Artificial Intelligence: Introduction to Knowledge Representation and Reasoning

Prof. Deepak Khemani

Department of Computer Science and Engineering

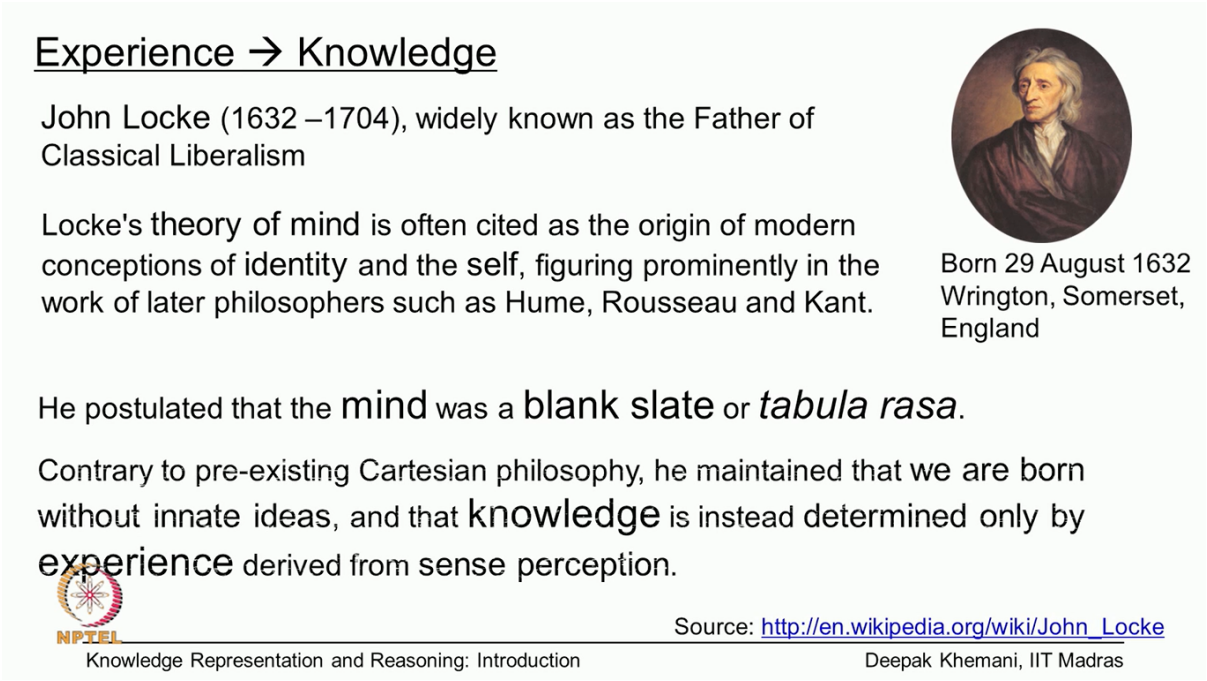
Indian Institute of Technology, Madras

Module – 01

Lecture - 02

We were looking at the history of AI and we were trying to see how this idea of knowledge representation has come. The idea of representation essentially and trying to figure out how they do thinking or how we do reasoning essentially. So let's look at some more people on the way. John Locke who is also known as the father of liberalism. His theory of mind is often cited as the origin of modern conceptions of identity and self; I mean how do people think of themselves, figuring prominently in the work of philosophers as Hume, Rousseau and Kant.

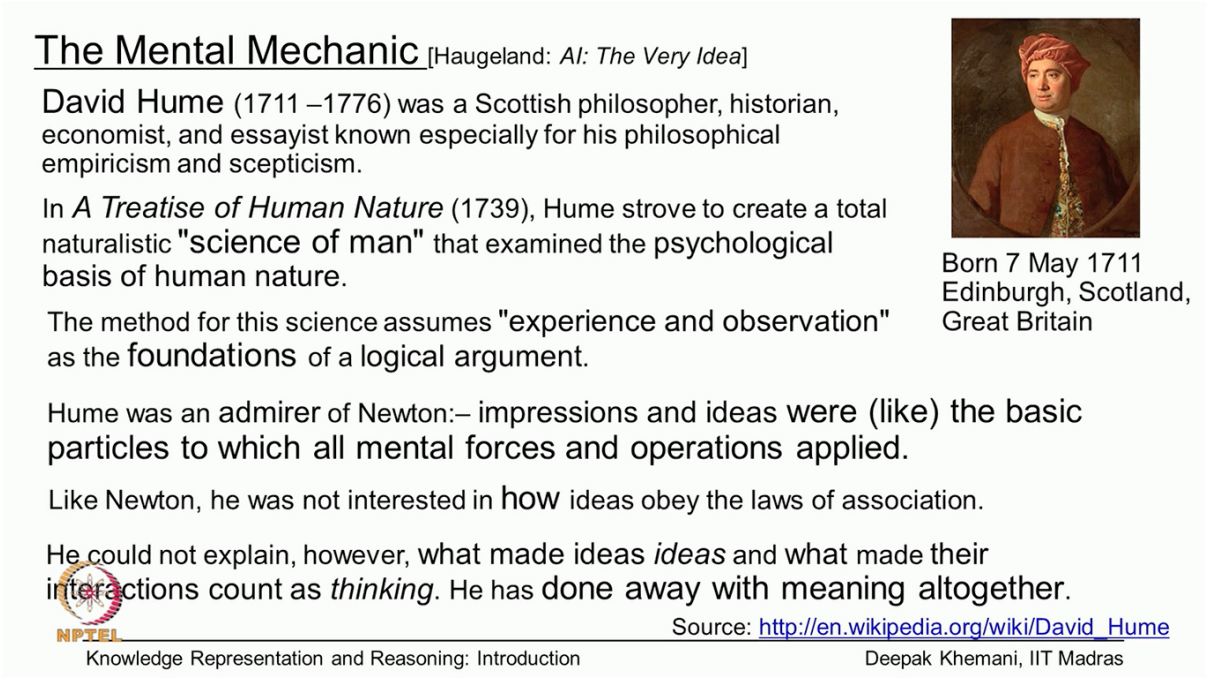
(Refer Slide Time: 00:51)



