Artificial Intelligence: Knowledge Representation and Learning

Prof. Deepak Khemani

Department of Computer Science and Engineering

Indian Institute of Technology, Madras

Module – 01

Lecture - 01

Introduction

Welcome to this course on **Knowledge representation and reasoning**. This is a course which is a kind of a complementary course to another course which I teach and some of you have done, which is on artificial intelligence. In artificial intelligence we focus more on problem solving methods and things like search and so on. And in this course we focus more on representation issues so, knowledge representation, and I feel that knowledge representation is the core of an intelligent agent essentially.

We will see during the introductory lectures that if you want to think of building an intelligent agent, the first thing that you have to worry about is that the agent, you will be able to represent the environment that it is in, the domain that it is in, the world around itself, it should be able to represent its goals and it should be able to reason about its own plans and you know all that is necessary. And at the core of all this is knowledge representation.

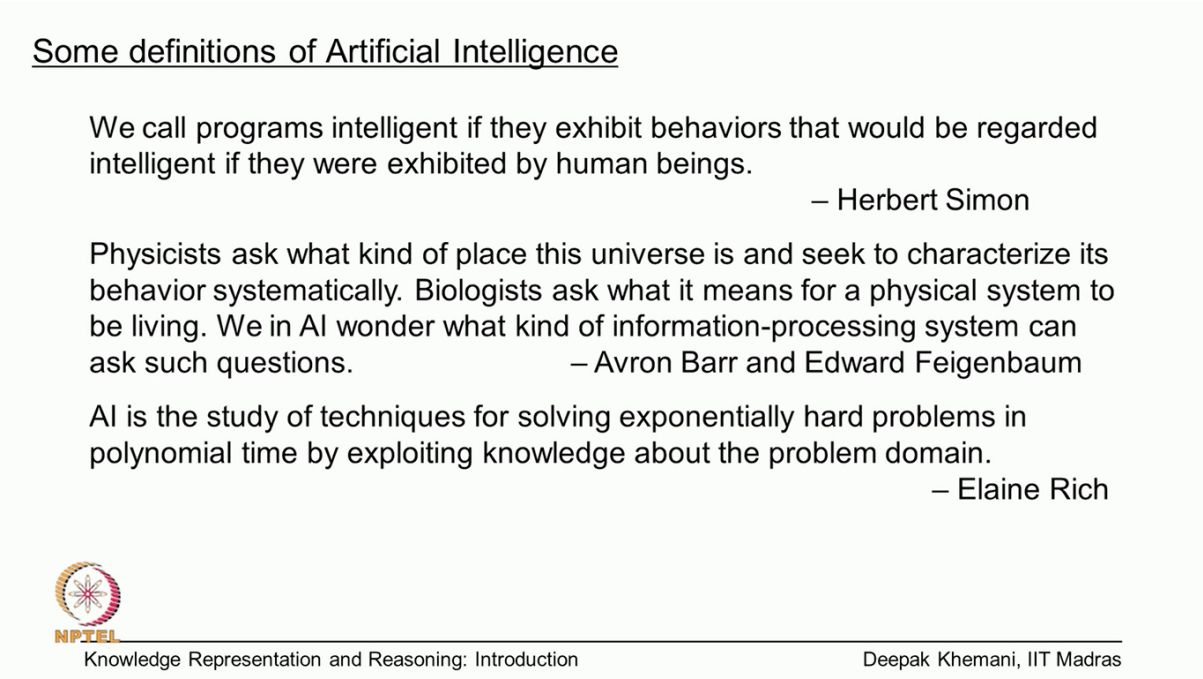
We will be recording these lectures is half an hour modules and initially for the introduction part I will mostly use slides then after that I will have a normal writing session on classroom and in between in one or two lectures where I have lot of text to be shown, I will again revert to slides session. So, let's just start with the beginning of, this is something which I also did in the course on AI but since knowledge representation is the core component of AI, we will just kind of revise this, and also it will apply for people who have not seen those lectures or have not done that course.

So we will start with some basic definitions of AI, one of the oldest is by Herbert Simon, is that we call programs intelligent if they exhibit behaviours that would be regarded intelligent

if they were exhibited by human beings. So this is the most standard way of thinking about AI is that can you make programs do what people do and of course underlying is that assumption is that what people do is an intelligent activity. And that's one of the oldest definition. Simon was one of the founders of AI. Feigenbaum also is one of the old timers. He said that Physicists ask questions about this universe and seek to characterize its behaviours systematically. Biologists ask what it means for a physical system to be living.

We in AI wonder what kind of information-processing system can ask such questions. We are interested in building these systems which will be able to look around the world and ask these kind of questions and so on. Elaine Rich who wrote a book which was very popular in the mid-80s I think, she wrote and then it was augmented with Rich and Knight and then subsequently as Rich, Knight and Nair. He gave a more computational definition to AI; It says that AI is a study of techniques for solving exponentially hard problems.

