

On Evidence

Table of content

1. Introduction	1
1.1. <i>Initial examples</i>	1
2. Evidence in the disciplines	4
2.1. <i>Example 1: Mathematics</i>	4
2.2. <i>Example 2: Astronomy – Evidence for the existence of Jupiter’s moons</i>	5
2.3. <i>Example 3: Astronomy – Evidence that the Earth is the centre of our planetary system</i> ..	7
2.4. <i>Example 4: Physics – Evidence for the non-existence of the aether</i>	9
2.5. <i>Example 5: Astronomy – Evidence that Pluto exists</i>	9
2.6. <i>Example 6: Astronomy – Halley’s comet</i>	12
2.7. <i>Example 7: Chemistry – The theory of Phlogiston</i>	14
2.8. <i>Exercises</i>	21
3. A commentary on evidence	25
3.1. <i>Examples of evidence which supports a false hypothesis</i>	26
3.2. <i>Examples of necessary and sufficient evidence: Multiple lines of independent evidence</i>	31
3.3. <i>Other examples</i>	37
3.4. <i>Exercises</i>	39
4. How do we find evidence for things we cannot see?	41
5. A complete example	45

1. Introduction

Here we will look at the theme of evidence in science. Evidence is all about “evidence in favour of a hypothesis or theory”. Such evidence requires data which comes from performing experiments. Experimentation, data and evidence which we ourselves gather, on the basis of our own work, falls under the category of primary research. We then become the primary source for other people. Since the IPM is not about us conducting our own research we will have to rely on experimentation, data and evidence gathered by others. If we are reading papers from people who have gathered their own evidence on a subject then we consider these papers *primary* sources. If we are reading papers from people who have not gathered their own evidence but are referring to other peoples’ evidence then we consider these papers *secondary* sources.

1.1. Initial examples

In terms of applied/physical sciences (where I also include statistics and applied maths) we need to ask the question, What is evidence? For example, what is the evidence that the following exist:

- the moon: The evidence is that we can see it with our own eyes. People since ancient times have known that it exists simply because
 - it is close enough and large enough to be seen directly,
 - everybody shares the same ability to be able to see it, so everyone can confirm other peoples’ observation of it.
- the moon’s of Jupiter: These can be seen with a telescope, i.e. an optical device which extends the ability of our own eyes. So the evidence that they exist as moons, and not stars or separate planets, is again visual.

Note, however, that when Galileo first discovered them (in January 1610) there was debate among people as to whether or not they really existed. This is because the telescope Galileo was using was homemade and of low quality. He made the lens himself, and the technology in his day was such that the lenses were not as perfect made as they can now be made. Problems such as distortion of image, false colour effects, impurities in the glass, etc., could be present.

So how was Galileo able to determine that these were moons of Jupiter and not specks of dust, impurities in, or distortions of, the lens? Also, how did he know that they were not

stars? Because he saw that these objects changed positions quickly over a few weeks, i.e. they moved quickly (which stars do not do). Secondly, he noticed that as Jupiter moved across the sky so the objects moved with Jupiter. Thirdly he noticed that the objects changed position with respect to each other and with respect to Jupiter. Fourthly, he noticed that the objects sometimes moved in one direction and sometimes moved in the opposite direction. As such he concluded that these objects were revolving about Jupiter. In other words, they were moons. This discovery provided evidence in support of the Copernican system and showed that everything did not revolve around the Earth.

- an electron: This cannot be seen with the naked eye. So what evidence is there for its existence? Well, originally (in the late 1800s) it was detected as the deflection of a beam of energy created inside a glass tube, with this beam being subjected to an electric field. If the beam deviated from its straight line path then this indicated the presence of a negatively charged particle. The beam did deviate, and the electron was discovered. More on this later. Today electrons can be seen with computer enhanced imaging machines (optical microscopes are not powerful enough for this);
- a magnetic field: This cannot be seen at all optically. The evidence for magnetic fields lies with the effects they have on other objects. For example, magnetic fields attract certain metals, they attract opposite poles and repel like pole, etc.
- The current and voltage in a circuit: Again these cannot be seen with any optical instrument. And again, the evidence of their existence can only be found by the effect they have on other objects, namely that currents can turn on light bulbs (the higher the current the greater the amount of electrons flowing through the light bulb, hence the brighter it is), and that voltage represents the strength of the current, which is a force which can be seen by how much the needle on a voltmeter is deflected (the higher the voltage the greater the strength of the current).
- Heat: This cannot be seen optically. The evidence for it comes from the effect it has on objects. For example, heat creates warmth, heat burns, heat melts ices, heat boils water, etc. The distortion of the air just above a flame is evidence of heat but not direct observation of heat. It is observation of the effect of heat on the air, by the agitation of air particles that cause the air to shimmer or wave around which is evidence of heat.

