that the soul can do things that violate the laws of physics, then I admit I am stumped. There cannot be a soul which is isomorphic to a brain, if the soul can read the future clairvoyantly, in a way that is not in any way explicable by physical law. On the other hand, if one is interested in more modest forms of magic like telepathy, it seems to me that there is no reason in principle why we could not construct a device which would project subvocalised thoughts from one brain to another. As to reincarnation, if we are, as I am urging, a certain kind of functional structure (my identity is, as it were, my functional structure) there seems to be in principle no reason why that could not be in reproduced after a thousands or a millions years or a billion years. Resurrection: as you know, Christians believe in resurrection in the flesh, which completely bypasses the need for an immaterial vehicle. So even if one is interested in those questions even then one does not need an immaterial brain or soul stuff.” Thus, according to Putnam, there is a functional aspect of the human mind, which can be analyzed in multiple systems. Functionally, the mind is isomorphic with the system realizing the mental functions. That is to say, two systems are functionally isomorphic if there is a correspondence between the states of one and the states of the other that preserves functional relations. For example, if the functional relations of a computing machine are just sequence relations, that is, state A is always followed by state B, then for F to be functional isomorphism, it must be the case that state A is followed by state B in system one, if and only if state F (A) is followed by state F (B) in system two. The functional relations are data or printout relations.

In this connection, Putnam points out that the traditional mind-body problems are wholly linguistic and logical in character. All issues relating to mind-body problem are concerning the computing systems capable of answering questions about its own structure, and have nothing to do with the unique nature of human subjective experience. One kind of puzzle that is discussed sometimes in connection with the ‘mind-body problem’ is the puzzle of privacy. In the functionalist theory of mind, however, privacy as a category as a category disappears altogether as there are no ‘qualia’ any more linked with the human mind.

Putnam characterizes a Turing machine, as that which generates theories, tests them and asserts theories, and follows some rules. In particular, if the machine has electronic ‘sense organs’ which enable it to ‘scan’ itself, while it is in operation, it may formulate theories concerning its own structure and subject them to test. Suppose the machine is in a given state A, when, and only when flip-flop 36 is on. Then this statement, ‘I am in state A, when, and only when flip-flop 36 is on,’ may be one of the theoretical principles concerning its own structure accepted by the machine. Of course, here, ‘I am in state A’ is the ‘observation language’ for the machine, while ‘flip-flop 36 is on’ is ‘theoretical expression’ which is particularly interpreted in terms of the ‘observable’.19 Now all of the usual considerations for and against mind-body identification can be paralleled by considerations for and against saying that state A is in fact identical with flip-flop 36 being on. Putnam thus holds that if the mind-body identity theory were true then it would have to be true as a consequence of the meaning of psychological words. If we take the question whether light is electro-magnetic radiation of such and such wave/length, it would lead to the conclusion that this too was not a question of empirical fact but called for a ‘decision’ on our part, a decision to treat

electro-magnetic radiation in a certain way. Still, light is not identical with electro
magnetic radiation.

Now the question arises: Does a computing machine have intelligence, consciousness, and so on, in the way that human beings do? According to Putnam, since mind is a Turing machine, the whole human body is a physical system obeying the laws of Newtonian physics. The universe as a whole is a machine too. Thus, Putnam’s argument shows that the whole human body is at least metaphorically a machine.

Putnam has taken the robot to be a ‘psychological isomorphic’ to a human being. However, it can be seen that this is not actually possible, because the epistemological, metaphysical and moral arguments show that there is no isomorphic relationship between the humans and robots. If machines were conscious, they would have feelings, thoughts, attitudes, etc. But now the question is; is it really possible? If it is possible, then what are the necessary and sufficient conditions? Regarding this, Paul Ziff\(^\text{20}\) wishes to show that it is false that a robot is conscious. He begins with the undoubtful fact that if a robot is not alive, it cannot be conscious. Here, Ziff has given the semantical connection between ‘alive’ and ‘consciousness’ in English in view of the fact that the meaning of ‘alive’ is connected with that of ‘conscious’. A robot is not a living entity and so cannot be conscious. This semantic connection shows that a robot is not alive. Thus from Ziff’s argument it is clear that Putnam is wrong in holding an isomorphic relation between human being and robot.

