PHILOSOPHY OF ARTIFICIAL INTELLIGENCE

PHILOSOPHY OF ARTIFICIAL INTELLIGENCE

A Critique of the Mechanistic Theory of Mind

RAJAKISHORE NATH

Universal-Publishers
Boca Raton

Philosophy of Artificial Intelligence:
A Critique of the Mechanistic Theory of Mind

Universal-Publishers
Boca Raton, Florida • USA
2009

ISBN-10: 1-59942-905-5
ISBN-13: 978-1-59942-905-2

www.universal-publishers.com

Library of Congress Cataloging-in-Publication Data

Nath, Rajakishore, 1977-
Philosophy of artificial intelligence : a critique of the mechanistic theory of mind / Rajakishore Nath.
p. cm.
Includes bibliographical references and index.
ISBN-13: 978-1-59942-905-2 (pbk. : alk. paper)
ISBN-10: 1-59942-905-5 (pbk. : alk. paper)
1. Artificial intelligence--Philosophy. I. Title.
Q335.N376 2009
006.3--dc22

2009023423

♦

Dedicated to my
Grand Mother (*Late Lochan Nath*)
&
Grand Father (*Late Tauli Nath*)

♦

CONTENTS

ACKNOWLEDGEMENT

The book is my PhD research work. I express my heartfelt gratitude to my supervisor, Professor R.C. Pradhan, whose benign guidance has enlightened my path through the course of this work. I express my gratitude to Prof. Suresh Chandra for his continuous concern about my progress in research. My special thanks to Prof. A. K. Pujari, Department of Computer Science, University of Hyderabad, India, for his guidance in the initial stages of my research. I am thankful to other teachers in the department Prof. Amitabh Das Gupta, late Prof. Chinmoy Goswmi, Prof. S. G. Kulkarini, Prof. Raghuramaraju, Dr. P. K. Basu, Mr. Anand V. Wazalwar for their encouragement and interest in my academic pursuit.

I do also acknowledge the help and facilities that I have availed from IGM Library, University of Hyderabad, ICPR Library. I am also thankful to IIT Bombay for providing me with all the necessary facilities for executing this project. I thank my philosophy colleagues Professor P. R. Bhat, Ranjan K. Panda, Vikram Singh Sirola, C. D. Sebastian, Ratikanta Panda, and Siby K. George at IIT Bombay.

My special thanks go to Debasis Patnaik, Ramakanta Bal, Ch. Sudhakar, Saroj Kant Kar, Biswanath Jena, Samir K. Diabagh, Vineet Sahoo, Nachiketa Pradhan, and Pabitra Panigrahi for their timely help and constant inspiration.

To my parents and other members of our family, I shall ever remain indebted to them for constant source of inspiration, support, love, affection, freedom and encouragement, which has sustained my enthusiasm to carry out my work through its course. My wife, Rojalini (Pinky) has been a great source of inspiration for my academic life. I am most grateful to her.

I am thankful to Jeff Young of Universal Publishers, USA for taking up the publication of this book.

And to each and all of them, who have helped me complete this work and remain anonymous either due to preference or my ignorance.

Rajakishore Nath
June 1, 2009
IIT Bombay, India

FOREWORD

Dr. Rajakishore Nath's work entitled *Philosophy of Artificial Intelligence* is a serious philosophical study of the foundations of Artificial Intelligence (AI). The project of Artificial Intelligence has dominated the contemporary computer science and cognitive science in general. Philosophers of mind and cognitive science have been seriously involved in exploring the foundations, scope and the limits of AI because it poses many challenges to the accepted notions of mind and cognition. The present work is in continuation with the work already being done in this challenging area.

The emerging area of Artificial Intelligence covers under its scope the many-sided computer-simulated activities like information processing, natural language processing, machine vision, machine learning, robotics and so on. This area thus is well known for its study of the cognitive processes with the aid of computers and their high- powered simulation activities called "artificial intelligence". The history of AI is the history of man's creation of machines that can think, can reason out things faster than the human brains and can replicate the human brains. Such is the success of AI that it has been able to create a virtual world of images, intelligences, minds and other exotic entities. Therefore it is pertinent that philosophers study the basic presuppositions of AI and also point out the limits of such an ambitious program.

