EPISODE 6

# OBSERVATIONS IN SCIENTIFIC INQUIRY

*K P Mohanan and Tara Mohanan*

## 6.1 Introduction

As was made clear in the previous episodes, what distinguishes scientific inquiry from the other forms of inquiry is its empirical grounding — the commitment to seek knowledge by observing the physical, biological, mental, and societal phenomena in the world around us.

In Episode 5, we gave an outline of theoretical science that can be pursued without experiments (whether in a laboratory or outside), and without the need for mathematics, either for the processing of numerical data or for the modelling in terms of equations. To complete the picture of scientific inquiry, we will now turn to what we left out in Episode 5.

## 6.2 Observational Inquiry

Let us go back to Fig. 2 in Episode 5, repeated below as Fig. 1:

| | |
|---|---|
| Premises of a scientific theory | **(EXPLANANS)** |
| ↓ (*derivation*) | |
| Conclusions derived from the premises | **(PREDICTION)** |
| ↕ (*match*) | |
| What is puzzling in our observations<br>Observational Generalisations | **(EXPLANANDUM)** |
| ↑ (derivation) | |
| Patterns in our sample | **(OBSERVATIONS)** |

***FIGURE 1***

To identify or detect patterns in samples, we need to first gather samples. A ***sample*** is a collection of ***observational reports***. We detect patterns in samples and then ***generalise*** them to the population or to the category.

The relation between observational reports (the raw ***data***) and ***observational generalisations*** come under ***observational inquiry***. In many cases, the data that serve as ***evidence*** for a claim are the results of experiments. For instance, take the question:

> Is the particular drug an effective cure for diabetes?

An investigation of the question calls for an experiment in which, taking a population of diabetes patients, a sample is given the drug, and those outside the sample are not.

Even though many experiments are done in a laboratory setting, not all of them need that setting. Some, called 'field experiments,' are done outside the laboratory. [NOTE: See the Wikipedia entry on Field Experiments at https://en.wikipedia.org/wiki/Field_experiment ]

There are also situations that do not lend themselves to experimentation. The discipline of astronomy, for instance, came out of the meticulous non-experimental observation of the behaviour of stars and planets. To prove that the earth would move in a straight line outside the gravitational field of the sun, we cannot move the sun away from the solar system to observe the motion of the earth. Nor can we change the axis of the rotation of the earth to find out what happens if the axis is perpendicular to the plane of revolution.

In sum, scientific inquiry calls for the following types of observational inquiry:

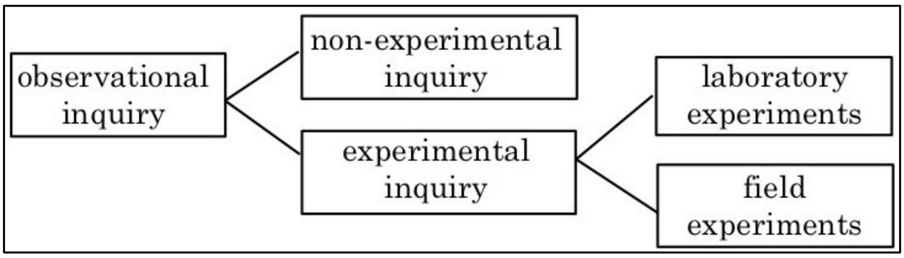


***FIGURE 2***

Along another dimension of observational inquiry is the distinction between the quantitative and the qualitative. The stereotype of data in many courses on research methodology in the sciences is quantitative data, that is, data coded in numbers, subject to the arithmetic operations of addition, subtraction, multiplication, and division. This approach often ignores the need for qualitative or categorial data. When we make the observational report, "Zeno is five feet eight inches tall," we are stating it in measurable numerical quantities. But when we say, "Zeno is male," we are stating the report in terms of category membership, which is qualitative, not quantitative. Likewise, when we say, "Insects have six legs," even though we are referring to a number, it is not subject to arithmetic operations. It makes no sense to say that the mean number of legs of arthropods (a taxon that includes insects and spiders) is seven. Nor does it make sense to do a statistical research project on the number of hearts in the animal kingdom.