(Refer Slide Time: 3:34)



Now those of who have good at substructural complexity would know that you can characterize problems according to the difficulty of how you solve them; for example SAT is exponentially hard; at least 3SAT onwards they are exponentially hard; TSP is even harder and so on.Now if you want to solve those problems; obviously the problem itself is by definition hard so you can't hope to solve it optimally in any reasonable amount of time unless of course it is an exceptional case or it's a small problem in which case the

exponentiation doesn't show off. We are interested in to find methods which in polynomial time will give us reasonably good solution. So for example when you are solving TSP you are not looking for the optimal solution but the solution which is reasonably close to optimal; and there are other communities which look at for these kinds of goals for example approximate algorithms and things like that but AI essentially focuses on the use of heuristic methods and knowledge and things like that. Charniak and McDermott, two authors who have another very popular book and I have used it quite often myself.

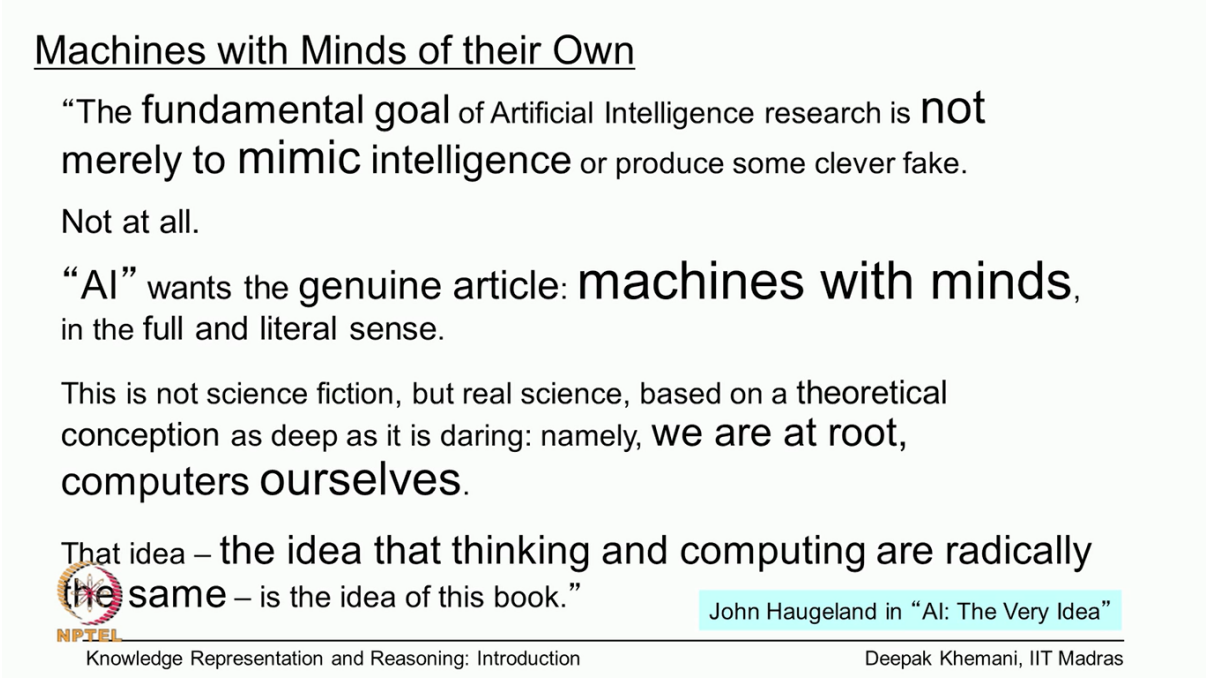
They say that AI is the study of mental faculties, it's a study of how humans think so it's more like cognitive science attitude towards AI as to how do we think, what are our mental faculties and the study is done by building computational models which will validate our theories essentially. So if we say this is how we remember things, this is how we store our memory, this is how we solve problems. Then if you can show a program that does something similar; then it is a kind of a validation.

So by and large AI has had interests from two kind of communities; one is the people who want to study intelligence or intelligent behaviour or what is intelligence; the cognitive science kind of; cognitive psychology kind of outlook and the others who want to build useful systems so you want to build a program which will control your robot on mars planet so for example NASA had the robots landed on Mars and a lot of AI went into that thing because mars is the place which is some 30 odd minutes away in terms of light distance. So you can't press a button here and immediately get a reaction from the vehicle; it takes some 30 odd minutes or whatever the actual time is for the signal to go there. So obviously you need to have to build autonomous systems; you need build in some of the techniques that we are talking about.

