**Research Methodology**
**Prof. Soumitro Banerjee**
**Department of Physical Sciences**
**Indian Institute of Science Education and Research, Kolkata**

**Lecture - 05**
**Causality Part 01**

Today we will talk about one of the central doctrines in science: Causality. Causality is almost at the basis of all science. Because whenever something happens, we believe that there must have been a cause and we try to find the cause, and much of science is built by doing so: looking for the cause for something that has happened.

So, causality is one of the central doctrines in science. But even though it is such a central issue in science, for much of human history it was not clear how to define the cause of an event. And so, it has been subject of great controversy how to identify the cause of an event, what do I call a cause. But, through the millennia of attempts presently we have obtained some clarity and we do understand what can be seen as the cause of an event and that is what we will talk about. I will take you through the course of human history through which this idea developed.

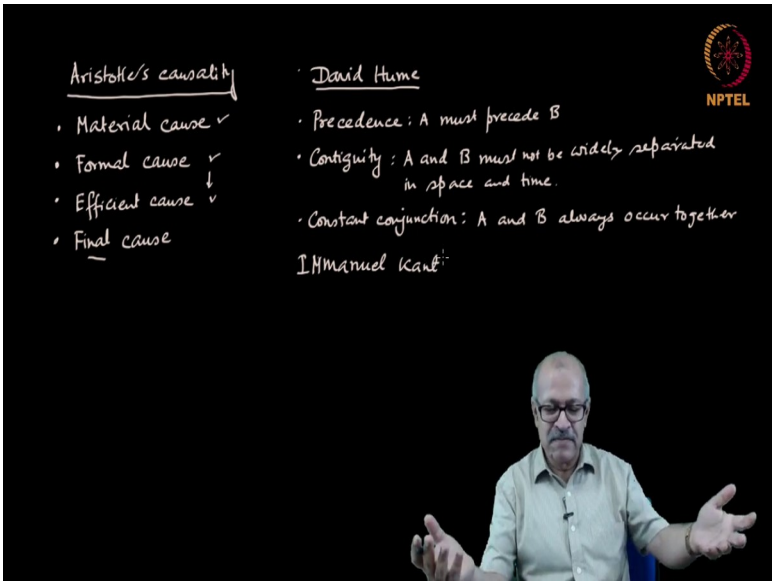
It is not difficult to see that some rudimentary idea of causality must have been there for any human action to take place. And for that matter, even the ancient people at the dawn of human civilization, they also must have had some rudimentary idea of causality. Tigers cause death and therefore stay away from tigers. The seed of today causes the tree of tomorrow. And so, place that seed where you want the tree of tomorrow. That is the start of agriculture.

It is not difficult to see that some rudimentary concept of causality must be there in order for these things to develop. Even for a normal day-to-day life of an ancient living in the hunting gathering society, some rudimentary idea of causality, what is causing what, must have been obtained. Otherwise it is practically impossible to even lead a day-to-day life.

But, these are intuitive unrefined ideas of causality. We see the first refined and articulated idea of causality in the Greek period, in the writings of Aristotle. Aristotle saw that whenever we are trying to figure out the cause of any event, we can look for the

cause from four different angles. He specifically gave an example to illustrate what these four different angles are and I will cite the same example.

(Refer Slide Time: 04:01)




Let us start our discussion of causality with Aristotle. Aristotle's causality. He said, suppose there is a marble statue and you are trying to find the cause of the marble statue. How would you look for the cause? You might say that the statue is made of marble and therefore, the marble must be the cause. Because if the marble were not there, that particular kind of white stone—if that were not there, then the statue would not be there.

So, the material with which the thing is built is a cause and he said it is the 'material cause'. So, what is the material cause? If you are looking for the cause of anything, there is some material with which it is done, and if that material were not there the thing would not be possible and so, that material is a cause.

In case of the sculpture the white stone, marble, is a cause. That is the material cause. Then he said that the sculptor, the artist who made the sculpture, must have a have the idea of the form of that sculpture in mind before he started to work on the stone. Only if one has an idea of what he is going to make, then only one can start working on the stone. So, the *form* of the sculpture must be there in the sculptor's mind in order for him or her to make the sculpture, and that Aristotle said, is the 'formal' cause.