He postulated that the mind was a blank slate; we are born with blank slates or the tabula rasa as he called it; which is different from other people who believed that we are born with certain inmate ideas and concepts and notions, for example Noam Chomsky; the celebrated linguist, he believes that we are born with universal grammars and essentially what you do when you grow up in different society, that you tune that particular grammar to that particular language but he notion of the grammar he says we are born with.

As opposed to that Locke believes that we are born with an empty slate or tabula rasa and essentially what we know is determined by the experience derived from sense perception. So as we see the world around us we create concepts about the world and represent knowledge and so on and so forth. So Hume was one of the people who also exposed this; a Scottish philosopher; notice that we are still way back in time 1711-1776 he was; was like most people in those time he dabbled in many things, philosopher, historian, economist and so on. He was a follower of newton as we will see and his Treatise called the science of man he tried to examine what human beings are essentially, the psychological basis of human nature as he said and his method of science assumes that experience and observation as the foundations of a logical argument. He was an admirer of Newton: Newton of course we know for the law of gravity amongst many other things; and he felt that just as Newton could explain the notion of physical motion ideas were like the basic particles to which mental forces and operations apply.

(Refer Slide Time: 02:57)

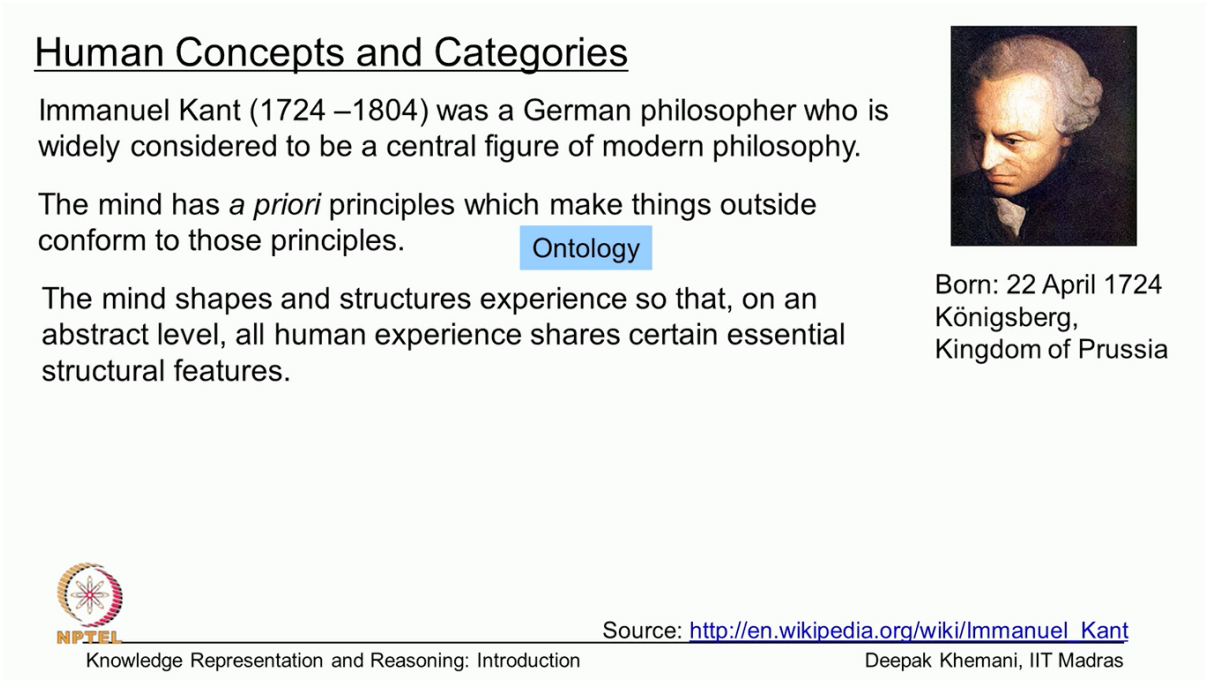


Just as we can say that two bodies which have certain mass; they exert a gravitational pull on each other and based on that we can explain why planets are going around the sun and that kind of thing. He says that mental particles which is that Hobbs has postulated also obeyed certain laws which determines what happened with particles. Like Newton he was not interested in how this happened. Newton never questioned as why is gravitational force; of course modern day physicists are still trying to understand the gravity and seems to be one of the last twenty years of physics, but Newton was not interested in explaining why gravity happens. He said if you assume that this is the laws of gravitational pull then you can explain everything. Likewise, Hume said ok we can postulate laws of reasoning then we can just accept that it happens because of that. So he is thinking of thoughts obeys some laws but he is not able to associate them with ideas; what makes them ideas so the interaction between such particles counts as thinking; so the notion of meaning he has lost altogether. So the question

of meaning; how does a manipulating agent, how does a thinking agent gets the meaning out of representation is a very hard questions in some sense that we still trying to grapple with this

So Immanuel Kant we saw was the follower of lord; when he talk we can hear the first pinches of what we now call as the Ontology. He says the mind has a priori principles which make things outside conform to those principles.