The definition which I like most is due to this guy called John Haugeland. He is philosopher by profession and maybe that's why he gives this definition. He says the fundamental goal of Artificial Intelligence is to not mimic intelligence or not mimic human intelligence but to produce some; not to produce some clever fake but it wants the genuine article; it wants machines with minds, in the full and literal sense of the world. So he said that I don't want to see that my machine can play smart chess or it can talk to you in natural language, or it can analyse data and extract patterns. I am not interested in the outcome, I want to try to create machines which you would accept have minds of their own. That of course is a very hard question to this thing but we we will sort of circumvent the notion of definition of minds and

so on here. And then he goes on to say that at the heart of this is the conception as deep as daring is it, we are at root computers ourselves. So that of course addresses the question of can machine be intelligent and if you are also a machine then of course you can be intelligent. So the idea that thinking and computing are radically the same is the idea behind his book. His book titled " AI: The Very Idea". It's a philosophical outlook to AI, it is not the technical book, it will not give you algorithms; the kind of algorithms, the we will be studying, but tries to answer question like these.

(Refer Slide Time: 7:22)



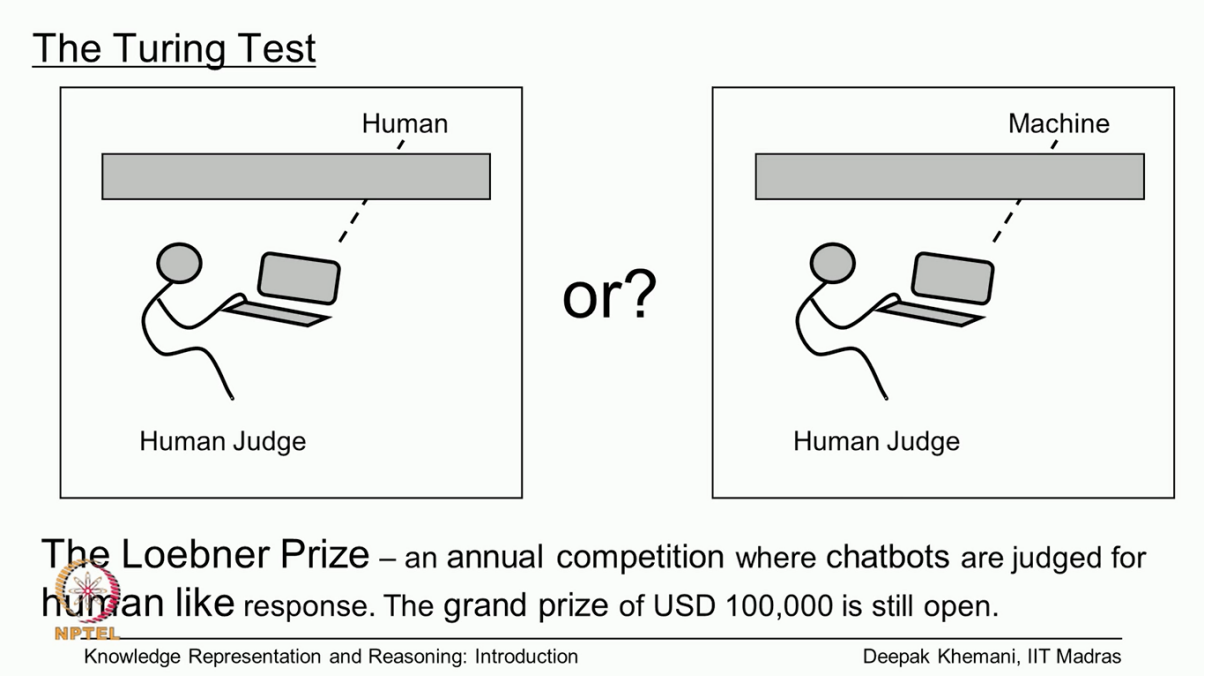
There was a big debate about can machines think in the 1950s and so on and Alan Turing that who you all know post this test which we call as the Turing test but which he called as the Imitation game. Now you must have heard about the movie which have just out; in fact the movie is called as the imitation game and the movie is really about Alan Turing; it's a kind of biopic about Alan Turing. But Alan Turing himself called this test an imitation game which is a kind of consistent with Simon's definition of intelligence that if you can exhibit behaviour which human being would exhibit then you would accept them to be intelligent, except that in turing's test the behaviour was exhibited by means of the text data; so you type in a question and system types in an answer so it is like a chat box you interact with.

Now in computing prediction is very hard; at all times people have predicted all kinds of things about this will happen in so many years; this will happen in so many years and they

have often proved to be wrong especially when it comes to AI kind of prediction for example Turing said that in about fifty years it will be possible to program computers with a storage capacity of about 10 to the power 9. Now he is talking in 1950s, in 1970s also you got machines with 64 kB RAM and the hard disk was 20MB or something like that. He is talking in those times that to make them play the imitation game so well that an average interrogator will not have a reasonably good chance of making out whether it is talking to a man or a machine.