Given the above remarks, we expand Fig. 2 as Fig. 3:

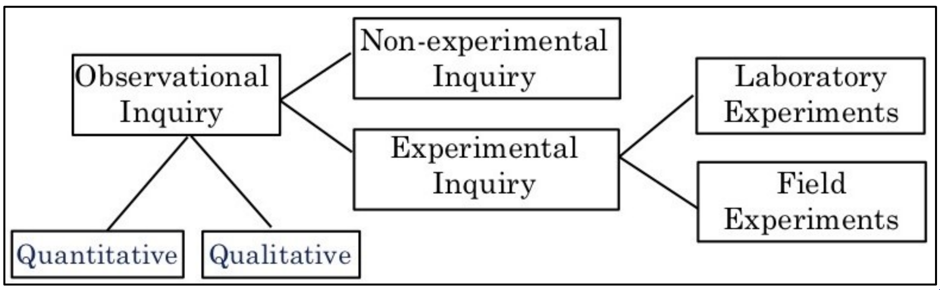


***FIGURE 3***

[NOTE: We have not said anything about inductive reasoning using numerically coded quantitative data. This comes under under ***statistics***, which makes a distinction between ***descriptive statistics***, which describes patterns in the sample, and ***inferential statistics***, which generalizes from the sample to the population.]

The process of observational inquiry involves five stages:

A. The formulation of a ***question*** to investigate
B. Consideration of how to gather ***evidence relevant to the question*** (e.g., experiment design)
C. ***Gathering*** the data (e.g., implementation of the experiment)
D. Arriving at an ***answer*** to the question on the basis of the data, and
E. Critical evaluation of the sampling and the ***reasoning*** from the data to the conclusion.

We now turn to a brief discussion of the general considerations in A, B, D and E in experimental inquiry.

## 6.3 Experimental Inquiry

### *6.3.1 Glass Shattering*

Suppose you wish to investigate the question of what causes things to break. The first step would be to clarify what you mean by 'break' in this question. Do you mean the breaking of a whole into two pieces, as it happens when you bend a dry twig or when you pull a sewing thread in opposite directions, or do you mean the breaking of a whole into many pieces, as in the case of a glass bowl shattering when it falls from a table and hits the floor. Let us pick the latter concept and rephrase our question "What causes things to shatter?"

Let us take an example. Suppose Zeno is sitting across you at a table. He accidentally hits a glass bowl on the table. It falls from the table, hits the floor, and shatters. In a letter to a friend, you report what you have observed as follows:

| | |
|---|---|
| Observational Report: | Zeno accidentally hit a glass bowl on the table.<br>The bowl fell to the floor.<br>It broke into a number of fragments. |

What caused the glass bowl to shatter? We might postulate one of the following reasons:

1. If a person hits a glass bowl, it shatters.
2. If a glass bowl falls to the floor, it shatters.
3. If a glass bowl collides with the floor, it shatters.

These are formulations of correlations between two events. They say nothing about causation. If we bring in causation, the formulations would be:

4. Hitting a glass bowl causes it to shatter.
5. Falling from a table to the floor, causes a glass bowl to shatter.
6. Colliding with the floor causes a glass bowl to shatter.

Let us examine each of these carefully.

a. Suppose a person accidentally hits a glass bowl on a table. It moves to a different place on the table. Do you think the bowl would shatter?

b. Suppose a person hits a glass bowl on a table with a hammer? It does not fall to the floor. Do you think the bowl would shatter?

c. Suppose the top of the table is one foot from the floor. A person accidentally hits a glass bowl on a table. It falls to the floor. The floor has a thick carpet on it. Do you think the bowl would shatter?

d. Suppose the glass bowl is made of glass, two inches thick. In scenarios (a)-(c), would the bowl shatter?

e. Suppose the bowl is made of clay, not glass. In scenarios (a)-(c), would the clay bowl shatter?