The theory that proposes to provide a complete description of our psychological states as a Turing machine is a utopian project. Putnam realized this later because this sort of utopianism is an illustration of what is called ‘scientism.’ It is based on speculations regarding scientific possibilities. The problem is that it is completely unclear just what possibility has been envisaged when one speaks about robotic consciousness. While arguing against AI, the later Putnam points out that pessimism about the success of

AI in simulating human intelligence amounts to pessimism about possibility of describing the functions of the brain.

Moreover, the later Putnam mentions that functionalism is incompatible with our semantic externalism because the mechanistic view of the mind does not square with meaning and representation developed within a semantic theory. The semantic theory possesses an externalist relation between meaning and the external world. Putnam takes meaning, not as a mental or psychological content, but as a content conditioned by the external world.

Putnam has rejected the computational view of mind on the ground that the literal Turing machine would not give a representation of the psychology of human beings and animals. For him, functionalism is wrong in holding the thesis that propositional attitude is just a computational state of the brain. For example, to believe that there is a cat on the mat, is not the same thing as that there is one physical state or a computational state believing that there is a cat on the mat. Therefore, it is not right to hold that propositional attitudes are semantically or conceptually reducible to computational predicates. According to Putnam, this is impossible because propositional attitudes express to the intentional states, that is to say that they refer to various states of affairs in the world. Thus, according to him, the functionalist is wrong in saying that semantic and propositional attitude predicates are semantically reducible to computational predicates, which can be realized in a physical system like the human brain. There is no reason why the study of human cognition requires that we try to reduce cognition either to computations or to brain processes. The reductionist approach to functionalism gives an inadequate picture of the human mind.

V. Dreyfus’s Argument Against AI
In ‘What Computers Cannot Do’ Dreyfus argues that research in artificial intelligence was based upon mistaken assumptions, which included psychological, epistemological, biological and ontological assumptions about the nature of human knowledge and understanding. We will see now what these assumptions are.

The psychological assumption is that the mind can be viewed as a device operating on bits of the mind according to formal rules. Thus, in psychology, the computer as a model of the mind is conceived of by the cognitive scientists.

The epistemological assumption is that all knowledge can be formalized in terms of logical relations, and more exactly in terms of Boolean functions, i.e., the logical calculus which governs the way the bits are related according to rules.

A biological assumption is that the brain has neurons, which operates so as to process information in the brain according to a neural network.

The ontological assumption is that the computer model of mind presupposes that all relevant information about the world, everything essential to the production of intelligent behaviour, must in principle be analyzable as a set of situation-free determinate elements.

The psychological, epistemological, biological, and ontological assumptions have this in common: they assume that man must be a device which calculates according to rules on data which take the form of atomic facts. Dreyfus argues that all these

---

assumptions can be criticised on philosophical grounds. Each of the assumptions leads to conceptual difficulties. As, he says, “... among philosophers of science one finds that an assumption that machines can do everything that people can do, followed by an attempt to interpret what this bodes for the philosophy of mind; while among moralists and theologians one finds a last-ditch retrenchment to such highly sophisticated behaviour as moral choice, love and creative discovery, claimed to be beyond the scope of any machine.”

The assumption that machines can do everything that human beings can do is definitely false as the human capacity exceeds that of the machine. All the above-mentioned assumptions are definite because they assume more than they can prove. The idea that the human mind functions like a digital computer is, according to Dreyfus, inadequate and misleading.