And his game is described in his paper called the Computing machinery and intelligence; and If you just search for it you will get it online quite easily. So this is a turing test. There is a human judge sitting on a teletype or chatbox or now a days may be on a smartphone or something and chatting with someone; only thing is the judge doesn't know whether he or she is chatting with a human being or with a machine and the turing test says that if the machine can fool the judge reasonably often then it has passed the turing test and periodically we get claims; even in 2014 we had a claim from Warwick university saying that the turing test has been passed and so on and so forth.

(Refer Slide time 9:51)



There is a prize called the Loebner Prize which is an annual competition in which chatbots participate. If you just lookup the Loebner Prize you will probably get to see the chat sessions for 2014 and you will find some of them to be quite impressive. I have not copied them here but I will leave it to you to look at them; there is a money to be won if you win the prize specially.

You must have heard about the program called ELIZA which was one of the earliest chatbots which was developed by Joseph Weizenbaum of MIT. ELIZA is named after Eliza Doolittle who was a working class character in George Bernard's play Pygmalion and so on. ELIZA had some very simple rules to manipulate language. It would take a sentence, manipulate it and throw it back to you. It had no notion of understanding, it had no representation about the world and things like that; It was just a simple, you might say twister of sentences. It could take a sentence and convert into a question and do that kind of things without really going into the meaning of words.

And there was a popular version called Doctor which was running; trying to behave like a psychotherapist and there is a small story about how a Russian scientist who was visiting Stanford interacted with the program and this is the text from that interaction; So I will ask you to read it yourself. If you can make out the color, there is a bit of purple in there; that is those bits of sentences which indicate that how that program is really twisting the sentence. So the visitor says I am a bit tired, that's all and the program which is a doctor says why do you think you're feeling a bit tired. You see it has just taken the phrase from there and added

why do you and it sounds as if it is a therapist because human beings are very gullible; they tend to believe all kinds of things.

But Weizenbaum was so disturbed by people's responses; so for example his secretary would confide all her life problems to the program and she was aghast to know that Weizenbaum could read those responses. He eventually wrote a book which says that we should limit the powers of the computer. Much earlier people had been trying to build real mechanisms to do reasoning so our focus is on knowledge representation and reasoning; so for example if you add two numbers you are doing some kind of reasoning. It I say 37 plus 12 is 49 then somehow I have come to the conclusion that it is 49. Now if you want machines to do this sort of things then arithmetic was one of the first thing that people tried to tackle. You have to first somehow represent those numbers and then you have to devise algorithms which will do addition.

Much before computers came people were interested in this sort of a thing and there is a lot of folklore behind this; so there is a nice book by Pamela McCorduck called Machines Who Think; and I would advise you to read it if you are interested in the history of AI and things like that; so these Arab astrologers were credited with constructing what is called as a Thinking machine called the Zairja and the principle behind that was to generate ideas by mechanical means with the help of a technique called breaking down; from which the word the algebra was derived and by combining number values associated with the letters and categories, new paths of insights and thought were created.

The Europeans were suitably impressed and Spanish Catalonian missionary called Ramon Lull; he decided to build his own device which he somewhat largely called Ars Magna and his goal which you can read in italics was to bring reason to bear to all subjects; by subjects we mean people around us and in this way, arrive at truth without the trouble of thinking or fact finding. If you had a machine which could do this for you then why do you want to take the trouble of thinking. Of course now we have google who does all these things. As I said some of the earliest things that were happening in arithmetic and some of the names you see are actually the well-known names from the world of science.

So we start with Pascal after whom the language pascal is also named amongst other things. He invented a mechanical calculator using some set of gears ; the system of gears in 1642 remember much before you were born. He went through 50 prototypes and made this machine which is called pascaline. It could add and subtract two numbers directly and

multiply and divide them by repetition. This is a kind of picture of pascaline which you can find from the web. So I have given all the sources of where the picture I have taken from; most of the stuff are from Wikipedia. He was given right, exclusive right to make and sell those machines but he managed to sell only about 20 of them so it was becoming bad of business proposition for him. So his startup in some sense did not work.

Leibniz you all know; in many areas we have heard easily. He worked on his calculator after pascal and he devised what he called as the stepped drum. We will see a picture which was used for three centuries. So you can see that there is a drum which is rotating and it is driving; it is rotating a red coloured rod which is called gear fitted on it and if you think carefully you will see that there those projections on the drum which are of different length so depending on where the gear is placed, in one rotation of the drum the gear will rotate after a certain amount of time. So we have some way of measuring a number you might say, some mechanism for representing a number. So he build this device and it could do division by repeated subtraction with 8 digit numbers.