And then he said that, of course, the sculptor himself—the artist himself—is a cause because if the artist himself or herself were not there, then the sculpture would not be built. So, the agent, in general the agent which actually does it, the agent which actually makes it happen, that is of course a cause and he called it the ‘efficient’ cause.

So, the difference between the formal cause and the efficient cause, according to Aristotle, was that the efficient cause is the person or the agent which has done it, the artist in this case, but the formal cause is the idea of the form of the sculpture the artist had in mind before he started working on the stone. And so, that form of the final thing is what is the formal cause.

And then he said that there must be some purpose for which this sculpture is built. Some end result to fulfill which this sculpture has been made, and that he said, is the ‘final’ cause. And in general, if you are looking for the cause of something, you have to figure out what is the ultimate objective of that to happen. What purpose does it serve. So, that is the final cause.

Aristotle gave this particular example, but he said that in general, whatever cause you look for, whatever events you investigate, you look for the cause from these four different angles and you will find four different types of causes. So, that was Aristotle’s idea of causality.

The Greek period was followed by the Roman period and the Roman period was followed by what is known as the Dark Age, the medieval period in Europe. In between Christianity developed and during the medieval period that became the dominant form of thinking. During that period Aristotle’s ideas actually were accepted and integrated into the church dogma.

In that period, the final cause became the most important, because whatever happens, people wanted to see the hand of God in everything that happens and that was seen as the final cause. So, the final cause became all important during the medieval period. After the medieval period came the Renaissance, during which we have giants in various fields of human endeavor. We have Copernicus, Bruno, Galileo in field of science, we have Leonardo da Vinci, Raphael, Michelangelo and a large number of artists, similarly in literature. So, culture sprouted in all directions.

The scientists of the time, obviously, had to worry about causality because they were investigating natural phenomena without the baggage of the belief systems of the earlier period. Naturally they had to worry about the cause of various things. But, we do not see a full treatise on causality in the writings of the renaissance scholars.

Only in Galileo's writings we find a rejection of the final cause. He said that I do not believe in any final cause. He rooted only on the material cause and the efficient cause. He pointed out that the formal cause—the form or the idea that the sculptor had is a part of the person. So, the formal cause actually can be subsumed in the efficient cause. So, he was actually saying that when you look for causes look for the material cause and the efficient cause. What it is made of and who made it. But beyond that, we do not see much treatise on causality in the Renaissance scholar's writings.

Following renaissance, there was a period which is called the 'period of enlightenment'. During that period, we see the first good treatise on the problem of causality—a philosophical treatise on the problem of causality from a British philosopher called David Hume.

Hume wrote a book titled 'A Treatise on Human Nature' in which there was a reasonably good exposition on the problem of causality. His attempt was to develop some way of defining causality which can be tested. He said that if I say A is the cause of B, then the following conditions must be satisfied.

First, 'precedence': A must precede B. In order for A to be called the cause of the effect B, A must precede B. Second, he said 'contiguity', which means A and B must not be widely separated in space and time. So, let me write: A and B not separated in space and time, that is, they have to be contiguous. For example, if a murder has been committed the police would look for the murderer in the immediate neighborhood of the place. It follows from common sense. And third, his point was, if A is to be called a cause of B, then another condition has to be satisfied, which is called constant conjunction.

Constant conjunction means A and B always occur together. That means, whenever A happens, B must happen. Or, whenever B happens, before that A must have happened. Whenever such constant conjunction is observed, we can say that A is the cause of B. He said that, if these three conditions are satisfied, then I can call A to be the cause of B.

Notice that all these three conditions are testable. Whether A is preceding B or not: testable. Whether A and B happened more or less at the same place or not: testable. And constant conjunction: whenever A happens B happens or not that also can be tested.

So, he developed testable criteria to define causality and that was a big step because now things were moving in the direction of testability, moving in the direction of materialism. His step was in that direction. However, soon it was noticed that his criteria of defining causality had some crucial flaws.