Dreyfus in his famous article ‘Misrepresenting Human Intelligence’ points out that the research in AI has misrepresented the nature of human intelligence because it emphasizes that the computers have capacity to understand language processing, pattern recognitions, the problem solving, etc. But this is only a poor ‘imitation of’ what human beings can naturally do. As Dreyfus writes, “The subsequent failure of every attempt to generalize micro-world techniques beyond the artificially restricted domains for which they were invented has put an end to the hopes inspired by early micro-world successes an brought AI to a virtual… And the prospects for programming a digital computer to display our everyday understanding of the world were looking less bright all the time.

Cognitive scientists were discovering the importance of images and prototypes in human understanding.\textsuperscript{23}

Dreyfus points out that the AI field of research dedicated to using digital computers to simulate intelligent behaviour, soon came to be known as ‘artificial intelligence’. One should not be mislead by the name. No doubt an artificial nervous system sufficiently like the human one, with other features such as sense organs and a body would be intelligent. But the term ‘artificial’ does not mean that workers in artificial intelligence are trying to build an artificial man. Given the present states of physics, chemistry, and neurobiology, such an understanding is not feasible. Likewise, the term ‘intelligence’ can be misleading. No one expects the resulting robot to reproduce everything that counts as intelligent behaviour of human beings.

According to AI scientists, “any complete description of behaviour should be adequate to serve as a set of instructions, that is, it should have the characteristics of a plan that could guide the action described.”\textsuperscript{24} But, as Dreyfus argues, that what instructions could one give a person about to undertake the action? Perhaps some very general rules such as ‘listen to the instructions’, ‘look toward an object’, ‘make your selection’, etc. It is not clear why or how a complete description in psychology should take the form of a set of instructions.


Again, AI scientists say that human bodies are part of the physical world and objects in the physical world have been shown to obey laws which can be expressed in a formalism manipulable on a digital computer. To be more particular, if the nervous system obeys the laws of physics and chemistry, then it is bound to be a part of the physical world. Accepting the fundamental assumption that the nervous system is a part of the physical world and that all physical processes can be described in a mathematical formalism which can in turn be manipulated by a digital computer, one can arrive at the strong claim that the behaviour which results from human ‘information processing,’ whether directly formalizable or not, can always be indirectly reproduced on a digital machine. Against the above view, Dreyfus argues that every form of information processing cannot in principle be simulated by a digital computer. Therefore, the strong claim that every form of information processing can be imitated by a digital computer is misleading.

However, when Minsky and Turing claim that man is a Turing machine, they cannot mean that humans are a physical system. Otherwise it should be appropriate to say planes or boats are Turing machines, because their behaviour can be described mathematically formulable laws. The human systems are category-different from the physical systems because they cannot be described in terms of purely physical laws.

Arguing against the epistemological hypothesis, Dreyfus says, “is there reason to suppose that there can be a formal theory of what linguists call pragmatics? There are two reasons to believe that such a generalization of syntactic theory is impossible: (1) An argument of principle; for there to be a formal theory of pragmatics, one would have to have a theory of all human knowledge; but this may well be impossible. (2) A descriptive objection: not all-linguistic behaviour is rule-like. We recognize some linguistic expression as odd-as breaking the rules- and yet we are able to understand them.”

---

25 Ibid., p.198.
More clearly, there are cases in which a native speaker recognizes that a certain linguistic usage is odd and yet is able to understand it. For example, the phrase ‘the idea is in the pen’ is clear in a situation in which we are discussing promising authors. But in fact, an idea cannot be in the pen, because obviously an idea is not a physical object.

As we know programmed behaviour is either arbitrary or strictly rule-like. Therefore, in confronting a new usage, a machine must either treat it as a clear case falling under rules, or as arbitrary. A native speaker feels he or she can recognize the usage as odd, not falling under the rules, and yet can make sense of it, give it a meaning in the context of human life. These usages which are arbitrary are likely to be understood in the context of human activities.