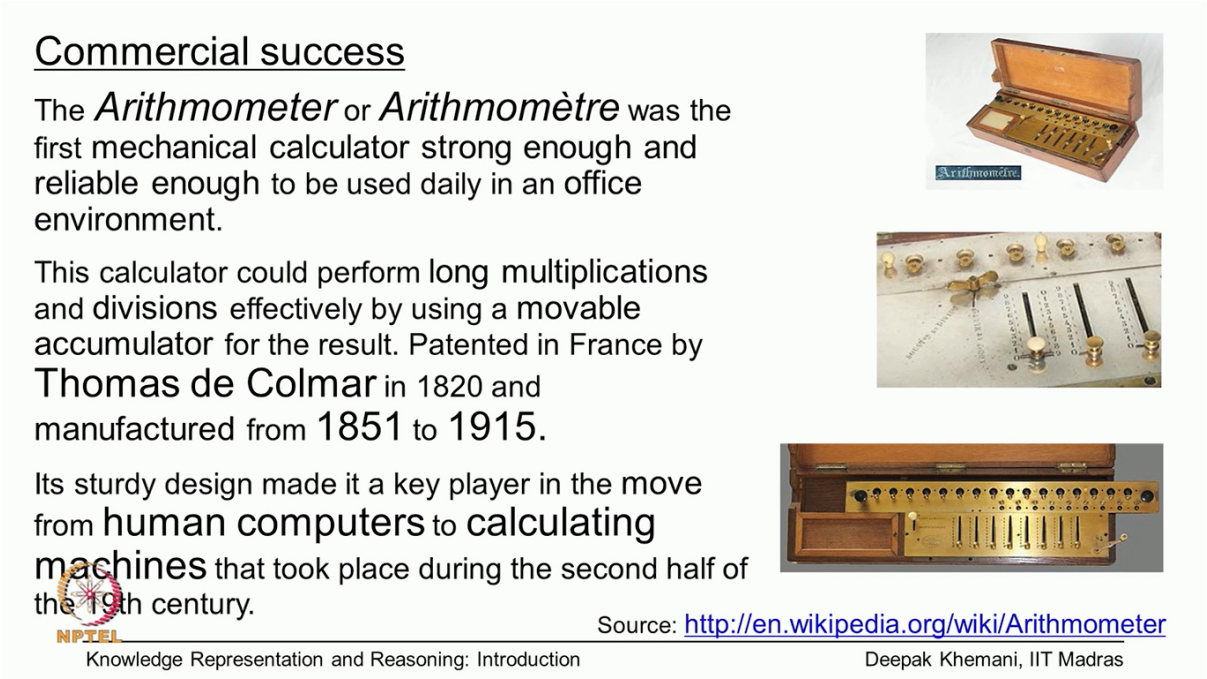
Some of the earliest machines that were built like computers were also of 8 bit machines. Leibniz was as we just said he was not just a mechanical engineer; he was a philosopher by which we mean that he wanted to understand everything in the world. And he said that human reasoning could be reduced to calculations of a sort and he says that the only way to rectify our reasonings is to make them as tangible as those of the Mathematicians.

Mathematicians were known to be precise and he said that the reasoning can also be precise. And so we can find the error at a glance, and when there is dispute among people, we can simply say: Let us calculate and resolve the dispute instead of taking out sword and fighting and punching; all kinds of uncivil methods of resolving the disputes, we can just sit down. So there are two principles of his logic; that our ideas are compounded from very small number of simple ideas which is something that we are now very familiar with, and complex ideas proceed from these simple ideas by uniform and symmetrical combination, analogous to arithmetic multiplication.

So we can understand this uniform and symmetric combination saying that some kind of algorithm; that if you have a basic alphabet; if you have a basic set of atomic units then by applying algorithms you can construct more complex things out of it. There was another device called Arithmometer built by Thomas de Colmar which was a really sophisticated device and it was manufactured as you can see in the middle paragraph till about 1915 which

is just about a 100 years ago after which of course electronic machines started coming to the market.

(Refer Slide time 18:46)



Calculating machines that took place during the second half of the 19th century. So we are interested in representation and reasoning so let's see what Ada Lovelace has to say. Ada Lovelace was a collaborator of Charles Babbage who was credited by many to be the first; the designer of the first computer. A computer is different from a calculator in the sense that it has the stored program. You can change the program and it behaves differently. Calculator can do only one thing and computer can do many things.

Ada Lovelace was somebody who worked with him and she could see even at that time that mechanical; the potential of Babbage's mechanical computer extended far beyond just number crunching. She writes about The Analytical Engine; this is one of the two machines that he designed; he never build The Analytical Engine, but he had a good design and much later people built it after him. She says that it might act upon other things besides number, were objects found whose mutual fundamental relations could be explained by those of who can relate; find relations between the elements; then abstract science of operations, and which should be also susceptible of adaptations to the actions of the operating; basically you can

represent things in some way, represent relations in some way then you can apply the algorithms on top of that and you will be able to do interesting stuffs.