For example, let us take the issue of precedence. One German philosopher Immanuel Kant, in his book, pointed out the problem of this precedence requirement. He said that, suppose there is a sofa and you place a heavy iron ball on it; it will make a dent. What caused the dent? Obviously, the pressure of the ball. But the pressure or the ball resting on the sofa and the dent appearing happened at the same time. It is not that you first place the ball and then after some time the dent appears. And therefore, Kant made the point that, you cannot say that A must precede B. Rather you should say that B cannot precede A. The effect cannot precede the cause.

Let us come to the second issue, contiguity: the cause and the effect should be contiguous. They should not be widely separated in a space and time. But, take the event of the tide. The tide is caused by the attraction of the moon. Therefore, the cause, which is the moon, is quite distant from where the effect happens, and they are not contiguous.

The third, constant conjunction. Suppose you want to check whether mosquito bites cause malaria. Let us take this case. Do mosquito bites cause malaria? Now, if mosquitoes containing the parasite bite a hundred people, maybe 90 will be afflicted by malaria, but 10 will not, because of their own body resistance. Therefore, the constant conjunction (whenever A happens B happens) will be violated. Cannot we then say that mosquito bites cause malaria?

According to Hume, since constant conjunction requirement is violated, you cannot say. Thus, we realize that that criterion is faulty. We have to do something else in order to work out the causality. So, that was the situation after David Hume.

Another point was, Hume said that causality actually is our mental construct. We see things happening in succession, one after the other. We see cloud and then the rain and

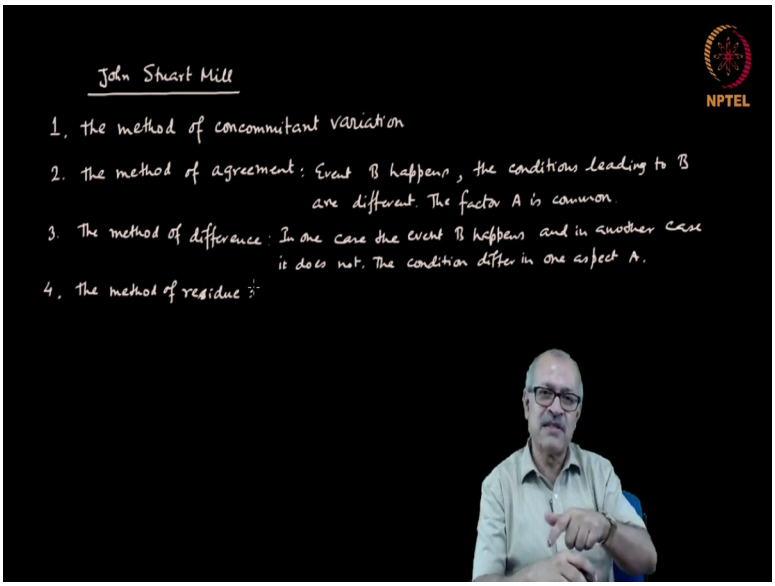
so, our mind forms a link between them. Our mind establishes a causal connection. Hume said, it is not really necessary that nature has such a causal connection. It is our minds finding a pattern in the events in nature.

But Kant made the point that we see B following A always, and then we see a pattern in nature, that is because there is a physical causal connection between A and B. A is actually causing B. It is not just arbitrary things happening in succession. So, Kant's point was that there is a physical way by which A is influencing and making B happen and that is why causality is actually a physical phenomenon.

There is causality in nature and that is why we see a succession of certain things, one after the other. That was Kant's point.

How do we look for causality? How do we plan experiments to find causality in nature? After Kant, in the middle of the 19th century, this point was illuminated by John Stuart Mill.

(Refer Slide Time: 21:28)




So, now we come to the work of John Stuart Mill, a famous philosopher of the 19th century. He proposed certain criteria by which one can design experiments and one can test causality. One can specify what causes what.

If, in a specific situation, you are trying to find out what caused a specific event, John Stuart Mill gave the prescription as to how to look for the cause. The 1st is the method of

concomitant variation. What is it? He said that if there are two variables A and B and whenever A increases you see that always B either increases or decreases monotonously then you know that there is a causal connection between A and B.

So, whenever A is increasing, it is a question of concomitant variation. So, A is varying, and B is varying concomitantly: either whenever A increases B also increases or whenever A increases B always decreases. When that happens, then you would say that there is a causal connection between A and B. You cannot say A causes B because it is possible that the opposite is true: B causes A. But there is a causal connection, that you can assert. So, that is the method of concomitant variation.