Following above argument Dreyfus, quoting from Weizenbaum, says: “I call attention to the contextual matter . . . to underline the thesis that, while a computer program that ‘understands’ natural language in the most general sense is for the present beyond our means, the granting sense is for the present beyond our means, the granting of even a quite broad contextual framework allows us to construct practical language recognition procedure.” To see this we must show that Weizenbaum’s way of analyzing the problem separating the meaning of the context from the meaning of the words used in the context is not accidental but is dictated by the nature of a digital machine. In our everyday experience we do not find ourselves making such a separation. We seem to understand the situation in terms of the meaning of the words as much as we understand the meaning in terms of the situation. But for a machine this reciprocal determination

must be broken down into a series of separate operations. Since Weizenbaum sees that we cannot determine the sense of the words until we know the meaning of the context, he correctly concludes from a programmer’s point of view that we must first specify the context and then use this fixed context to determine the meaning of the element in it.

The Dreyfus critique therefore is not addressed against computer per se, but against one particular way of programming them. Dreyfus seems willing to grant that machine intelligence can replace human intelligence. This shows the limits of artificial intelligence as a programme.

VI. Penrose’s Arguments

Roger Penrose, While arguing against AI, says, “when I assert my own belief that true intelligence requires consciousness, I am implicitly suggesting (since I do not believe the strong AI contention that the mere enaction of an algorithm would evoke consciousness) that intelligence cannot be properly simulated by algorithmic, i.e. by a computer, in the sense that we use that term today. For I shall shortly argue that there must be an essentially non-algorithmic ingredient in the action of consciousness.” 27 His suggestion is that unconscious actions of the brain are ones that proceed according to algorithmic rules, whereas the conscious acts of the mind are non-algorithmic. Penrose discusses this nature of consciousness and computation, and provides an answer to the question whether our conscious awareness of happiness, pain, love, aesthetic sensibility, will, understanding, etc. can fit into a computational model of mind. His argument consists in the following proposition:

(a) All thinking is computation, that is, all cognitive acts can be mathematically computed.

---

(b) Physical actions of the brain can be simulated computationally, but this computational simulation itself cannot evoke awareness.

(c) Awareness cannot be explained by physical, computational, or any other scientific terms.\(^\text{28}\)

Awareness, understanding, consciousness, intelligence, perceptions, etc are all our intuitively given mental activities. These cannot be computationally explained according to Penrose. Thus, according to him, for example, ‘intelligence’ requires ‘understanding’ and ‘understanding’ requires ‘awareness’. Awareness is a basic feature of consciousness. These mental activities are basic to the human mind. Penrose remarks, “… a person’s awareness is to be taken, in effect, as a piece of software, and his particular manifestation as a material human being is to be taken as the operation of this software by the hardware of his brain and body.”\(^\text{29}\)

However, human awareness and understanding are not the result of computations undertaken by the brain. Understanding is the inborn activity of the human mind which cannot be simulated by a computer. Human understanding cannot be replaced by computer simulations. The strong AI, much against our ordinary understanding of the mental activities, tries to reduce them to computational functions. In the words of Penrose: “Thus, according to strong AI, the difference between the essential functioning of a human brain (including all its conscious manifestations) and that of a thermostat lies only in this much greater complication (or perhaps ‘higher-order structure’ or ‘self-referential properties’, or some other attribute that one might assign to an algorithm) in the case of a brain. Most importantly, all mental qualities – thinking, feeling, intelligence, intelligence,


understanding, consciousness—are to be regarded, according to this view, merely as aspects of this complicated functioning; that is to say, they are features merely of the *algorithm* being carried out by the brain.”

It is, therefore, obvious that the strong AI cannot explain the mental activities properly, because it misses the very non-computational and non-algorithmic nature of the mental activities. Penrose says that in the human mind, there is non-verbality of thought. In order to make his argument stronger, he quotes Francis Galton who said, “it is a serious drawback to me in writing, still more in explaining myself, that I do not think as easily in words as otherwise. It often happens that after being hard at work, and having arrived at results that are perfectly clear and satisfactory to myself, when I try to express them in language I feel that I must begin by putting myself upon quite another intellectual plane. I have to translate my thoughts into a language that does not run very evenly with them. I therefore waste a vast deal of time in seeking appropriate words and phrases, and am conscious, when required to speak on a sudden, of being often very obscure through mere verbal maladroitness, and not through want of clearness of perception. That is one of the small annoyances of my life.” Once it is accepted that much of conscious thinking can be of a non-verbal character, as described above, it follows that the non-verbal thought can never be computational in character.