As an example she uses the composition of music. She says that suppose, for instance, that a fundamental relations of pitched sounds, basically notes of music system in the science of harmony and of musical composition were susceptible to such expression; and by expression we should read representation and adaptations; and by adaptations we should read reasoning, the engine might compose elaborate and scientific. I am sure that some of you must be familiar that there are programs which now compose music; infact last year or the year before last, The Guardian ran a kind of survey in which they put up 6 pieces of music out of which I think one or two were composed by machines and 4 were composed by a human composer and they asked the readers to judge which one is composed by the machine. A kind of turing test you might say for music and interestingly the people could not figure out which one is composed by the machine. If you search the guardian site you might actually find it; the results.

Ok so let us now move towards representation. In medieval Europe, the idea about the world; so when people think about the world you know they are actually modelling the world essentially. It was based on Christian adaptation of greek ideas. So platonic notion of thing was that there is a creator who has got ideas, which are perfect ideas; you have heard about the thing like platonic love for example which comes from the same sources. Like perfect which is the creator's idea from which we derive human ideas and from which the world also is made basically. So that world is made up of corruptible materializations of God's idea.

If we have a chair which is perfect then it resembles God ideas and if it is broken then it doesn't resemble God ideas. So our thoughts are true; so our interest is in representation and reasoning, how can we talk about things that we represent and reason about though are true to the extent that they are accurate copies of God's ideas. So that was Plato; his disciple and successor Aristotle did away with the notion of a creator. He says that human ideas basically resemble the objects they stand for. This is kind of known as correspondence theory of truth.

If you are seeing something and thinking about that then the fact that you are thinking about that then is related to the fact truth value of the thing. And much later Wittgenstein the philosopher had created what he called as the picture theory of language in which he says that you create memories or thoughts which are kind of pictures of things that you see essentially. So let's look at what else the Greeks gave us , they gave us a thing called Syllogism which I

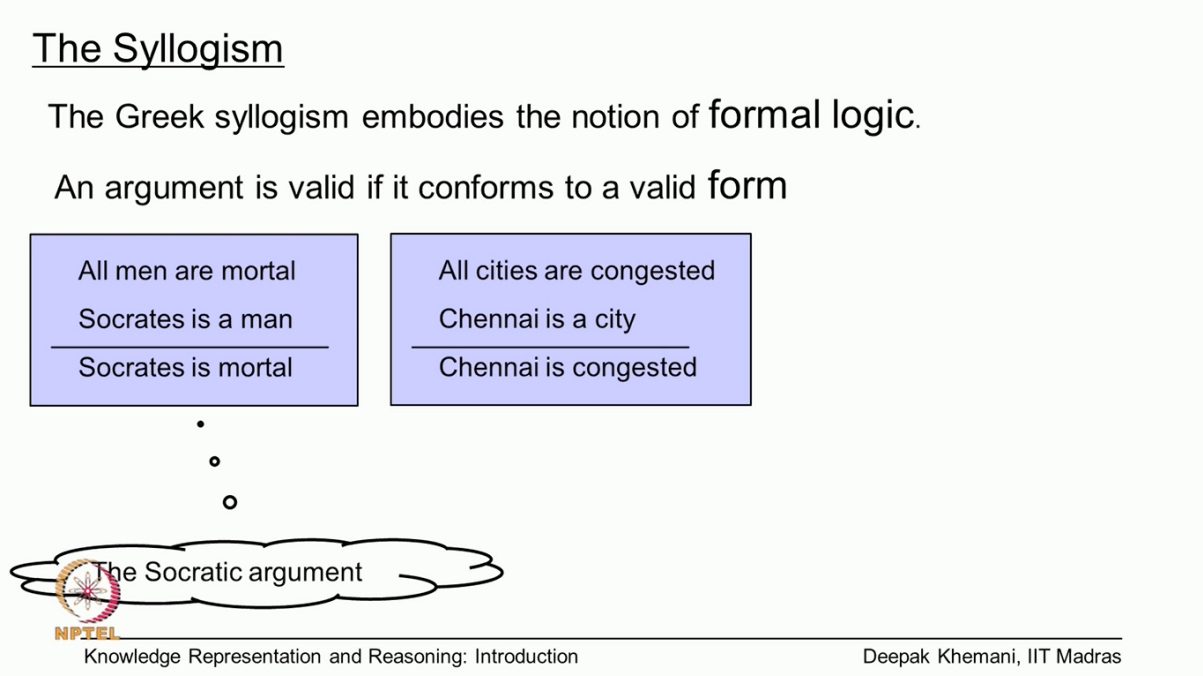
am sure you must have read about. So the greek syllogism embodies a notion of formal logic, this is one of the things we are going to be interested in.