Dr. Nath has undertaken the present study to precisely bring out the limits of AI and to show that the machine intelligence cannot go far in understanding human intelligence as such. AI has many operational successes, but also it has many conceptual constraints which delimit its scope. For example, machines like computers cannot replace the human brains which are biologically programmed to execute complicated cognitive processes. Besides, machines cannot have consciousness and creativity which the human brains have. The idea of machine consciousness is a derived notion that is attributed to the artificially constructed mechanical systems like the robots. The creativity shown by the machines is an 'as if' creativity which is imposed on the machines and is therefore not *sui generis*.

Dr. Nath has argued very aptly and systematically against the very idea of a computational mind. It has been widely accepted by philosophers and cognitive scientists that the human mind operates

like a compute-like machine and that there is no difference between the operational systems in the human mind/ brain and those working in the computers. Such a notion has become so fashionable that philosophers of mind have started talking about the mechanical cognitive processes in the human brain. They have gone to the extent of saying that there is no mind and that there is only the brain and its complicated mechanical functions. Following Alan Turing, many computer scientists, mathematicians and philosophers have approached the human mind from a computational point of view espousing various forms of functionalism.

The main thesis of the present work is to expose the limits of the computational theory of mind which has been the mainstay of the philosophy of Artificial Intelligence. According to this theory, human mind is a computing machine such that all its activities can be formally programmed and that every activity of the mind can be deductively inferred from a given set of inputs. This gives us the idea that human mind is syntactically organized and that its operations can be mapped through algorithms and rules of inference. This idea has been recently stoutly refuted by many great philosophers of mind like John Searle, Hilary Putnam, Roger Penrose and many others. Their main contention is that the human mind is more than a syntactic system and is open to creative functions which can hardly be predicted. Besides, as the Chinese Room argument given by Searle shows, the human mind has the capacity of understanding meaning and language use in a communicative situation which a purely syntactic mind cannot explain. Thus it has been shown that the idea of a computational mind is a limited and a very narrowly conceived mind which must be rejected in order to grasp the true nature of the mind.

Dr. Nath has studied some of the well known arguments against the computational theory of mind and has shown that there is enough scope to investigate the nature of mind in a non-computational way. A non-computational theory will have a place for the subjective experiences and qualia of the mind. Besides, it can show why materialism, behaviourism and functionalism must fail as serious philosophies of mind. The credit of Dr. Nath's work lies in arguing for a neo-Cartesian theory of mind with non-computational creativity and consciousness. His argument against machine consciousness and machine creativity is another significant feature of his work.

I am hopeful that this work will further deepen the debate on the computational approach to mind and its philosophical implications. It will also help to clarify the serious implications of AI for the understanding of the human mind and the human self.

R.C.PRADHAN
Professor, Department of Philosophy
University of Hyderabad
Hyderabad, India

INTRODUCTION

The book makes a critical study of the philosophical problems in artificial intelligence. In mid 19th century, the hypothesis that 'machine can think' became very popular, especially after Alan Turing's article, '*Computing Machinery and Intelligence*.' This hypothesis states that machines can think intelligently just like human beings. This view emphasizes that intelligent behaviour and its interrelatedness gives rise to abstract automation, which Searle calls Strong Artificial Intelligence. That is to say, an artificial, non-biological system could be a sort of thing that could give rise to conscious experience. Similarly, Marvin Minsky held that human mind happens to be nothing but a '*Meat Machine*.' Thus artificial intelligence has emerged as a scientific programme in the twentieth century.

The object of research in artificial intelligence is to discover how a program performs the remarkable functions that make up human intelligence. The work in artificial intelligence leads not only to the inversion of increasingly useful computers but also to an enhanced understanding of human cognitive processes. The book intendeds to span the full range of interests from classical problems in the philosophy of mind and the philosophical psychology through issues in cognitive psychology, to ideas related to artificial intelligence and computer science. While primary emphasis will be placed upon theoretical, conceptual, and epistemological aspects of these problems and domains, empirical and methodological studies will appear from time to time. The broadly interdisciplinary character of cognitive science finds expression in studies that range over art, design, learning, problem-solving and analogical reasoning, which are explored from the perspective of philosophy, psychology and AI. The questions asked in cognitive science are: What do we mean by 'intelligence'? What is the mechanism used in the cognitive processes that is required to produce it? Such an investigation results in a deeper understanding of human intelligence and mind.