Thus mathematical activity is a very tiny area of conscious activity that is indulged in by a small minority of conscious beings for a limited fraction of their conscious lives. There is a vast area of human consciousness which does not follow the mathematical rules of computation. This non-computational consciousness is that which

---


31 This is quoted by R. Penrose, in his *The Emperor’s New Mind*, p.424.
that allows us to become directly aware of something. This direct awareness plays a very important role in our mental life as we have already mentioned. Thus human understanding and conscious awareness cannot be reduced to computational processes following algorithms. There is something essential in human understanding that is not possible to simulate by any computational means.

Some philosophers believe that consciousness is a computational property, but the fact is that not even scientists, nobody, know how to design a conscious machine. McGinn interprets the concept of machine in two ways, i.e. in the narrow sense and the wide sense. The narrow sense refers to those machines, which are constructed by human beings such as motorcars, typewriters, pocket calculators, office computers, etc. In these machines, consciousness cannot be found. In the wide sense of the word ‘machine’, there are mechanical devices, which are the artefacts or the intentional products of some kinds of intelligence. In this connection, McGinn puts forward the following questions: (i) Could a human artefact be conscious? (ii) Could an artefact of any conceivable intelligence be conscious?

The first question concerns whether human beings can produce a conscious artefact with his superior technological power. It is like asking whether we shall ever travel to another galaxy. The second question raises the issue of whether the concept of an artefact is such as to eliminate the possession of consciousness. McGinn does not rule out the possibility that an artefact could be conscious. According to him, “Suppose there were an intelligence clever enough to create beings physically just like us (or bats). Then I think this intelligence would have created conscious beings. Or consider the doctrine of creationism… If we are the artefacts of God, this is not a reason to suppose ourselves unconscious. After all, there is a sense in which we are artefacts: for we are the products
of natural selection operating upon inorganic materials to generate brains capable of subserving consciousness.”

In the wider sense, the human beings are artefacts of nature and are conscious. Even then, all artefacts like tables and chairs are not conscious. Consciousness is an intrinsic property of organisms, and so in the strict sense, only organisms are conscious. That is, only living things can be conscious, and so a conscious being must be animate, organic, and alive. As Wittgenstein puts it, “…only of a living human being and what resembles (behaves like) a living human being can one say: it has sensations; it sees, is blind, hears, is deaf, is conscious or unconscious.”

There is a conceptual link between being conscious and being alive. According to this view, a conscious being either must be alive or must be like what is alive, where the similarity is between the behaviour of the things in question. In other words, only of what behaves like a living thing we can say that it is conscious. Our concept of a conscious state is the concept of a state with a certain sort of behavioural expression. We cannot really make sense of a conscious stone, because the stone does not behave like a conscious being. The point is that being biologically alive is not the same as being conscious, but it is necessary that a conscious being should behave like a living thing. Thus, instead of identifying consciousness with the material composition of the brain, we should identify it with certain higher-order properties of the brain, which manifest in conscious behaviour. For example, pain is a higher-order property of physical states,


which consists in having a certain pattern of causes and effects, and certain outward behaviour.

Now coming back to the problem of AI, it goes without saying that machines do not have consciousness. The so-called artificial intelligence does not entail consciousness. The computing machines of AI are limited in a way that human beings are not, so that it is out of the question for a conscious mind to arise merely in virtue of computation.

**VII. Argument Against CRTM**

As we have seen in chapter-II, Fodor combines both the computational theory of mind and the representational theory of mind in order to develop a new theory of mind called the computational representational theory of mind (CRTM). He has laid emphases on the computational theory of mind, as he understands mind in terms of the computer model. His claim that the mental content can be reconstructed in terms of certain formal structures shows his syntactic approach to the mind. We shall see some of the philosophical problems involved in Fodor’s thesis that supposes that mind is a system of rule-governed symbol manipulation.