Our base for formal representation is going to be formal logic and the interesting thing about the formal logic is its formal; it depends only upon form and doesn't depend upon content essentially. And there is a notion of an argument. We say that an argument is valid; we will discuss these thing in more detail when we actually look at logic. An argument is valid if it conforms to a valid form. It's a form which is important not the content. It is not about you are saying what is important, it's that, it you say 2 or 3 things which are the premises and then the conclusion is true if it follows the certain acceptable forms .

Now the greeks gave some 19 different syllogisms out of which one we are mostly familiar with which is also known as the Socratic argument and you would have surely heard this. If you have given the first two sentences above the line is premises and below the line is the conclusion; you say that All men are mortal and then you say Socrates is a man then you are allowed to conclude that Socrates is mortal. And this argument is actually known as a Socratic argument.

It talked about Socrates who you might are forced to drink poison because of his beliefs. That's why they are talking about his mortality. But if you look at this argument which is about Chennai; it could be about any other city. If we accept the fact that all cities are congested and we say Chennai is a city then we can conclude that Chennai is congested.

(Refer Slide Time 25:02)



So both these arguments have the same form; All xs are ys, some z is an x therefore z is a y. As long as we believe in the first two, we are allowed to believe in this conclusion and that is a valid form. So this form is a valid form; it is one of the 19 valid forms that the greeks has given us; we will not look at those, we will look at more modern logics and rules that we use here.

So one last example; it you believe that All politicians are honest; if you believe that Sambit is a politician then we have no choice but to believe that Sambit is honest because all three arguments have the same form. So in a valid argument, if the premises are true then the conclusions are necessarily true. So if the first two sentences are; if you accept the first two sentences then you must accept the third sentence.

So more on representation; here we have a picture of Galileo which I have also taken from wikipedia; there is the picture of Galileo with his telescope; very well known for his telescope but what we are thinking about here is that he is talking about taste, odour, colours and such things which are related to perception. You know we smell flowers, we taste food, we see colours; he says that tastes, odours, colours, and so on are no more than mere names as far as the object in which we identify them so if I see you wearing a red T shirt, then the Red is a name, it is a concept which I have and they reside in our consciousness so they are represented in our consciousness. Hence if the living creature were removed all these qualities would be wiped away.

The fact that I see your shirt as red is something to do with me essentially. I mean your shirt may or may not be red. Of course you can argue that it reflects a certain fraction of the spectrum of light and that particular combination of fraction may be called as red; and that kind of definition you can try to give but our perception of red or smell or taste is in our mind essentially and it is something important to us for the representation point of view.

So he says philosophy; basically talking about the world around us is written in this grand book, the universe and it is written in the language of mathematics. See everybody is obsessed with mathematics because it is precise, you can rely on it essentially and it's characters are triangles, circles and other geometric figures so Galileo, we know him for his laws of motion so we have equations like v equal u plus at and so on and so forth but the notion of variables didn't exist when Galileo was thinking about them. He thought about motion in terms of geometry.

So if you see there is a one of the equation is ½ at square or something like that. You can think of half at square as the area of triangle in which one side is *at*; t is time and a is acceleration and other side is time essentially. So if you can visualize the triangle in which one side is time and other side is time in to acceleration then you can sort of compute something. So he used to think in terms of geometry; so he says that motion could be represented in geometry. So he showed that geometry could be used to represent the reasons of a motion. This is something very interesting for us; we want to think about motion but we can use a representation which is basically a geometry.

Thomas Hobbes which many people call as the grandfather of AI; he extended this idea; he was a political scientist and philosopher; so he says that thinking is the manipulation of symbols. Galileo has said that all reality is mathematical in the sense that everything is made up of particles, and our sense of smell or taste was how we react to those particles. So when we react; when we sense or smell we are reacting with some particles which are emanating from the food or whatever that we are smelling. Hobbes extended this notion to say that thought too was made up of particles which the thinker manipulated essentially.