The first chapter discusses artificial intelligence (AI) in general. The main aim of artificial intelligence is to reproduce mentality in machines. Some artificial intelligence scientists have tried to show that human mind is itself a machine. They have argued that machines that think and act like humans are the systems that think and act rationally. This work leads not only to increasing use of computers, but

also to an enhanced understanding of human cognitive processes which constitute what we mean by 'intelligence'. What is needed is a deeper understanding of human intelligence and the human mind. The basic tenet of the computational model of mind is that the brain is just a digital computer and that the mind is a software program. The functionalist program has been strongly influenced by analogies drawn from strong-AI and computer science in general; both in its general outlook and in several of its specific applications to problems about the nature of mind. The functional state of the mind is like a computational state of a computer. A computer program can be described as a functional organization of the hardware. The functionalists argue that mental states are like the 'information processing' states of a computer. According to the computer functionalism or artificial intelligence, the brain is a computer, and the mind is a computer program implemented in the brain.

In the second chapter an attempt is made to understand the philosophical presuppositions of artificial intelligence; especially the nature of mind as presupposed by artificial intelligence. Artificial intelligence as a programme in cognitive science is based on many theoretical presuppositions like the distinction between natural and artificial intelligence, the information-processing character of mental representations, the idea of mind as a computer, and so on. The nature of cognition is a complex process which needs to be studied in order to understand the nature of artificial intelligence. On similar lines the chapter explores mental representations like beliefs and thoughts which constitute the broad domain of cognitive science. These representations explain how cognition takes place in the human mind. Cognitive science (including cognitive linguistics and cognitive psychology) has brought about a cognitive revolution in the study of mind. Here, we can undertake two important developments in cognitive science. One is the representational theory of mind, and to accept the representational theory of mind is to accept that mental representations are very much like the internal representational states of a digital computer. Two questions have to be answered in this connection: What kinds of representational systems are employed in cognition? What is machine intelligence or artificial intelligence? This gives rise to the computational representational theory of mind (CRTM in short) as this theory has answered these questions—which are the other theory. CRTM is based on two fundamental assumptions; the first is Fodor's Language of Thought

(LoT) hypothesis, and the second is the view that psychological explanation is both intentional and nomological.

Fodor's CRTM is very much related to computationalism and functionalism because CRTM is an upgraded version of both the theories. According to these theories, the mind is a device capable of performing particular sorts of operation. A state of mind resembles a computational state and is at least, to some extent, sharable in principle by any number of material systems. If we understand computation at a fairly general level, the idea that mental processing is computation, is a serious empirical hypothesis rather than a metaphor. Thus to talk about minds and mental operations is to talk about machines and their states. If there are no functional distinctions between mind and machine, the machine has the functional capacity of creativity and competence. However, program can construct hierarchical plans, ascribing them to the individual characters, according to the sorts of motivation one would expect them to have. It can give one character a role in another's plan. NET talk is one of the famous and important examples of machine creativity and competence. If this is so, we will find out how AI scientists are explaining mind from the reductionists' point of view.

Chapter three explores the limits of artificial intelligence, because a subject as intermingling and fascinating as AI could hardly escape the notice of philosophers. A critical survey will be undertaken with special reference to John Searle's, Chinese Room Experiment, Dreyfus's Challenge to AI, Gödel's Argument, and works of other philosophers like Putnam and David Charmers, and criticism of the computational representational theory of mind. They have argued for the reality of mind as distinct from a machine.

The first argument is based upon the general understanding of the 'state-of-the-art' and the 'state-of-the-attitude' in artificial intelligence. And the second argument is based on the concept of 'artificial intelligence' that seems a contradiction because the word 'intelligence' is something natural and is the quality of a conscious mind, as we understand it. Therefore, it is better to designate computers' computations as artificial information processing. The brain in its intrinsic nature has the capacity of intentional operation. Man is a conscious being, who cannot be reduced to patterns of any matter; man is qualitatively different from the rest of the nature. If we reduce it to matter, we lose its value as a human being.