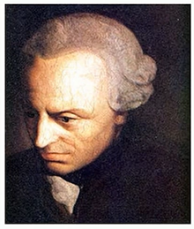
(Refer Slide Time: 04:46)



Now this is a school of philosophy which says that what we see out there is essentially controlled or determined by what we have in our head already. And Kant says that we have apriori principles and what we see from outside we conform to this principle; we will see the idea of scripts some time later in the course to show how we interpret whatever we are seeing or hearing in terms of preconceived notions that we have. He goes on to say that lines shapes and structure experienced so that on an abstract level, all human experience shares certain essential structural features. The concepts of space and time are integral to all human experience, as are our concepts of cause and effect. We are born with the notion of space and time. Now this is interesting, he says that we don't have direct experience of the world, what he called as the nominal world but what we do experience is the phenomenal world as conveyed through our senses.

(Refer Slide Time: 06:03)

Human Concepts and Categories



Born: 22 April 1724
Königsberg,
Kingdom of Prussia

Immanuel Kant (1724 –1804) was a German philosopher who is widely considered to be a central figure of modern philosophy.

The mind has *a priori* principles which make things outside conform to those principles.

Ontology

The mind shapes and structures experience so that, on an abstract level, all human experience shares certain essential structural features.

The concepts of space and time are integral to all human experience, as are our concepts of cause and effect.

We never have direct experience of things, the *noumenal* world, and what we do experience is the *phenomenal world as conveyed by our senses*.

Human concepts and categories structure our view of the world and its laws.

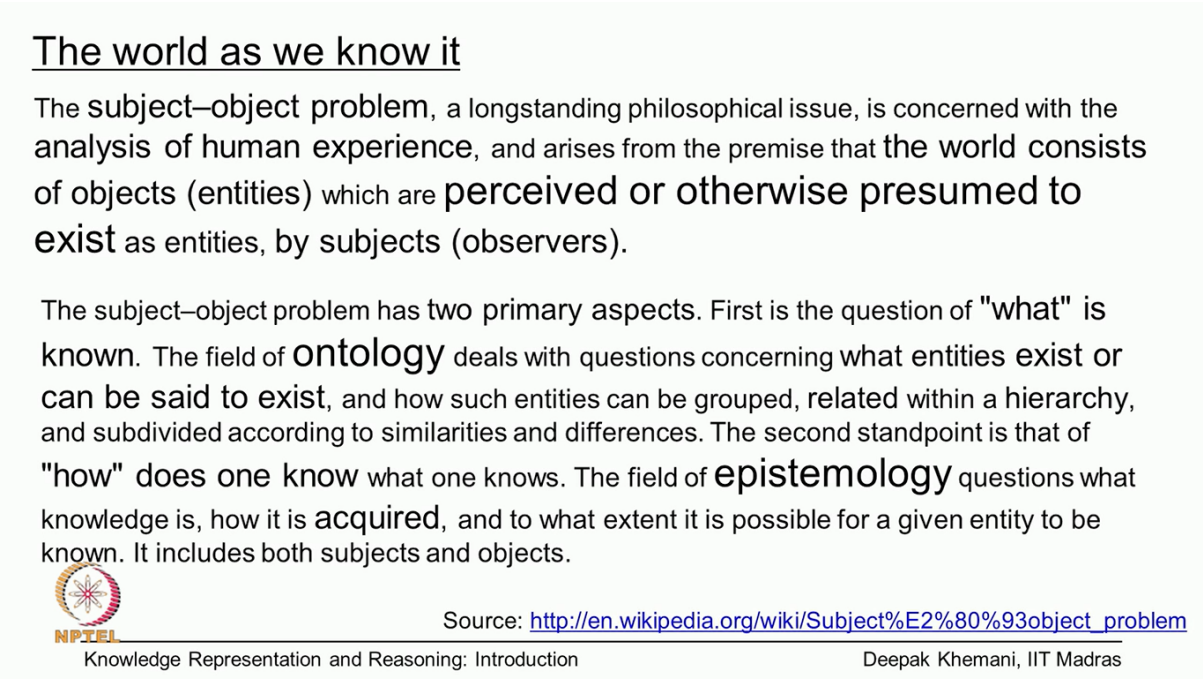
Source: http://en.wikipedia.org/wiki/Immanuel_Kant

Now that's a pool which has lot of followers. So finally human concepts and categories structure our view of the world and its laws. Concepts and categories are the terms that we use predominantly when we talk about ontologies and we will do that little bit later in the course. So the subject-object problem, subject seeing an object, a longstanding philosophical issue; so remember the correspondence theory which Aristotle posed and Wittgenstein accepted the picture theory of memory, we are still addressing the same problem, how can a subject look at an object and make sense of that object is concerned with the analysis of human experience, and arises from the premises that the world consists of object.

it is a premise which are perceived or otherwise assumed to exist as an entity. We will see if we can discuss this a little bit later some time but when we see a tree, is the concept of tree out creation or it is out there essentially. For a tree you might be willing to say that it's a tree out there but what happens when you see the clouds and things like that; when we see a cloud are we influenced by our notion of cloud essentially. So the subject-object problems has two primary aspects, these are modern day terms that we will use. First the question of what is known.

The field of ontology deals with the questions concerning what entities exist or can be said to exist. The word ontology was used by the philosophers; what is this world, what is it made of, what is an apple, what is a cow, what is a cheeta and concepts essentially; and how such entities can be grouped, related within a hierarchy, and subdivided according to similarities and differences. This is what we will do when we construct a taxonomy of some kind essentially.