- Black holes: This cannot be seen optically (yet? Or has an image of a black hole now been made?). The evidence for black holes lies in the creation of accretion disks, the bending of light as light passes close to a black hole, etc.

Exercises

Consider the following example, and then do the exercises that come after them.

“I have evidence that plate-tectonics explains earthquakes, volcanic eruption and mountain ranges”.

The evidence is that earthquakes are caused when one continental plate slides under another plate, which causes huge friction and energy to be released via an earthquake; The evidence is that volcanic eruptions occur because huge temperatures occur between the plate below and the plate above which melts the rock into magma. The huge temperature cause huge pressures whereby the magma can no longer stay trapped between the plates and therefore explodes up through the Earth to produce volcanic eruptions; Secondly, the evidence is that mountain ranges are caused when two plates collide without either one sliding underneath the other. Huge amounts of rock are therefore lifted up to form mountain ranges (for example the Himalayas were created when the Indian plate collided with the Eurasian plate, and the Alps were created when African plate collided with the European plate).

- 1) “I have evidence that my software works correctly”. What type of evidence would you need to say this?
- 2) “I have evidence that my antivirus software blocks all current known viruses”. What type of evidence would you need to say this?
- 3) “I have evidence that my information security protocols work correctly”. What type of evidence would you need to say this?”
- 4) “I have evidence that the conclusions based on my big-data analysis are correct”. State some conclusions relevant to big data, then ask yourself what type of evidence you would need to say this.
- 5) “I have evidence that *<some outcomes, theory, practice or product, etc.>* works, is correct or does *<state what it is that is supposed to happen>*”. Fill in the comments in italics with aspects from your own discipline. What type of evidence would you need to say this?

2. Evidence in the disciplines

2.1. Example 1: Mathematics

Let us consider the three categories of statistics, pure mathematics and applied mathematics:

- a) When analysing large amounts of data statistics does not deal with trying to prove the truth of individual cases but only in the truth of trends or patterns of data as a whole. However, we do want to make conclusions about individual cases or groups. In this situation statistics deals with the probability that such individual case or groups are likely to occur. Hence statistical evidence is never about obtaining a definitive or certain result. Instead we speak of results which are significant, or results in which we have a degree of confidence.

Generally speaking, statistical evidence is the interpretation of data which has been organized, summarised and/or tested in particular (statistical) ways. There are two main ways to demonstrate statistical evidence:

- *Visual*: Histograms or bar charts are visual organisations of data from which we may draw conclusions about the nature of the data. These conclusions form the evidence we use to confirm/support (or deny) our hypothesis.
Other ways of visually organising data for statistical purposes include scatter diagrams, lines of best fit, and box plots. Each diagram has its own form of analysis associated with it, i.e the analysis required to produce said diagram (we will not go into the analyses in these notes, although they may appear in other C3 STEM notes), and each of these diagrams are used to draw conclusion designed to act as evidence in support of our hypothesis.
- *Analytical*: Specific tests (methods of analysis) exist for analysing data to see if an individual data point is significantly different from usual, or to see if a particular experiment has had a significant effect or not. The tests measure what are called levels of significance and intervals of confidence (all based on probabilities). In this situation statistical evidence is a number (derived from the data) that is used as justification, or evidence, for or against our hypothesis.

- b) In pure mathematics there is no such thing as data, hypothesis, experimentation, evidence, or theory. There are only definitions, theorems, proofs and examples. The “evidence” that a theorem is true is given by the algebras used, which is itself simply a logical, deductive system of analysis based on axioms and prior theorems (which

themselves have been proved to be true).

- c) The case of applied mathematics is slightly different to that of pure mathematics. In terms of the mathematics itself, everything mentioned in **b)** above applies. But applied mathematics is designed to model natural phenomena. Therefore, from the real world we collect data, form hypotheses, conduct experiments, develop theoretical models (our mathematical theories about the world) and gather evidence in favour of our theoretical models. In this case we will have evidence for or against a hypothesis. All the physical science uses mathematics in the sense of this being applied mathematics.

2.2. Example 2: Astronomy – Evidence for the existence of Jupiter’s moons

When Galileo built his telescope he was able to see what we now know to be the moons of Jupiter. But his lenses were quite poor quality. This probably means that there could have been a lot of chromatic (and other optical) aberration. So the claim that he saw moons around Jupiter was open to question. No wonder some people didn’t believe him when he said he had seen moons around Jupiter, since the small white dots which Galileo said were moons could easily have been due to optical effects (dirty lenses, chromatic aberrations, reflections internal to the lenses, etc.).

