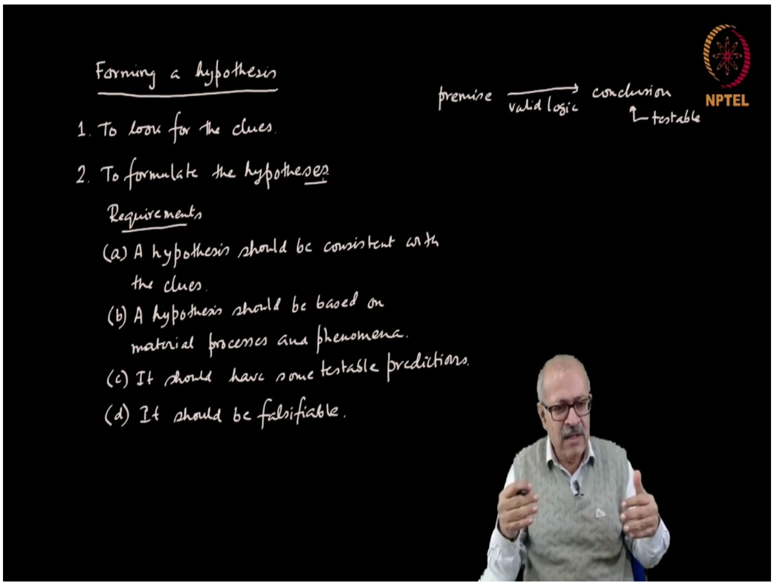
Research Methodology
Prof. Soumitro Banerjee
Department of Physical Sciences
Indian Institute of Science Education and Research, Kolkata

Lecture - 27
Proposing a Hypothesis - Part 01

In the act of doing science, we ask questions and the questions are of the type of ‘how did it happen?’, ‘why does it happen?’, ‘when did it happen?’ -- that kind of questions. Something happened, or something is observed, a phenomenon is observed, and then we ask such questions. Then, I said, that the standard way of answering or attempting to answer the question is to form hypotheses.

(Refer Slide Time: 00:58)



Today we will be discussing the topic: forming a hypothesis. How to form a hypothesis? What are the scientific ways of forming a hypothesis and so on and so forth. But we have to understand that, in answering a question, the standard method is to make scientific guesses. Not one scientific guess but many scientific guesses, and to eliminate the wrong ones.

In this pursuit, the method adapted by a scientist is quite similar to the method adapted by a detective, when such a person is called to find out the perpetrator of a crime. Just imagine a situation like Sherlock Holmes; he is sitting in his chamber and there are footsteps on the stairs, somebody comes in and reports a crime has happened last night. Sherlock Holmes goes there and does something. What does he do?

He, on the first day, looks for clues: a footprint here, a bit of finger-print there, maybe cigarette butt put somewhere, or some object located in a place where it is not supposed to be located, so on and so forth. Those things he observes and then comes back home. These form his clues.

The clues, on the basis of which, after he comes back home, he would form hypotheses. He would guess that this must have happened, that must have happened. When the clues are insufficient, then he would be able to formulate many hypotheses that are consistent with the clues. So, he has a bank of such hypotheses and then his task could be to eliminate the wrong ones.

And then his logic, line of argument, would be that 'if this hypothesis is true, then I must find this', 'if that hypothesis is true, then that must happen' and so on and so forth. Then he goes and checks. And that way he eliminates the wrong ones and finally, the day he announces who is the criminal, that day he has basically eliminated all the other hypotheses and zoomed onto one. This is how science also works.

Whenever we have a question in mind, we first try to find out the initial information that is available about that event or a phenomenon. People may have observed it, people may have studied it, people may have experimented on it, have measured something. So, those form our clues . When we have the clues in hand, then we proceed to form the hypotheses.

So, the first task in forming the hypotheses, the first task is to look for the clues. When we have the task of looking for the clues, we basically try to find out what is already known about that phenomenon or event. For example, if you have question regarding the origin of the solar system, then the initial clues would be whatever had been observed: that all the planets move in elliptical orbits with the sun at a focus, the fact that the sun is much bigger than the planets and so on and so forth. The observed things. The fact that the solar system is almost co-planner, all the planets move in the same plane, in the same direction. Not that one is moving in this direction, another is moving in that direction. These are the initial clues on the basis of which a scientist would provide the hypothesis.

So, the first step is to look for the clues, and then you have to formulate the hypotheses. The second step is to formulate the hypotheses.

I will use 'hypotheses' that is a plural of hypothesis. When the clues are not sufficient to zoom on to one, then there would be a large number of hypotheses that would be able to satisfy the clues.

But when you formulate a hypothesis, there are certain immediate requirements. Unless you satisfy those requirements the hypothesis cannot be called scientific.

As I said, a hypothesis is also a guess. There has to be a difference between a scientific guess and a wild guess. Every wild guess is not scientific. So, there has to be some requirements to be fulfilled by the hypothesis in order to be called scientific, in order to attract the attention of other scientists. Otherwise people will ignore that guess.