The 2nd is the method of agreement. Imagine two situations in which some event B happens. The event B happens, and the situation preceding to that are different. Let me write. The event B happens in many situations, and the conditions leading to B are different. But one particular factor A is common to all of them. So, the event B happens in various situations, the situations differ, but there is a common factor between all these situations: the factor A. In that case Mill says that you might call A to be the cause of B. This is the method of agreement. That means, all the situations agree on one point, the existence of the factor A. Then you might say A is causing B.

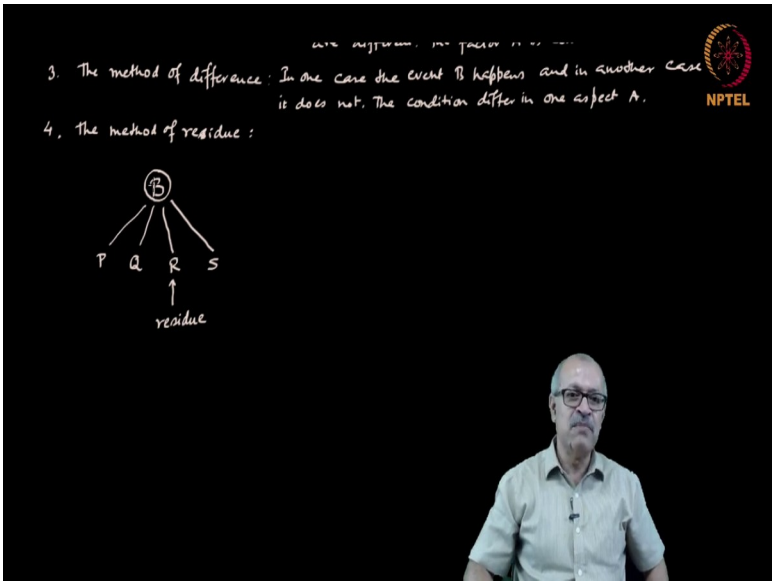
The 3rd is the method of difference. The method of difference is where the event B happens in one case and does not happen in another case. So, it happens and does not happen and the antecedent situation, the conditions prevailing before the occurrence of the event B, are more or less the same in both cases, but they differ in only one aspect. So, in one case the event B happens and, in another case, it does not. The conditions prevailing before this happening and not happening are more or less the same, they only differ in one aspect. The conditions differ in one aspect, let us call it A. So, in one case A was there and B happened, in another case A was not there and B did not happen. But otherwise the conditions were the same. In that case you might call A to be the cause of B. So that is the method of difference.

And the $4^{th}$ is the method of residue. There are situations where there can be various factors existing in the condition preceding a particular event: various factors say P, Q, R, S. These are various factors that are existing in the condition preceding an event B, which is happening.

Now, from prior research you know what is caused by P and that is not B. You know what is caused by Q and that is not B. You know what is what can be caused by S and that is not B. But B actually happened. Now, in the antecedent things there is one thing, say R, whose result you do not know, whose effect you do not know.

Thus, there is a residual condition existing prior to that event, whose effect you do not know. But you find that something else happened, B happened and B cannot be the result of P or Q or S. Then you might say then the residual condition that was existing, R, is the cause of B.

(Refer Slide Time: 29:41)




So, in this situation you have the event B happening and before that the situations prevailing had the aspect, let us call it P, another aspect, let us call it Q, another aspect, let us call it R, another aspect, this is called S. All these were there in the condition prevailing before the occurrence of B. So, B is the effect whose cause we are trying to find out.

And from prior research you have clear idea as to what can be caused by P. So, you know that B cannot be caused by P, you know B cannot be caused by Q, you know B cannot be caused by S, because you know what can be caused by S, Q, P. But you do not know the effect of R. This is the residual part, whose effect is yet unknown.

You know that B cannot be caused by P, Q, S, but you do not know what is the result of R. Something else has happened which cannot be caused by P, Q and S. So, R is the residue. Then you might say that R is the cause of B. So, that is the point of the method of residue.