(Refer Slide Time: 08:02)



The second standpoint is that how does one know what one knows. The epistemological problem is how is knowledge acquired and what do we know essentially; in some sense it is talking about facts essentially. Ontology is taking about concepts. What are the kind of concepts we have, I have a concept of chair, I have a concept of light; whereas this is particular light, that is a particular chair, that is epistemology essentially so knowledge based systems can also be called as the epistemological systems, and the bounds of our own mind. This is Kant we are still talking about this thing. He claimed to have created a Copernican revolution in philosophy. Copernicus created a revolution by telling us that what you see is not what really out there, you see the sun going round our earth, rising and setting but it is really the earth which is rotating essentially; and similar notion was coming was the notion of mind essentially so if you see, what you see in the mind is not really out there essentially.

But we are bounds by our own minds and moral philosophy of the autonomy of practical reason; it says that we are constrained by our reasoning abilities. I will skip some of these stuffs but the highlighted stuffs in blue. He is talking about the minds process, the product of rule-based activity so our interest is not only in the representation but in the reasoning. How can we operate upon those representations to come up with new things, new ideas, or new conclusion and so on and Kant is also talking about the mind process, that the mind has processes which work on these representations which for some he has called as rule-based activity, for us rule-based activity will be a more precise term which we will see. Just a quick recap of how AI got its name. It is credited to John McCarthy and Marvin Minsky and Claude Shannon.

McCarthy was one who was supposed to have suggested it. They organised a conference called the Dartmouth conference in 1956 which was to be two month, ten-man study of artificial intelligence on the basis of the conjecture that every aspect of leaning or any other feature of intelligence can in principle be so precisely described that a machine can be made

to simulate it. So this is the sentence where the term Artificial Intelligence was used first by McCarthy. And you can find this description in the history of Dartmouth Conference in a full chapter in the book called Machines who think by Pamela McCorduck. So these are the organizers McCarthy, Minsky, Rochester, an IBM engineer, Claude Shannon as we all know gave us the concept of information theory. We will look at some work by McCarthy and Minsky much later in the course when we look at the notion of default reasoning or something called circumscription which is due to McCarthy, Minsky we will look at the idea of structure knowledge; you know how we group together the things which eventually lead to this thing called object oriented programming.

The show stealers in Darthmouth Conference were two relatively unknown people, Herbert Simon and Alan Newell who apparently according to McCorduck atleast almost as an afterthought essentially. They were the ones who created the biggest impact; along with J. C. Shaw also from RAD Corporation; all this is the part of CMU or Carnegie Mellon University. They build a program called the logic Theorist. Now we will look at such programs to start with. It was the first program deliberately engineered to mimic the problem solving skills of a human being and by this we mean logical reasoning the kind of syllogism that we saw; given some premises what else can you conclude; can a program do it; that was the program Logic Theorist, one of the earliest theorem provers. We will spend some time looking at theorem proving in this course. But in fifties and sixties when the Darthmouth Conference was going, they were already demonstrating this Logic Theorist. It went on to prove several theorems in Russell and Whitehead's celebrated (Principia Mathematica finding shorter and more elegant proofs for some) some comment about creativity you might say essentially. So these are the two characters Simon and Newell. Herbert Simon is probably the only noble prize winning computer scientist but he got the noble prize for economics not for computer science as we don't have a noble prize for computer science. And his student Newell; both of them set up a big school in Carnegie Mellon University which was very influential. So they become leading figures in AI research at CMU; they wrote a program called General Problem Solver which basically characterise heuristic search and some problem solving strategies like mean-ends-analysis. But more importantly they provided the information processing approach to AI, whatever you call it as classical AI now a days. One shining example at CMU was a program called SOAR which we will look at little bit later in this course and maybe we will have to write some programs; it's a programming system or cognitive modelling system which we will use to write some programs developed by John Laird I think.

(Refer Slide Time: 13:47)

Simon and Newell

Simon and Newell went to become leading figures in AI research and founded a strong group at CMU.

Their program General Problem Solver (GPS) was a pioneer in the use of heuristics in search and adopted a human like approach called means-ends-analysis.

Their work defined the Information-Processing approach for AI.

At CMU one shining example was the development of the SOAR cognitive architecture by John Laird.

Now this is the one of the basis for what we want to do is the physical symbol hypothesis given by Simon and Newell. So when we are talking of representation now a symbol is the perceptible something that stands for something else. Alphabet symbols, numerals, road signs, musical notation; all these are symbols. If I write a squirrel on a black board then you perceive it as a number seven for example, now it is squirrel which a symbol which is standing for number seven so somehow we manage to represent the notion of number seven by some strokes of chalk or pen. A symbol system is a collection of symbols it could be a pattern-words, arrays, lists, even a tune. And a Physical Symbol System is something a symbol system which obeys laws of certain kind so remember the kind of stuff that Hume is talking about.