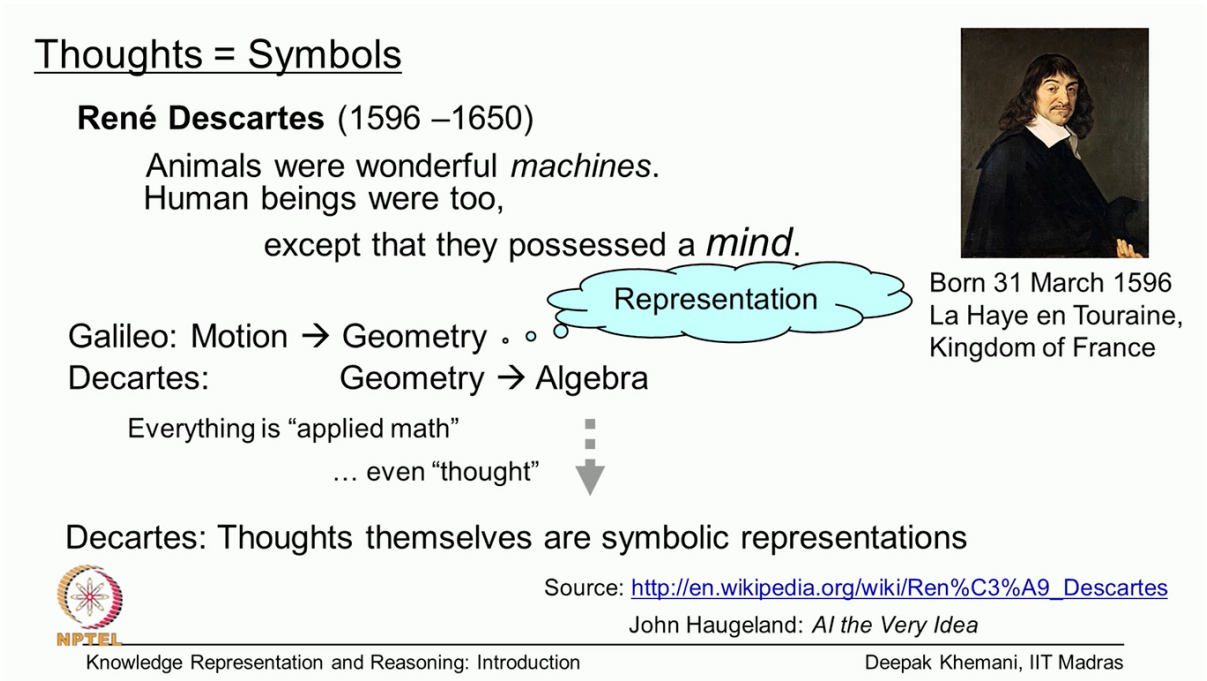
So you see we are heading towards the idea of representation; however he had no answer as to how can a symbol mean anything? So if I write a word c80, it means something to you. How does the meaning arise? That's the question I feel is still difficult for us to evolve. Remember Hobbes was saying that thinking is equal to computation. Somewhere he says that by reasoning he says I understand computation and to compute is to collect the sum of many

things added together at the same time or to know the remainder; basically some kind of mathematical algorithm; arithmetic algorithm; he says that thinking is also like that; you create particles and you reason about them. This I have taken from Stanford Encyclopedia.

Hobbes was influenced by Galileo. Just as geometry could represent motion, thinking could be done by manipulation of mental symbols. So for those of you who know Calvin and Hobbes; the character Hobbes was named after Thomas Hobbes which will increase our respect for him. Let's also talk about Rene Descartes; we know Descartes from different angles, Cartesian coordinates and other things. He said that the Animals were wonderful machines. Human Beings were too, except that they possessed a mind. Descartes was what people call it as a dualist. He believed that he mind was separate and the body was separate and he ran into all kinds of problems, philosophical problems. Unlike Socrates who had to lose his life but Descartes had to only grapple with problems.

So just like Galileo said motion can be represented by geometry so we are talking about representation, Descartes said geometry can be represented by algebra. Instead of drawing a line you can write its equations. You can use variables and things like that and you can express geometry in terms of algebra. Everything is applied math; even thought so everybody is extending this idea of math to thinking essentially. Descartes says that thoughts themselves are symbolic representations.

(Refer Slide Time 31:32)



Again if you read the book by Haugeland you will get some of these insights. Now there is a paradox of mechanical reasoning. If reasoning is the manipulation of meaningful symbols according to rational rules then who is manipulating the symbols? That's the problem essentially. The problem is it can be either mechanical; that what the people say or it can be meaningful, how can it be both? How can a mechanical manipulator pay attention to meaning? We will may be discuss this as we go along little bit because that's what are going to talk about. We will be able to represent things in some language and then we will be able to write programs to manipulate that language. But we already know that we can write programs to add up numbers. Does it know the meaning or not is the question that we will look at gradually.

But the people like Descartes were ridiculed by people, other people; they said its some faculty of will, it's some transcendental ego or homunculus, a little man; you can imagine the little man manipulating the symbols in your head. It's the kind of people making fun of Descartes. But some very interesting ideas in these concepts have been given by Douglas Hofstadter, who many of you would have heard about and I have listed three books by him here . Godel, Escher, Bach which was published in 1979 or 80 and got a Pulitzer prize; then a book called Mind's I where I is the single letter I and more recently and his most fascinating book according to me is called I am a strange man. It's a philosophical output as to how thinking is possible in the first place. So I will stop here and in the next lecture we will come closer to logic and representation and talk about things that we will be interested in.