The first requirement is: a hypothesis should be consistent with the clues. So, if you have initially located your clues, then the hypothesis should be such that, if that course of events did happen, then the clue that has been found should be true. So, the clue should be a logical consequence of the hypothesis.

When we did logic, we showed how to form the logical structures: a premise leads to the conclusion. If the hypothesis is true, the clues should be the conclusion. That way we have to formulate the hypothesis. So, initially we find clues, and on the basis of the clues we formulate the hypothesis. And when the clues are insufficient, there would be a large number of hypotheses that would be able to satisfy the clues.

I said that the hypothesis should be consistent with the clues, which means that the clues should be the logical consequence of a hypothesis. The logical structure should be very clear and that should be based on what is already known about the laws of nature.

So, a hypothesis, added to the known laws of nature, leads to a conclusion and the clue should be that, should satisfy that. Now in some situations, it is possible that the intermediate course of logical progression, which means the use of certain laws of nature, it may be that at some point of time a particular law of nature is not known and therefore, the hypothesis does not completely corroborate with or correlate with the clue. It is possible.

In that case, should we abandon the hypothesis? No. But we should have some doubt about it, we should look at it with some suspicion. But if the scientist very clearly states that this hypothesis will lead to the clue, if this is a law of nature, this is a missing piece of the jigsaw puzzle, and then people might look for that particular law of nature. If that law of nature is found, then the hypothesis will become a sound scientific hypothesis. Till then, it is just a provisional hypothesis.

So, a hypothesis should be consistent with the clues, but at any point of time it is possible that it may not be completely consistent, in the sense that some piece of information might be missing. Then, till that gap is filled, that hypothesis will be taken as a provisional hypothesis, not at par with the other hypotheses that may have found.

The second thing is that a hypothesis should be based on material processes and phenomena. This is the demand of materialism. We cannot hypothesize that a particular event happened because somebody willed it, it happened because some miracle, etc. Those things are out of the question when we deal with science.

So, a hypothesis should be based on material processes. For example, when people started forming hypotheses about the origin of the solar system, a very strong current was that the solar system had been created by somebody. But scientists said that, no, let us try to formulate hypotheses based on material processes and phenomena, and that is how we finally came to an understanding about the origin of the solar system.

The third requirement is that, it should have some testable predictions. We should have testable predictions. Testable means, as I have already said—if you have a premise, you apply valid logic, and you come to some conclusion—that one has to be testable. So, this structure is important. That premise, in that case, is the hypothesis and if you apply valid logic on that hypothesis, then you should be able to arrive at a conclusion. That would be the expectation from the hypothesis, and that expectation should be experimentally or observationally testable. That is a hard requirement. Why? Because, it should be falsifiable.

In the last class, I explained the concept of falsifiability. This is exactly where it is applied. When a hypothesis is proposed, it should be conceivable how to falsify that, under what observations a hypothesis or a theory would be proved false.

At this point notice one thing: that every good theory is restrictive; in some sense prohibitive. A good theory prohibits certain things from happening. The example of a good theory is Newton's theory. It prohibits things from going upwards, an apple will not go upward, it prohibits that. A simple falsifiability criterion would be that, if we ever see an apple going upward on its own, Newton's theory is definitely false.

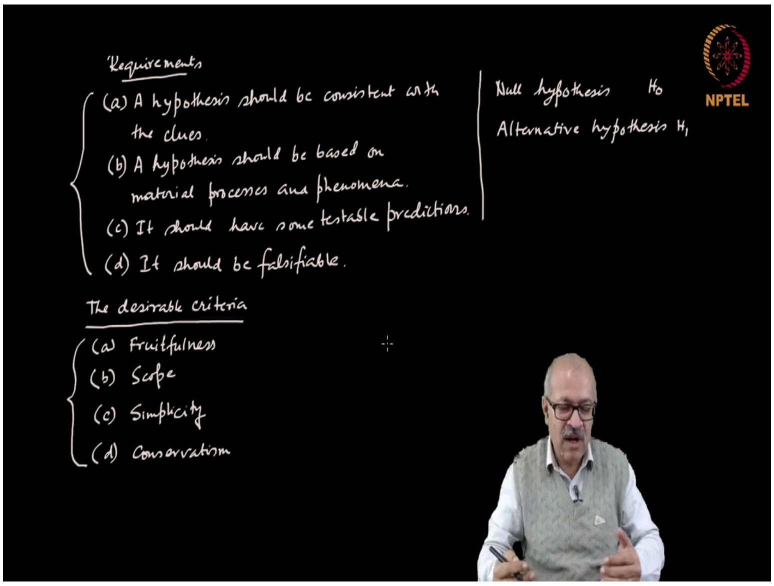
It prohibits any motion of a particle or an object under centre force in any fashion other than an ellipse. A circle is a special case of an ellipse. Nothing other than a conic section; it has to be a conic section. There are a few types of conic sections. They are possible, but nothing else. No other form of motion other than conic sections are possible. If you ever see any motion under a central force which is different from a conic section, then definitely the theory is false. So, these are easily testable, falsifiable theories.