Think about issues of this kind (use your imagination combined with your prior experience). On the basis of your reflection, would you be tempted to revise the statements in (4)-(6), and perhaps combine multiple causal factors? Come up with an explanation on the basis of your thinking. In your explanandum (what you wish to explain), include why a bowl made of glass or clay shatters, but one made of metal, rubber, or plastic does not.

Now design an experiment (or a combination of experiments) to test your explanation.

### *6.3.2 Hotness of Chillies*

Miko and her younger brother Jomo (ages 12 and 10 respectively), got some hot chilli peppers, and planted their seeds under a tree in their backyard. When the chillies were ready to be eaten, Miko was disappointed because they were not hot. Why were the chillies not hot?

Jomo's suggestion was that the plants were not watered enough. He wrote down his *hypothesis*:

J: If the plants don't get adequate water, their chillies won't be hot.

Miko disagreed. She wrote down her hypothesis:

M: If the plants don't get adequate sunlight, their chillies won't be hot.

Each of these hypotheses could be either true or false. So, Miko and Jomo decided to conduct two experiments, one to test J, and another to test M.

In each of the two hypotheses, there are two **VARIABLES**: In J, they are the *hotness of chillies* and *water*, and in M, they are *hotness of chillies* and *sunlight*. In J, the hotness of the chillies depends on the amount of water the plant receives, and in M, the amount of sunlight it receives.

To design an experiment to test the hypotheses, they needed four groups of chilli plants. The four groups, A, B, C, and D, would differ in the amount of water and sunlight they receive. The differences would be as follows:

| WATER \ SUNLIGHT | Adequate | Inadequate |
|---|---|---|
| Adequate | Group A | Group B |
| Inadequate | Group C | Group D |

To interpret this two-by-two matrix with ease, we look at the relation between the four different groups and the two factors, water and sunlight:

Group A: gets adequate sunlight and adequate water.

Group B: gets inadequate sunlight and adequate water.

Group C: gets adequate sunlight and inadequate water.

Group D: gets inadequate sunlight and inadequate water.

For Hypothesis J, the plants in groups C and D would be the **EXPERIMENTAL GROUPS** (receiving inadequate amounts of water), while A and B would be the **CONTROL GROUPS** (receiving adequate amounts of water).

If the chillies that receive inadequate amounts of water (C and D) are consistently less hot compared to those from Groups A and B, then we can say that the causal factor is water, and that Hypothesis J is correct.

For Hypothesis M, the plants in groups B and D (receiving inadequate amounts of sunlight) would be the experimental groups, and groups A and C (receiving adequate amounts of sunlight) would be the control groups.

If the chillies from plants that receive inadequate amounts of sunlight (B and D) are consistently less hot compared to those that receive an adequate amount of sunlight (A and C), then hypothesis M is correct.

**Learning point**:
To maximize the benefits from our inquiry, it is important to formulate our hypothesis clearly and precisely.

This design says nothing about the nature of the soil as a possible factor in the hotness of the chillies. Suppose the soil used to pot some of the chilli plants contained nutrients that increased the hotness of the chillies. If so, our conclusion about the factors influencing the hotness of the chillies would be unjustified, because the relevant variable here is the soil and its contents, not water or sunlight. Think about how you would eliminate the effects of such variables if you were doing this experiment.

Here is another factor that one needs to pay attention to when interpreting experimental results. Suppose after the plants yield chillies, both Jomo and Miko taste the chillies. Jomo finds the chillies from groups A and B to be hotter than those from groups C and D. Miko finds the chillies from groups A and C to be hotter than those from groups B and D. They are surprised that their findings are contradictory. Could it be that the subjective nature of their taste of the chillies, combined with each researcher's eagerness to support his/her own hypothesis, resulted in the differences in their observational reports? How would you design a strategy to satisfactorily avoid the problem? We leave you with that question to chew on.

## 6.4 Non-Experimental Observational Inquiry

Let us now look at an example of scientific thinking in everyday life.

### *6.4.1 A Controversy at the Breakfast Table*

*[Ammu (A), her mother, Tara (T), and father, Mohan (M) are tasting sugar-coated banana chips from a package.]*

M: (bites into a chip) Hey, these chips are great, they're really crisp!