As an example, consider the photographs below. What do you see? How do you know the smaller white dots are not effects due to your telescope/binoculars, etc. (dirty lens or bad quality lens)? How do you know that these smaller white dots are not stars, or planets revolving around the Sun?





We can summarise the situation above using six categories taken from *Understanding scientific reasoning* (5th edition), R. Giere, J. Bickle, R. Mauldin, 2006, Thomson Wadsworth Publishing. The categories are Real world, Model, Prediction, Data, Negative evidence, Positive evidence:

- *Real world*: The real world phenomena under investigation are objects moving across the face of Jupiter.
- *Model*: The Ptolemaic model tells us that all objects in the sky move around the Earth.
- *Prediction*: If these objects are indeed moons around Jupiter they should always return to the same positions at set time intervals (because moons revolving around a planet usually come back to the same point after a set time period).
- *Data*: Direct visual evidence of the position of the white dots.
- *Positive evidence*: The fact that these small white objects kept returning to the same location at the expected time. This showed a regularity of behaviour which could not be explained away once all negative evidence had been sorted out.
- *Negative evidence*: The white dots could be specks of light, or chromatic aberration, or reflection internal to the lenses, or badly calibrated telescope, or any other effect due to the poor quality of the lenses. This would then mean that the white dots are simply artefacts of the instrument itself.
- *Model (updated)*: In hindsight we can now suggest that the behaviour of these objects fits with the Copernican theory of a Sun centred planetary system.

As Allan Franklin says (*Experiment right or wrong*, 1990) “[...] though one might believe, as Cremonini and other seventeenth-century Aristotelians are reported to have believed, that Galileo’s telescope could create specks of light, it is extremely unlikely that it could create

specks of light that behave as a planetary system with consistent motions and eclipses.” Allan Franklin goes on to add that “It would be even more astonishing if such **artefacts** obeyed Kepler’s third law [...]”. Although Kepler’s third law was not known to Galileo at the time, such a law provides an explanation for the regularity of these “objects” which could not in all probability be attributed to the creation of specks of light.”

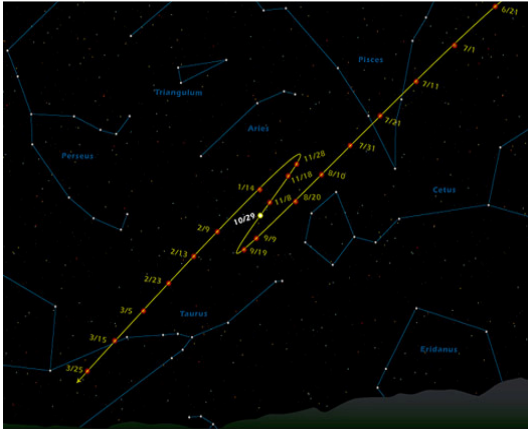
The truth is that there was always evidence for the existence of the moons of Jupiter. It only depended on whether we were able to conduct the necessary experiment which would allow us to see them. In this case looking at Jupiter through a telescope was the relevant experiment to perform. But such an experiment then depended on having the correct equipment, designed and built to a sufficient quality so as to see the moons. The objects that were observed were indeed the moons of Jupiter (i.e. the data was correct). It was then a case of interpreting the data we saw, and the interpretation was correct because these objects were revolving in a regular manner which could only be accounted for by these objects being moons.