(Refer Slide Time: 14:51)

Physical Symbol Systems

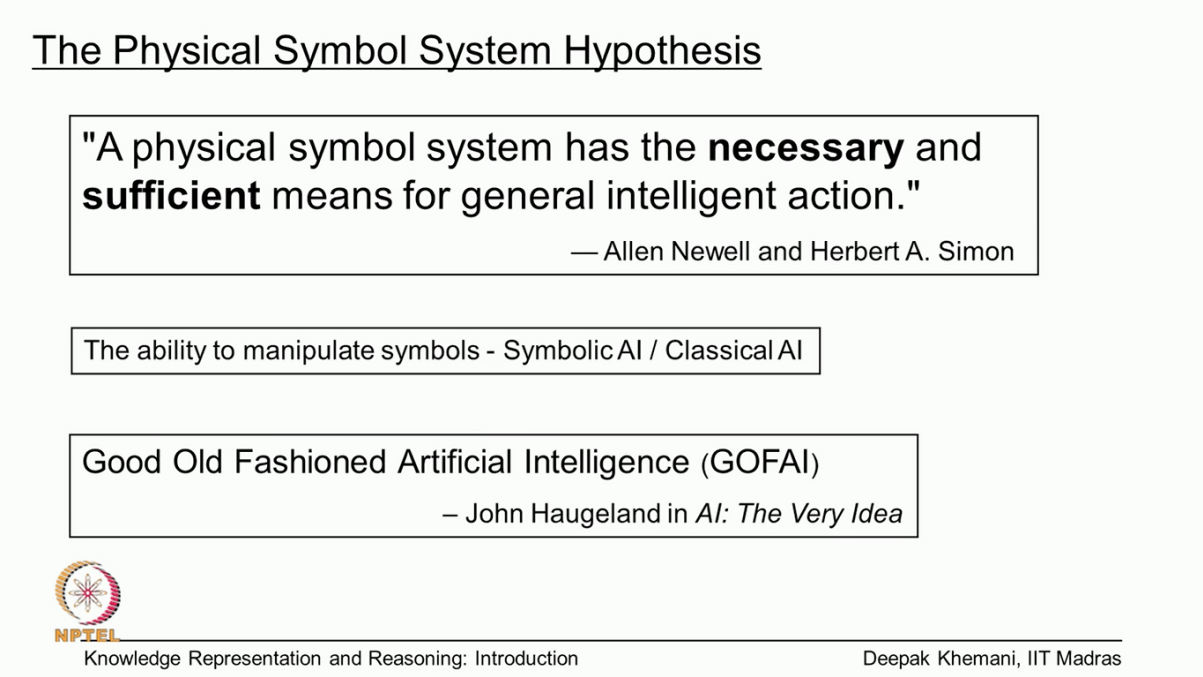
Symbol : A perceptible something that stands for something else.
- alphabet symbols, numerals, road signs, musical notation

Symbol System: A collection of symbols – a pattern
- words, arrays, lists, even a tune

Physical Symbol System: That obeys laws of some kind, a formal system
- long division, an abacus, an algorithm

That you manipulate symbols using certain rules just like gravity controls the physical motion some rules which control the manipulation of symbol system. So the algorithms, we are familiar with algorithms so we will also use the term algorithms, long division for example. So the Physical Symbol System Hypothesis, it's a hypothesis, you cannot talk of it is a law or something. What Simon and Newell say is that a physical symbol system is necessary and sufficient to produce an intelligent behaviour which is of course different from what some people called as strong AI which says that you have to have a certain kind of architecture you have to have a certain kind of hardware for AI or for the machines to be able to think. What Newell and Simon are saying is nothing, the ability to represent something and to manipulate those representation is enough to create intelligent system and that's known as the physical symbol system hypothesis and the whole of classical AI is built on that basis.

(Refer Slide Time: 15:55)



Haugeland calls it as good old fashioned artificial intelligence which is as I said in contrast to strong AI and other approaches to interesting stuff I should say of intelligence, statistical approaches or neural network. In neural network you don't have a representation, you don't know how you represent something. Somehow there is a complex network of neurons and connection between the neurons and the weights that we attach with those connections somehow captures the concepts that we are trying to capture so of course we can learn or train a system to know what is a, what is b, what is c, you can do character recognition but we don't know how the system is represented whereas in classical AI we have our very clear explicit representation which we work with. Ok so since we are talking about agents representing the world around them, let's ask some basic questions. What's really out there, what is the meaning of objective reality.

If you look at it from the fundamental point of view one thing is clear that everything in physical world is made up of some number of fundamental particles. We may dispute, we may debate about what are the fundamental particles but let's assume for the time being that

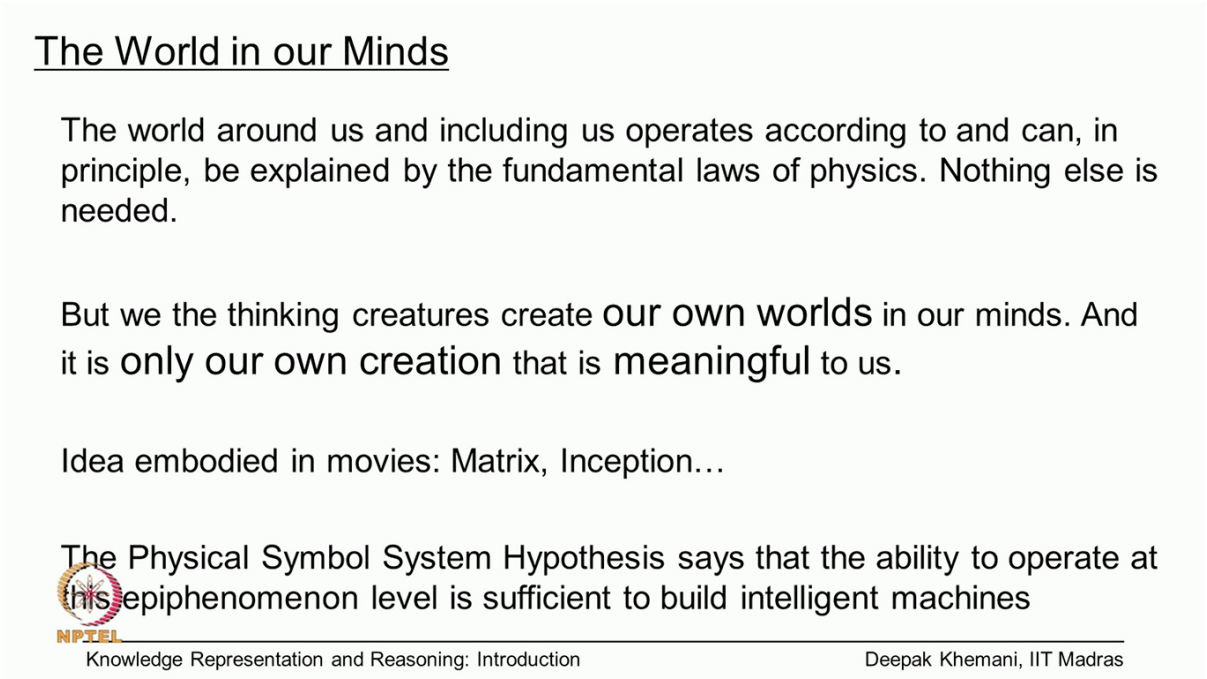
we accept something like atom to be the fundamental, of course we know that atom is not a fundamental particle, it is made up of protons, neutrons and electrons and they in turn are made up of still smaller particles but let's not worry about that but at some level we assume that there are some fundamental particles unless we believe in something called idealism in which we say that everything is the world of ideas and matter is only a manifestation of ideas but let's not get in to those philosophical things. Let's assume that the world is made up of fundamental particles.