T: (bites into a chip) They're just hard, not crisp.

A: (bites into a chip) Hard!

T Do we have different meanings for the word 'crisp'?

A: What do you mean, different meanings? We're all speaking English, aren't we?

T Let's see. Try this. (holds up her right hand) How many fingers do I have on this hand?

A: Five, of course, silly!

T You know, Amz, the word 'finger' has two meanings. One is your meaning: a digit on the hand. It includes the thumb. The other meaning is, the digits on the hand other than the thumb. So if my meaning of the word 'finger' excludes the thumb, then I have four fingers.

A: Oh, right, I get it.

M: You know, many debates, even among academics, arise from the same word being interpreted in two different ways.

T Listen, there may be another explanation for our different judgments on the chips. May be our sensory experience of them was different. Perhaps because you have stronger teeth…

M: Right! Remember the chutney we had yesterday? You thought it was exploding hot, and I thought it was bland. It was the same chutney, but our *experiences* of it were different.

T: Take another example. Imagine Jay and Viru in the same room. Jay finds the room cold, Viru finds it warm. The *objective reality* of the temperature in the room is the same, right? But their *experiences* are different. The difference would make perfect sense if Jay has a fever and Viru doesn't.

A: (*tries another chip*) HEY! I just got a crisp one. They're not all hard.

T I too got a few crisp ones, and a couple of hard ones.

A: Papa, you try some more too.

T: So based on one chip each, *you* concluded that they were all crisp, and *we* concluded that they were all hard… Extremely defective sampling!

M: Ammu, what Amma is talking about is called ***anecdotal evidence***. Exactly like meeting a six-foot tall Australian and a five-foot tall German and concluding that Australians are taller than Germans. We make such hasty conclusions all the time.

### 6.4.2 *Looking for an Explanation*

*[Why were some chips crisp and others hard? Was the difference systematic? Or was it random? Was there a pattern in the difference? Human beings have a compulsive need for making sense of their experience.]*

T Look, the sugarcoating on the chips is not even. Looks like the ones with more sugarcoating are the hard ones, and those with less sugar are crisp.

M: Ah! So we have some data. The packet has both crisp and hard chips. And we now have an *intuitive explanation* for the data: the difference in hardness vs. crispness depends on the amount of sugarcoating.

A: But how do we know how much sugarcoating there is? We can't judge the amount by just looking. How do we measure it?

T You're right, Amz, we have no way of measuring the amount of sugar on the chips.

M: That makes the claim *untestable*. Every time we find a hard chip, we can claim that it has a lot of sugarcoating, and every time we find a crisp chip, we can claim that it had less sugarcoating. 'Untestable' means that there is no way to find out if it is true or false. Suppose I tell you that a headache is caused by headache demons. Whenever anyone has a headache, I say, "Ah, a headache demon has possessed him!" And when the headache is gone, I say, "Oh, the headache demon has left him."

T: So our claim cannot be shown to be false, it is not *falsifiable*, unless we can find an independent test.

M: Otherwise, nothing would stop me from cooking up mythologies as explanations.

T: Right. For any theory, if there is no way to even imagine what kind of a situation or what kind of evidence would tell us if it is false, the theory is unfalsifiable.

### 6.4.3 *Developing an Intuition into a Testable Theory*

A: So, saying that sugarcoating causes hardness is like saying that headache demons cause headaches?

T: Unless there is some independent way of observing sugarcoating. Also, suppose you find a chip with a lot of sugar coating, and it is not hard. That would be a ***counterexample*** to the claim that the hardness of the chip has to do with the amount of sugarcoating.

A: What's a counterexample?

M: "A creature without wings cannot fly," right? Now suppose you found a living thing, say X, that doesn't have wings but can fly. Our principle says, X can't fly. So X's ability to fly proves the principle to be false. X is a counterexample to the principle.

T Here is a different situation. Suppose you found a living thing that has wings but cannot fly, like ostriches and penguins. What does the principle say about them?