Chapter four establishes a non-computational view of mind. As opposed to the non-computational view, the computational view of

mind says that mind is a machine. The mind, asserting to this view, is nothing but the brain consisting of computational processes, that go on inside the brain. In this chapter, we shall examine the Cartesian view of mind because Descartes has given a systematic and scientific explanation of the existence of mind and its nature. In the Cartesian nature of mind, there is no place for computationality in the human brain as the thought act is due to the subjective thinking thing, which is the self. The existence of thinking thing is same as the existence of a subjective thinking thing, because it is the subject, who thinks. This shows that all these subjective activities are non-computational because subjective activity presupposes semanticity. It opposes the idea of mind as computational because mind is beyond the computational processes. The above argument is similar to Descartes' view, regarding the non-computationality of the human mind. Based on the above argument, we can examine Chalmers' conception of mind which endorses a non-computational view of mind. Chalmers criticizes the materialistic theory of mind and establishes dualism between conscious mind and the non-conscious matter. He calls this dualism as 'naturalistic dualism' because it is based on the premise that matter and consciousness are ultimately explainable by appealing to natural laws. Following Descartes and Chalmers, in the book we shall argue that the phenomenon of consciousness can arise only in the presence of some non-computational physical process taking place in the brain. That is to say that the non-computational property emerges from physical properties. In the last two sections, we would examine arguments in favours of non-computational theory of mind and the consequences of holding a non-computational theory of mind. Since living human brains are ultimately composed of the same material, satisfying the same physical laws, as are the inanimate objects of the universe. We must presume that such non-computational processes would also have to be inherent in the action of inanimate matter.

Chapter five of the book deals with the problem of creativity and consciousness. Creativity and consciousness are some of the puzzling features of the human mind. Both the concepts 'creativity' and 'consciousness' are logically linked because a conscious human being alone has the power of creativity. However, recently there has been a reappearance of interest in this area, principally in artificial intelligence and cognitive science. The book intends to explore the features of creativity, and how creativity is related to different cognitive faculties of the human mind. The second section explores the di-

mensions of creativity, especially the psychological dimensions of creativity, because creativity is also related to human psychology. The third section critically examines the question: Are there creative machines? Creativity is related to skills and abilities and also to ideas which are new and original. The ability to generate ideas and beliefs effectively *ex-nihilo* is the core of creativity. The most common reason put forward to support the claim that computers cannot originate anything is that they merely follow instructions. The fourth section deals with consciousness and creativity. Philosophers have treated consciousness as a mystery for a long time. In recent years, researchers from diverse subjects like psychology, neuroscience, computer science, physics, etc. are showing interest in the subject and are coming forward to share their findings with others. Consciousness is very much related to the creative activities because a human being cannot be creative without consciousness. This does not mean that a man who is conscious necessarily has the capacity of creativity. The fifth section deals with the concept of machine-consciousness. This section will raise the question: Is there a creative machine? The usual answer is 'No' because machine-consciousness is derivative in comparison to the human consciousness. We are so prone to count the robots of science-fiction films as conscious beings; they do not live, but act as if they do. We cannot find a conscious stone because the stone does not behave in ways we can recognize as expressive of its supposed consciousness. Being biologically alive is not a necessary condition of consciousness, but it is necessary that a conscious being should behave like what it is like to be a conscious being. In the last section, you will find that artificial intelligence fails as theory of mind because the way AI scientists explain mind is very much mechanical and deterministic. If we accept mind as a machine, then we are leaving unexplained the essence of human mind, i.e., the subjective qualitative experience. The mechanistic and scientific explanations of mind leave out the explanatory gap between mind and body, problem of subjectivity, consciousness, qualia, etc. The book thus aims at exposing the limits of artificial intelligence as a theoretical and scientific programme.

CHAPTER I

THE IDEA OF ARTIFICIAL INTELLIGENCE (AI)

The object of research in artificial intelligence (AI) is to discover how to program a computer to perform the remarkable functions that make up human intelligence. This work leads not only to increasing use of computers, but also to an enhanced understanding of human cognitive processes that constitutes 'intelligence', and the mechanisms that are required to produce it. What is needed is a deeper understanding of human intelligence and the human mind. In the first section, we will focus on the various definitions of artificial intelligence and organize it into the four categories—systems that think like humans, systems that act like humans, systems that think rationally, and systems that act rationally. In the second and third sections, we will explore the field of AI, and the issue of what computers can do, respectively. In the fourth section, we will argue for the computational model of mind. The basic tenet of this thesis is that the brain is just a digital computer and that the mind is a software program. In the last section, we will focus on the relation between AI and the functional theory of mind.