Similarly, all hypotheses should be falsifiable. So, these four are the demands, are the requirements. Without these, a hypothesis will not stand as a scientific hypothesis. It will just be a wild guess.

But as I told you, in most situations you have a large number of hypotheses that would be conforming to the clues. So, we have to always use plural hypotheses, a large number of hypotheses. When you have a large number of hypotheses, each claiming to be explanation of some phenomenon, you have to test. When you do the test, we normally prioritise the hypothesis, which one I will test first, which is a good candidate and so on and so forth.

When we prioritize, we have to have some kind of desirable criteria. These are not the requirements, but some desirable criteria. Let us enumerate the desirable criteria.

(Refer Slide Time: 17:16)



The first one is, fruitfulness. ‘Fruitfulness’ means, the hypothesis was proposed in order to explain certain event, but if that hypothesis explains only that event and nothing else then it is not fruitful. But if it has a larger applicability, so that it helps to explain more things than for which it was primarily proposed, then it is fruitful.

So, if there are a number of hypotheses that a scientist has to check, the one that has the largest fruitfulness, would be the one that is preferred by the scientist. So, other things being equal, one prefers the hypothesis that is most fruitful. So, what is fruitfulness? It is the applicability or the usefulness of the hypothesis in explaining phenomena other than the one for which it was initially proposed.

The second is scope. ‘Scope’ is where every hypothesis will have a testable prediction, but if a hypothesis has many testable predictions and another hypothesis has a small number of testable predictions, then it is easier or more meaningful to test the one that has a larger number of testable predictions.

So, it should be such that it should be testable under varied circumstances, and ‘varied circumstances’ means, for different situations you should have different testable predictions different experiments should be able to test it. Then it has a larger scope. And other things being equal, we will prefer the one that has a larger scope.

The third is simplicity. 'Simplicity' means the following. In common parlance you might think that the one that looks very simple, that is simplicity. No, it is not so. A simple hypothesis is the one that makes the least number of *a priori* assumptions. The more *a priori* assumptions you make, the more complex it becomes, it loses simplicity. So, again, other things being equal, a scientist prefers a hypothesis that is simpler, in the sense that it makes lesser number of *a priori* assumptions.

Finally, we have to come to the last one, that is conservatism. 'Conservatism' means the following. So far, through millennia of human history, we have built a body of knowledge. There are certain theories that have been tested, that are now a part of the human body of knowledge. And, the hypothesis that you are proposing should be consistent with the existing body of knowledge that has been tested and there are reasons for believing that these are dependable.

When a hypothesis is compatible with the existing body of knowledge, it is called conservative. Now, it may be that under certain situation, a scientist proposes the hypothesis which is not consistent with the existing body of knowledge. When that happens, scientists will look at it with some suspicion because there is some bit of uncertainty regarding the proposition it makes. Only when that bit of additional knowledge is obtained, then we say that this is a scientifically stated hypothesis. We do not jettison it, we do not abandon it right away, but we take it with a pinch of salt. We look at it with some suspicion because it is not consistent with our existing body of knowledge.

But, we do not abandon it right away because it is possible that some part of our existing body of knowledge may be wrong, may need to be corrected and, may be that hypothesis, when ultimately tested, would lead to that correction. Since that possibility exists, we do not outright reject such a hypothesis.

So, these are the mandatory requirements when proposing a hypothesis and these are the desirable criteria when proposing a hypothesis. Now as I said, in any given situation, there would be a large number of hypotheses, and then scientists will go ahead to test with the specific objective of rejecting the wrong ones, not proving the one that is correct because that cannot be done.

So, the idea is rejecting the wrong ones. All scientific experiments are carried out with the express objective of rejecting the wrong ideas. Now, whenever we propose a hypothesis, we always propose two hypotheses: one is called the ‘null hypothesis’ and the other is called ‘alternate hypothesis’.

If I am talking about a particular phenomenon, the null hypothesis would say the phenomenon does not happen, and alternative hypothesis will say the phenomenon happens. If I am talking about the cause of certain event, a particular cause, the null hypothesis will say that is not the particular cause, and the alternative will say that is the cause.

So, that way we have to formulate the null hypothesis and the alternative hypothesis. The null hypothesis is normally called H0 and this one is called H1. In some books you will find null hypothesis is given the symbol H and alternative H prime. There are various notations, but I prefer to use the notation of H0 and H1.

So, for every hypothesis there will be a null, there will be an alternative. And, when we actually test, we actually test the correctness of the null hypothesis. I will come to that when we come to hypothesis testing. So, ultimately we formulate the hypothesis and then we test the hypothesis.