Human mind is essentially a natural creative process. And the theory of mind is essentially concerned with mental states. Now the question is: How do we come to have new contentful mental states? How does mind as a system of rule-governed symbol manipulation explain the creative aspects of mental states? According to Fodor, mind consists in the application of the rule of substitution over a given set of symbols. Fodor explain this in the following passage: “one might think of cognitive theories as filling in explanation schema of, roughly, the form: having the attitude R to proposition P is contingently identical to being in computational relation C to the formula (or sequence of formulae) F. A cognitive theory, in so far as it was both true and general, would presumably explain the productivity of propositional attitudes by entailing infinitely
many substitution instances of this schema; one for each of the prepositional attitudes that the organism can entertain."\textsuperscript{34}

This is exactly what the functionalists in general are arguing for. If there are only computational processes in the internal codes, then it leaves out the essence of propositional attitudes. That is to say that there are no beliefs and desires, there are only mechanical or computational processes in the internal code. These computational processes embody the symbolic structures of mental states whose function determines the intentional content of the propositional attitudes. Underlying these computational processes are the neural states. These neural levels represent the mental states. Fodor argued, as we have already seen that mental states are nothing but brain states. Thus he argued, “the causal roles of mental states typically closely parallel the implication of structures of their propositional objects; and the predicative success of propositional-attitudes psychology routinely exploits the symmetries thus engendered…the structure of attitudes must accommodate a theory of thinking; and that it is pre-eminent constraint on the latter that it provides a mechanism for symmetry between the internal roles of thoughts and their causal roles.”\textsuperscript{35}

Fodor is concerned with an investigation into the syntactical structure of the mental states including the propositional attitudes. He has offered psychological explanations based on our folk psychology beliefs, desires, and the other propositional attitudes. However, Fodor is giving more importance to the formal features of the mental states and processes in view of his argument that the content of mental states can be exhausted by their syntactic form. On this account, therefore, mind is just a process of


\textsuperscript{35} \textit{Ibid}. 

33
rule-governed symbol manipulation and does not take into account the meaning of the formal symbols. However, mind cannot just be a manipulation of formal symbols, as it is something more than it. Searle provides a persuasive argument against the position that Fodor upholds. For him, rule-governed symbol manipulation is not a sufficient condition for mind. Take, for example, the processes involved in understanding a natural language. If understanding a natural language is just a rule-governed symbol manipulation, a computer instantiating an appropriate program must be capable of understanding a natural language.

We have seen in the last section that cognition cannot be a rule-governed manipulation of uninterpreted symbols because such a strategy fails to appreciate the distinction between syntax and semantics. Semantics is not intrinsic to the syntax, because understanding and meaning are real, independent of the syntax. Therefore, the computational model of mind is not sufficient to explain meaning of the symbols used by the mind. The human mind is more than a syntactical device. It is semantical in the sense that it has contentful states called the mental states.

The concept of ‘representation’ in the case of the computer turns out to be problematic. The computers cannot have representations the way the human mind has. The mental representations are intentional in character. As Searle rightly points out that such intentionality as computers appear to have is solely in the minds of those who program them and those who use them, those who send in the input and those who interpret the output. The point is that these symbols used by a computer are not intrinsically representational entities. That is to say, computational symbols states are not discovered within the physics; they are assigned to the physics. Though symbols are taken always as physical tokens, ‘symbol’ and ‘same symbol’ are not defined in terms of physical features. Therefore, it has the consequence that computation is not discovered in physics, but it is assigned to it. It follows that we cannot discover that the brain is intrinsically a digital computer, though we can assign a computational interpretation to it from the syntactic point of view. In view of the above, mind is intrinsically a computational mind is not a physical symbol-system with syntactic properties. It is more than a syntactic system, which needs semantic interpretation.