I. What is Artificial Intelligence?

It is difficult to give a precise definition of artificial intelligence. Some recent times artificial intelligence scientists have attempted to define artificial intelligence (in short AI) in various ways.

According to Haugeland, artificial intelligence is, "the exciting new effort to make computers think … machines with minds, in the full and literal sense"[1] For Bellman, it is "the automation of activities that we associates with human thinking, activities such as decision making, problem solving, learning…"[2] Charniak and McDermett define AI as "the study of mental faculties through the use of computational model."[3] And for Winston, it is "the study of the computations that make it possible to perceive, reason and act."[4] AI, for

Kurzweil is "the art of creating machines that perform functions that require intelligence when performed by people."[5] Rich and Knight say that AI is "The study of how to make computers thinks at which, at the movement, people are better."[6] For Schalkoff, AI is "a field of study that seeks to explain and emulate intelligent behaviour in terms of computational process."[7] Luger and Stubblefield hold it to be "the branch of computer science that is concerned with the automation of intelligent behaviour."[8]

Let us look at all the definitions from different angles. Haugeland and Bellman point out that artificial intelligence is concerned with thought process and reasoning. They have explained the mind as a machine that is completely associated with human thinking. That is to say, computers do think. But Schalkoff, Luger and Stubblefield are concerned with the behavioural aspects of systems. For them, computers behave as intelligently as human beings. Moreover, Kurzweil, Rich and Knight are concerned with measuring success in terms of human performance. For them, artificial intelligence can be attributed to machines, but it belongs basically to the human mind. Finally, Charniak, McDermett, and Winston are concerned with an ideal intelligence. They explain the mental faculties through the use of computational models.

To sum up, all the definitions of AI can be organized into four categories.[9] They are as follows:

(i) Systems that think like humans.
(ii) Systems that act like humans.
(iii) Systems that think rationally.
(iv) Systems that act rationally.

Now, we have to look at each aspect in detail.

(i) Acting Humanly: Turing Machine Approach

The Turing test, as named after Alan Turing, was designed to provide a satisfactory operational definition of intelligence. Turing defined intelligent behaviour as the ability to achieve human-level performance in all cognitive tasks to fool an interrogator.[10] In his '*Computing Machinery and Intelligence*', Turing says the new form of the problem can be described in terms of a game which we call the 'imitation game.' It is played by a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the interroga-

tor is to determine which of the other two the man and which the woman. He or She knows them by labels X and Y, and at the end of the game, he or she says, either 'X is A and Y is B' or 'X is B and Y is A.' The interrogator is allowed to put questions to A and B.[11]

Thus C: will X please tell me the length of his or her hair?

Now suppose X is actually A, then A must answer to the question. It is A's object in the game to try to cause C to make the wrong identification. His or her answer might, therefore, be 'my hair is singled, and the longest strands are about nine inches long.' However, because the tones of voice may not help the interrogator, the answer should be written or better still be typewritten. The ideal arrangement is to have a tele-printer for perfect communication. Alternatively, an intermediary can repeat the questions and answers. The object of the game for the second player (B) is to help the interrogator. The best strategy for her is probably to give truthful answers. She can add to her answer such things as 'I am the woman, do not listen to him,' but it is of no avail as the man can make similar remark.

Now, we can ask the question, what will happen when a machine takes the part of A in this game? Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between man and a woman? Turing's answers to these questions are more or less summed up in the following passage; "I believe that in about fifty years time it will be possible to program computers, with a storage capacity of about 10^9, to make them play the imitation game so well that an average interrogator will not have more than 70 percent chance of making the right identification after five minutes of questioning." [12]

What Turing had predicted then is a fact now—the machine or the computer can imitate human behaviour. It should be pointed out that Turing's beliefs about the capabilities and capacities of machines are not limited to such activities as playing the imitation game as successfully as human beings. Roughly speaking, the test Turing had proposed was that the computer should be interrogated in the place of human beings. Turing's test deliberately avoided direct physical interaction between the interrogator and the computer, because physical limitation of a person is unnecessary for intelligence. However, the so-called Turing test includes a-video signal so that the interrogator can test the subject's perceptual abilities. In order to pass through total Turing test, the computer will need computer vision to perceive objects and robotics to move them.