And once we have fundamental particles, the laws of physics are enough to explain the behaviour of these particles. So in principle at least you can understand the world in terms of particles that anything is made up of and apply the laws of physics but of course in practice as you can guess, that's not a good idea. For example if you are talking about human being then we are typically made up of ten raise to twenty seven atoms which is number that you should try to think about. It is not something that you can imagine quite easily. Just think about it terms of, let's say in terms of time how much is the ten raise to twenty seven seconds.

Now these ten raise to twenty seven objects between people like you can me interact with zillions of surrounding atoms so we breathe air, we eat food, vibration from air molecules sound waves and so on and so forth. Can we ever hope to understand the world in terms of these fundamental particles, and can we write down equations and solve them. Even if we could millions of equations and solve them what would we get out of this. We would only get the prediction of their location and movement. Whereas what we want to know is you know some high level what many people call it as epi phenomena. If you through a ball at somebody then somebody will stretch out his hand or her hand and try to catch it. Now this kind of reasoning you can't even imagine doing it with fundamental particle; so we, the world may be made up of anything but we as human being are forced to think of it in our terms; so this the idea of Kant that we have the ideas in our head and we see the world in those ideas. It's a necessity because obviously we cannot reason with ten raise to twenty seven particles. So we create the notion of concepts, we create the notion of people, chairs, lights, apple, banana everything and we reason at that level of representation.

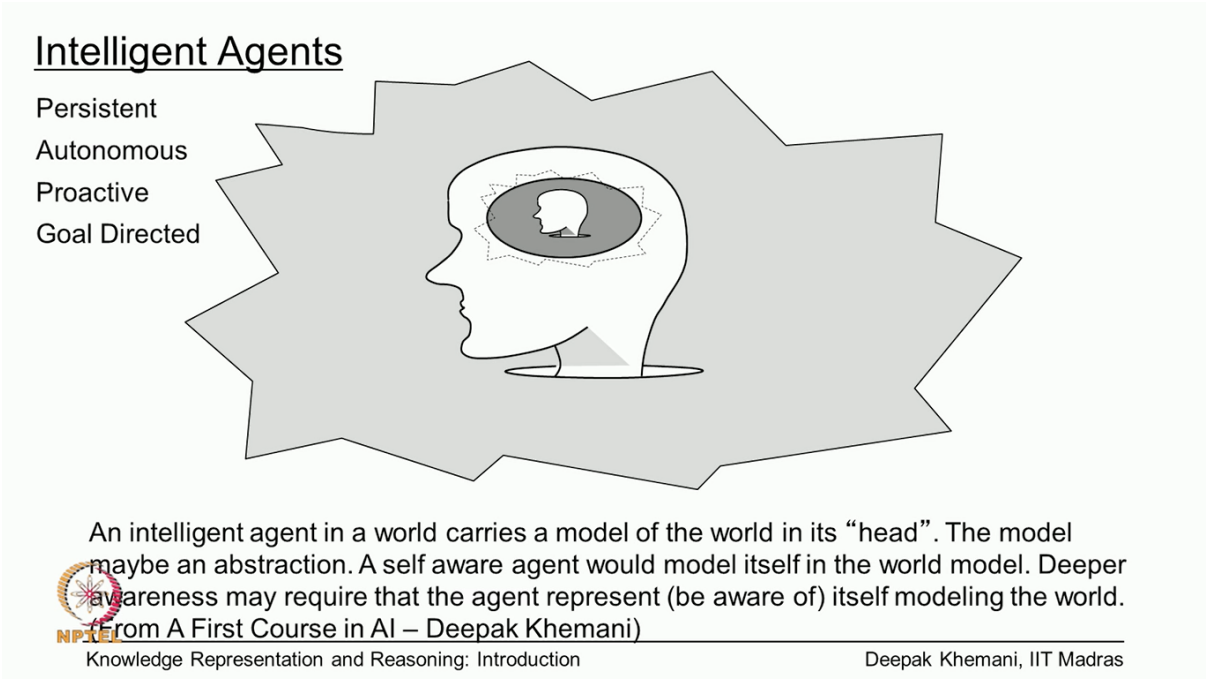
The worlds are in our minds, the world that we think are in our minds. In principle we can explain the world with physics so I can try to predict, let's say that this fielder on the cricket field is going to try to catch the ball by applying laws of physics but you can imagine how mind boggling that exercise would be, we tend to think at different levels essentially. We create our own worlds in our minds and these are the representation sets we are talking about; even as human beings essentially so we are not necessarily talking about machines. These are the kind of ideas which were exploited in the movies like Matrix for example or Inception where what's going on in the mind is not necessarily direct correspondence with what is out there in the world. So we cannot reason at that level, we cannot use physics to understand the world, we have to create our own scheme of things.