A: Let me think. The principle says that living things without wings cannot fly. It doesn't say that ALL living things with wings can fly. So it doesn't say anything about ostriches and penguins, does it?

M: Excellent. So it doesn't cover ostriches and penguins. It doesn't tell us why they can't fly.

A: I guess not.

T So this example wouldn't be a counterexample. It's simply that it goes against our expectation that anything with wings would be able to fly. That doesn't make the principle wrong, it's just that there's something missing, or incomplete. We just have to find an explanation for why ostriches and penguins can't fly.

M: Yes, the principle is inadequate by itself. This is why a theory is mostly not just a single statement, it is a set of statements that together provide an explanation.

Coming back to, "Only living things with wings can fly," this statement is admissible in science, because it is empirical, and falsifiable. That it is false is a different matter. By the way, it's important to remember that 'false' and 'falsifiable' are not the same. Something can be false, and yet unfalsifiable.

### *6.4.4 Testing the Predictions of the Theory*

A: Phew! Now I want to eat some more chips. *(Eats some, and examines several others carefully.)* Look at these, some chips are more dark brown. *(eats a few)* Hey, the dark ones are all hard, and the light ones are all crisp!

T: Hmmm. I wonder if that has something to do with the amount of sugar. Could it be that the ones with thicker sugarcoating are darker? If that's systematically so, the unobservable sugarcoating has found an observable *correlate*: its color…!

M: This looks like the beginning of a banana-chip-crispness theory. Can we crystallize the intuitions so far into an explicit testable theory? Amz, what are the patterns we've found?

A: The harder chips are darker.

T What else?

A: The darker chips have more sugarcoating.

T Anything else?

A: The sweeter chips have more sugarcoating.

M: Let me write down these intuitions. Our theory has three hypotheses or laws:

> (1) Theory 1
> a. The darker the chip, the more the sugarcoating.
> b. The more the sugarcoating, the harder the chip.
> c. The less the hardness, the greater the crispness of the chip.

M: These are *precise* and *explicit* statements. The advantage of stating an intuition as such a set of propositions is that we can *test* our theory by checking the predictions that follow from it.

T Then we can check if our intuitive understanding is on the right track, or if we are completely wrong.

### *6.4.4.1 Two Meanings of 'Prediction'*

A: What do you mean, check the predictions? You mean the theory can tell you about the future? (Points to the 'horoscopes' section in the Sunday magazine on the dining table.) Look at this. They are making predictions about what's going to happen next week for everyone… So is that a scientific theory?

T No, Amz, the predictions of a theory are not the same as predictions about the future.

A: But when you say, "I predict that it will rain today," and it actually rains, you say that the prediction came true. It's like a prophecy, right?

T Yes, but the word 'prediction' has a different meaning here. In the context of scientific theories, by prediction, we mean the observable state of affairs that follow as a logical consequence of the theory.

A: Amma!!! What does that mean?

M: See if this example from linguistics helps. In the English sentence, *John admires him*, the pronoun *him* cannot refer to John. Agreed? Take the principle: "An object pronoun cannot take the subject of the same clause as its antecedent." This principle correctly predicts that *him* cannot refer to John. This is a logical consequence of the principle.

A: Oh, okay, I think I get it…

## 6.5. Towards a Theory of Crispness in Sugarcoated Banana Chips

M: So now, going back to what we started writing. We said:

> (1) Theory 1
> a. The darker the chip, the more the sugarcoating.
> b. The more the sugarcoating, the harder the chip.
> c. The less the hardness, the greater the crispness of the chip.

If we put together the first two statements in our theory: "the darker the chip, the more the sugarcoating, and the more the sugarcoating, the harder the chip," we get a prediction. Can you state it?

A: The darker the chip, the harder it will be.

T Excellent. And we found this prediction to be correct. The dark chips were hard. And the darker the chip, the greater its hardness.

M: So premises (1a) and (1b) together predict a ***correlation*** between the observable color of the chips and their hardness. (writes)

> (2) Prediction 1: The darker the chip, the greater its hardness.

A: That's right. There are no dark chips that are crisp.