Again, the issue of acting like a human comes up primarily when artificial intelligence programs have to interact with people. Then the expert system explains—how it came to its diagnosis, or a natural language processing system has a dialogue with a user. These programs must behave according to certain normal covertness of human interaction in order to make them understood. The Turing test shows that machines can interact with human beings the way human beings interact amongst themselves. Which in a nutshell means machines can behave the way the human beings do.

(ii) Thinking Humanly: The Cognitive Modelling Approach

The interdisciplinary field of cognitive science brings together computer models from Artificial Intelligence and experimental techniques from cognitive psychology to try to construct precise and testable theories of the workings of the human mind. And if we say that a given program thinks like a human being, we must have some way of determining how human beings think. For that, we need to get inside the actual workings of the human mind. Stuart Russell and Peter Norvig say that there are two ways to do this: through introspection—trying to catch our own thoughts as they go by—or through psychological experiments. Once we have a sufficiently precise theory of the mind, it becomes possible to express the theory as a computer program. If the program's input/output and timing behavior matches human behaviour, that is evidence that some of the program's mechanisms may also be operating in humans.[13]

Now it is almost taken for granted by many psychologists that "a cognitive theory should be like a computer program."[14] But we know that cognitive science is the science of mind. Therefore, cognitive scientists seek to understand perceiving, thinking, remembering, understanding language, learning, and other mental phenomena. Their research is remarkably diverse, ranging from observing children's mental operation, through programming computers to do complex problem solving, to analyzing the nature of meaning. In order to appreciate the work in artificial intelligence, which is a necessary part of cognitive science, it is necessary to have some familiarity with theories of human intelligence. The cognitive scientists introduce the notion of machine intelligence and emphasize the relationship between human and machine intelligence. The aim of artificial intelligence is to develop and test computer programs that exhibit characteristic human intelligence. The most fundamental contribution of symbolic computational modelling has been the physical symbol system hy-

pothesis. According to Newell and Simon, "a physical symbol system has the necessary and sufficient means for general intelligent action."[15] By 'necessary' we mean that any system that exhibits general intelligence will prove upon analysis to be a physical symbol system. By 'sufficient' we mean that any physical symbol system of sufficient size can be organized further to exhibit general intelligence. Lastly, by 'general intelligent action' we wish to indicate the same scope of intelligence as we see in human action; that in any real situation behaviour appropriate to the events of the system and adaptive to the demands of the environment can occur within some limits.

However, the ability of computer simulations to model such process is interpreted as a proof of the broader claim that a symbol system is at the centre of human intelligence. In this hypothesis, it shows that intelligence is an essential aspect of machines. If the machines have the capacity of intelligence, intelligence is the essence of human cognitions. Therefore, machines have cognitive capacity like the human beings. In the cognitive modelling approach, thus, human beings and machines show the property of being intelligent.

(iii) Thinking Rationally: The Laws of Thought Approach

'Right thinking' is the inferential character of every reasoning process. Aristotle in his famous syllogisms provided patterns of argument structures that always give correct conclusions from given correct premises. In the syllogisms, the Laws of Thought play a vital role because these laws give the right explanation of a syllogistic inference. There are three Laws of Thought recognized by the logicians. These have traditionally been called the law of Identity, the law of Contradiction, and the law of Excluded-Middle.[16] These Laws of Thought are appropriate to different contexts. The formulations appropriate as follows:

a) The law of Identity asserts that if any statement is true, then it is true. This law asserts that every statement of the form $P \supset P$ is true, and that every such statement is a tautology.

b) The law of Contradiction asserts that no statement can be both true and false. This law asserts that every statement of the form $P \,.\, \sim P$ is false, that is, every such statement is self-contradictory, and its negation is logically true.

c) The law of Excluded-Middle asserts that any statement is either true or false. This law asserts that every statement of the form $P \vee \sim P$ is true, that is, every such statement is a tautology.