(Refer Slide Time: 21:01)



And what Newell and Simon have said is that the physical symbol system hypothesis is enough essentially. So this is what I was talking about an intelligent agent, what is an intelligent agent. An intelligent agent is something, so if you see this diagram, you see a face essentially, so that's a face of an agent and inside the face, inside the agent's head as you might say is the model of its world which is a greyish circle that you see.

(Refer Slide Time: 21:30)



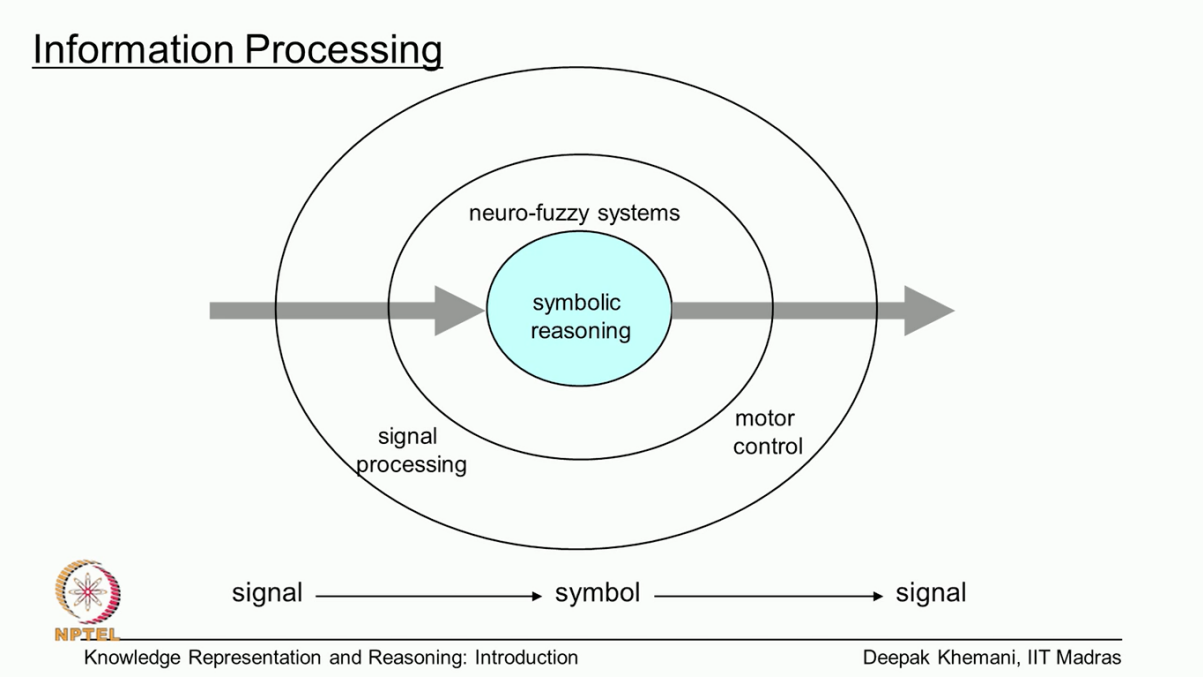
So intelligent agent has to create some sort of a model of the surrounding of the world it is operating in it. And a self-aware agent, because many people have argued that self-awareness is critical to intelligence; that humans are intelligent because we are self-aware, we have consciousness and this kind of things, but we can sort of just say that self-aware agents will have the model of themselves inside the model of the world they have so the grey oval that you see is the model of the world that the agent have; it may not be a perfect model because we can never represent the world perfectly, it may be some approximation and that's true of all kind of modelling activities. We always approximate certain features and in that model if you create a model of yourself then you are self-aware essentially. So when we talk of intelligent agent we talk of persistent, autonomous, proactive, goal directed entities.

We can think of them as programs of people. So when we do information processing, of course when we interact with the world, when you are looking at me, when you are hearing what I am saying, we are not doing at the symbolic level, I mean what you get is rays impinging on your eyes and what we get in some sense analogue signal essentially, the sound we hear are basically vibration and so on and so forth. So there is a layer, this is my model of how thinking happens, this is the layer of signal processing on the left hand side when we sense the world, then there are some neuro-fuzzy systems which can convert signals in to symbols.

So for example a neural network can take a pixel array and convert in to a character A, can say this is character A character B and so on. It can convert signal in to a symbol and what we are concerned with in classsical AI is symbolic reasoning which is the inner core of reasoning then we can do some decision making and do actions which is on the right hand side so there

may be things like motor control, you move your arm if you want to throw a ball and so on and so forth.