T Given the second and third statements, we expect that the lighter a chip, the greater its crispness.

M: That's right, this is an expectation, not a prediction. (writes)

> (3) Expectation 1: The lighter a chip, the greater its crispness.

T Let's check if this is true.

A: Uh-oh, I just had a light chip that wasn't crisp at all, it was actually hard. So it's not true that the lighter, the crisper.

M: Don't be disappointed by the mismatch between the expectation arising from the theory and the actual observations. One of the qualities of a good theory is that it points us in the direction of new and interesting observations.

May be the dark chips are systematically hard, but the light ones show no pattern, they're randomly hard or crisp.

T M-kmmm, that's a cop-out. It can't just be random like that. Let's see. What about the thickness of the banana slices themselves? *(examines more chips, and is excited, and groups some light chips into two sections.)* Look at the two mounds of light chips, Amz. Do you see a difference?

A: Cool! You're right, Ma, I just ate a thick one and a thin one, both light ones. The thick one was hard, and the thin one was crisp. (eats some more) The thick ones are all harder, the thinner ones are all crisp!

M: Excellent! We're getting somewhere with our theory. Let's modify it by adding the new hypothesis. So this is what we now have. (writes)

> (4) Theory 2
> a. The darker the chip, the greater the amount of sugarcoating.
> b. the greater the amount of sugarcoating, the harder the chip.
> c. The less the hardness, the greater the crispness of the chip.
> d. The thicker a chip, the greater its hardness.

M: Now this yields an expectation: A light chip that is thin will be crisp. (writes)

> (5) Expectation 2: A light chip that is thin will be crisp.

A: And that is correct! Yippy, I was the one who figured out that pattern!

T (smiles at her) Actually, shouldn't this expectation extend to the dark chips too, because of our fourth statement? Among the dark chips too, the thinner ones should be crisper, and the thicker ones harder…

A: (examines dark chips) Yes, that's true! The thick dark ones are extra hard!

T Nice! So the theory makes falsifiable predictions. Thick-dark-crisp chips, or thin–light-hard ones, would be counterexamples to the theory. We haven't found any in our data. If we do, we'll have to either modify the theory, or abandon it.

M: See, Amz, we wouldn't even have thought of looking for the relation between thickness and crispness before constructing the theory. So, we formed the falsifiable theory based on the patterns we saw; the theory lit up new patterns by making a new prediction; we verified that prediction from new data, and the verification has increased our *confidence* in the theory.

It's only because of stating our understanding as explicit, logically connected, testable hypotheses that we explored factors and domains that we wouldn't have considered otherwise.

This is exactly how we construct scientific theories. Our theory of banana chips is *scientific* because it makes falsifiable predictions on observable reality.

## 6.5 Concluding Remarks

Episode 6 outlines what is common to mathematical and scientific inquiries, and identifies what makes scientific inquiry distinctive. Unlike mathematical theories, scientific theories have an empirical grounding, in that they seek to predict and explain what we observe in the world outside. This is diagrammatically expressed in terms of diagram 1 on page 2 and diagram 3 on page 3.

Having clarified that, the rest of episode 6 provides detailed examples of experimental inquiry and non-experimental inquiry. Our last example, that of scientific inquiry in everyday life (the theory of crispness in banana chips) integrates the discussion of theoretical inquiry in episode 5 and observational inquiry in Episode 6.

[NOTE: Perhaps a good sequel to the theory of crispness in banana chips would be our learning resource "On Statistical Thinking: Populations, Samples, and Representativeness," at (https://thinq-website.s3.ap-southeast-1.amazonaws.com/other/articles/3+Statistical+Thinking+in+Kitchen-Diced+Potatoes-2009+copy.pdf).

We hope that episodes 3-6 help you develop a deep understanding of two of the central the knowledge systems in higher education, namely, mathematical inquiry and scientific inquiry. In later episodes, we will take you on a similar journey through the humanities.

## Acknowledgements

We are indebted to Wee Lian Hee, Rahul Kulkarni, and M G Subramanian for comments on a draft of this article.