(Refer Slide Time: 23:51)



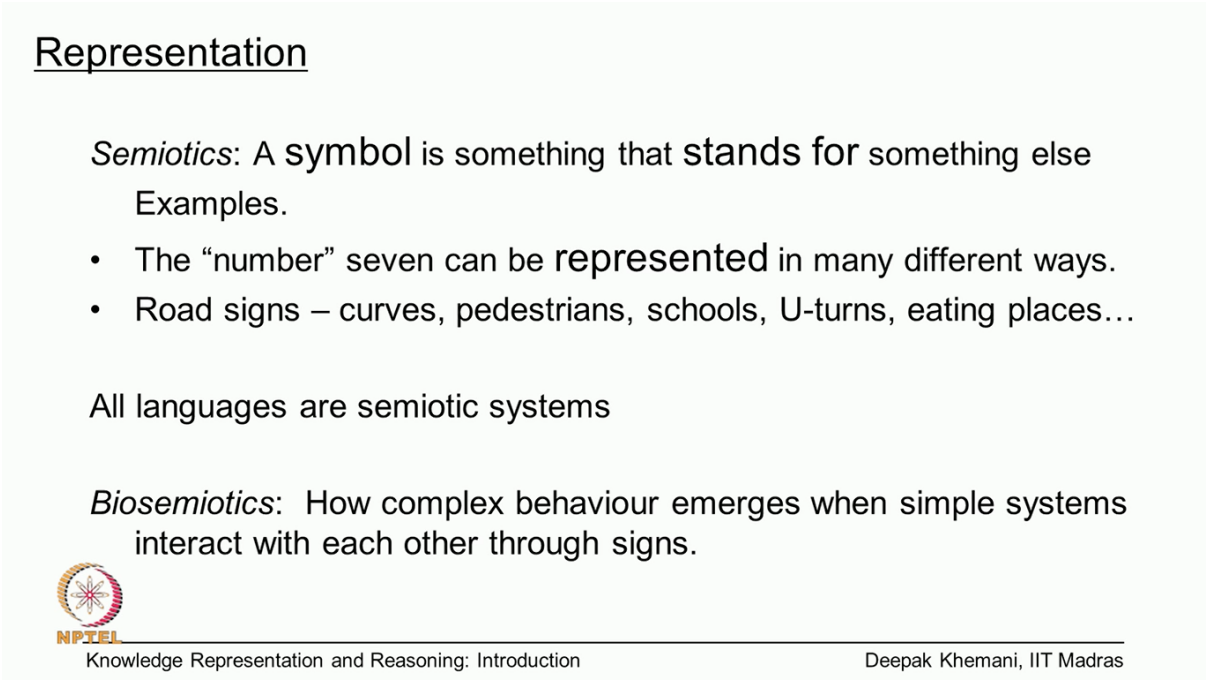
So AI has many many topics which are listed here including everything in some broad sense falls in to AI even if you talk about neural networks or pattern recognition or image processing or speech recognition or graphics or robot control but the dotted circle kind of represents more of classic AI the symbolic AI that I am talking about and within that also there are many things, for example qualitative reasoning, machine learning, memory, planning, adversarial reasoning or game playing. We will be focussing on what is in blue here which is predominantly knowledge representation, logic, ontology, semantics, certain amount of search for reasoning, handling uncertainty and little bit of natural language understanding. Let's just give a few definitions to these terms. So what do we understand by the world knowledge is basically whatever you know. Humans deal with knowledge of many kinds. We have models of the world we live in. We have models of ourselves in the world. We have knowledge of society, knowledge of facts, knowledge of how to do things; what is correct in one situation, what is not correct; all these is knowledge which we have to deal with essentially. Ontology we have already mentioned is knowledge about how the world is conceptualized or what are the conceptual categories in the world so there are apples, there are goats so on and so forth, that is part of the study of ontology. Of concepts and the relation between them. Epistemology is knowledge about what is true in this world essentially. A knowledge based system typically refers to systems that employ domain specific problem solving knowledge in some form or the other essentially. And a word about memory so we not going to be very much concerned about memory here but it's our repository of knowledge.

The knowledge we have sits in our memory essentially and human being's memory is dynamic essentially. We keep continuously learn and update our memory and so on and so

forth. So just let's repeat what we talked about representation. Just let me throw another word called Semiotics. Semiotics is the science of symbols essentially and according to Semiotics and symbol, we have already mentioned this when we talked about the physical symbol system hypothesis. A symbol is something that stands for something else essentially and we have said number seven can be represented in many ways, these are different symbols for a concept. Whatever number seven mean to you so just try to think about what is the number mean to you essentially and what is seven, I mean apart from the representation that we are so familiar with.

Road signs that we see all the times; curves, pedestrian, school; they tell you there is a school here, pedestrian crossing here or there is a U-turn you can take here or there is a eating place coming up. These are kind of signs essentially. All language that we work with are Semiotic systems, they are made up of symbols including the languages like English or Tamil or Hindi. All formal languages, programming languages like C plus plus or java. All are Semiotic systems in which symbol stands for something else so it's a collection of symbols which stands for something else. There is also term called Biosemiotics in which how complex behaviour emerges when simple systems interact through signs essentially.

(Refer Slide Time: 27:42)



So there are many examples about bees interact with each other, how ants and ant colonies leave small small messages for each other and that kind of things so all these are the Semiotic systems, they communicate through signs essentially. Here is a nice picture I took from this book called Conceptual Structures by John Sowa. Essentially to ask question that many people ask, why don't you use natural language, English or hindi for representation.

The reason is as you can guess, the richness of the language or the ambiguity of the language, verbosity of the language, impreciseness of the language. Which is why if you happen to read

any legal document where they try to be really precise you can see that most of us can't make any sense of any legal document because it is so verbose essentially. So it is good medium for communication, but here of course the diagram is saying that that is also not very successful. A poor policeman trying to give directions to someone. There are people who say Sanskrit is language of computers but we will not get in to that. The only thing you can say about Sanskrit is that it has very well defined rules, it's mostly unambiguous, largely unambiguous.