2.3. Example 3: Astronomy – Evidence that the Earth is the centre of our planetary system.

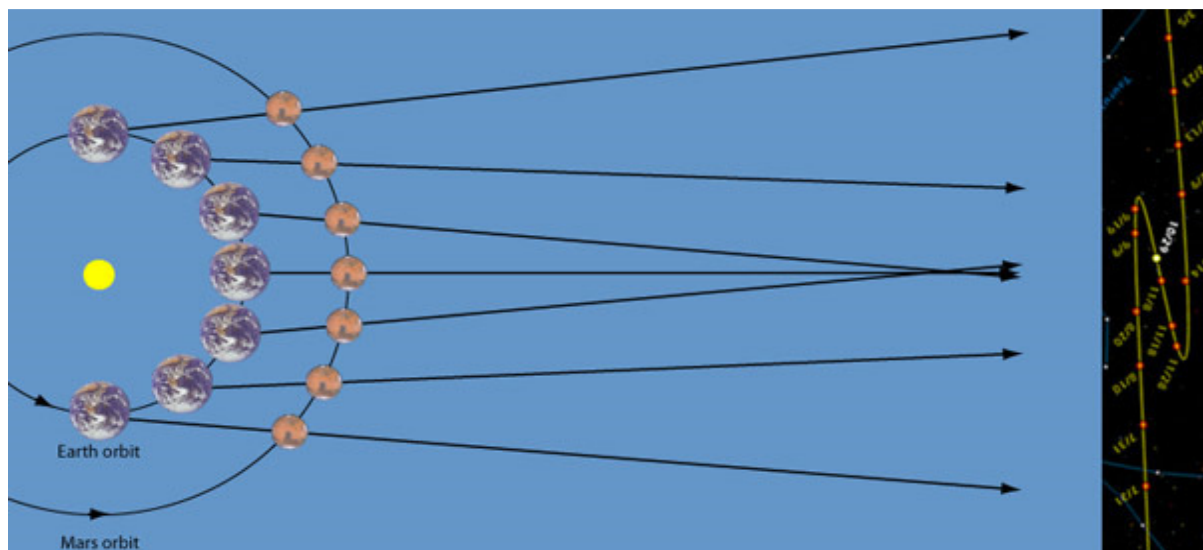
We believed that the earth was stationary with everything else revolving around us because we had some data (i.e. the motion of the planets and Sun) illustrating this. We saw these objects move relative to us on Earth so there seemed to be evidence that we were stationary and everything else revolved around us. We can summarise the situation above as follows:

- *Real world:* The real world phenomenon under investigation is the motion of the Sun, and planets around a stationary Earth.
- *Model:* Our model is that of the Ptolemaic Geocentric system whereby the Sun and the planets revolve around the Earth.
- *Prediction:* The sun, planets and stars revolve in the same direction around the earth in circular orbits and should return to positions previously seen.
- *Data:* The planets, stars and Sun.
- *Positive evidence:* All other planets, as well as the sun and the stars do move in the same direction around the Earth.
- *Negative evidence:* Mars sometimes moves “backwards” in the sky; Parallax: the (slight) shift in the apparent positions of stars due to the change in position of the earth.

In terms of the motion of Mars, people then saw something strange, a supposedly anomalous observation: at certain times every few years Mars started moving backwards in the sky as illustrated below.



How can this happen if everything moves around the Earth!? This retrograde motion cannot be explained if the Earth is at the centre of the system. It can only be explained if the Sun is at the centre of the system, with the planets revolving around the sun, as illustrated below.



Since it has always been True that the planets revolve around the Sun it is the case that there was always evidence for this fact. It only depended on whether we were able to conduct the necessary experiment which would allow us to deduce the correct explanation. In this case looking at the night sky over an extended period of time was the relevant experiment to perform. It was then a case of interpreting the data appropriately, and the interpretation was wrong because the motion of Mars would never behave as it does if the Earth was the centre of our planetary system.

2.4. Example 4: Physics – Evidence for the non-existence of the aether

There were two theories about the nature of light. It was either made of particles (Newton's theory) or it was a continuous wave (Huygen's theory). But it was known that waves could only travel in a medium/substance, not in a vacuum. So people came up with the idea of the aether, the medium through which light could travel. Aether was supposed to be a substance which was weightless, transparent, frictionless and which pervaded all of space.

In 1887 two physicists, Albert Michelson and Edward Morley, conducted an experiment to confirm the existence of aether. This would then provide evidence that light was a wave form. The following comes from "Doing Much About Nothing", Allan Franklin (2004), *Archive for History of Exact Sciences*, May 2004, Vol. 58, No. 4 (May 2004), pp 323-379:

"In the Michelson-Morley experiment, one of the most famous experiments in modern physics, the experimenters expected to detect a [...] change in an interference pattern produced by two light beams, one travelling parallel to the velocity of the earth relative to the ether, and the other perpendicular to that velocity, [...] They found no such [change]."

We can summarise the situation above as follows

- *Real world*: The interference patterns produced by two light beams.
- *Model*: Aether, this being the supposed medium through which light was travelling.
- *Prediction*: A change in an interference pattern produced by two light beams.
- *Data*: The data being sought was the interference pattern between two light beams.
- *Positive evidence*: None.
- *Negative evidence*: The expected interference pattern was not found.

2.5. Example 5: Astronomy – Evidence that Pluto exists

This passage is taken from <https://en.wikipedia.org/wiki/Pluto>

In the 1840s, Urbain Le Verrier used Newtonian mechanics to predict the position of the then-undiscovered planet Neptune after analysing perturbations in the orbit of Uranus.^[12] Subsequent observations of Neptune in the late 19th

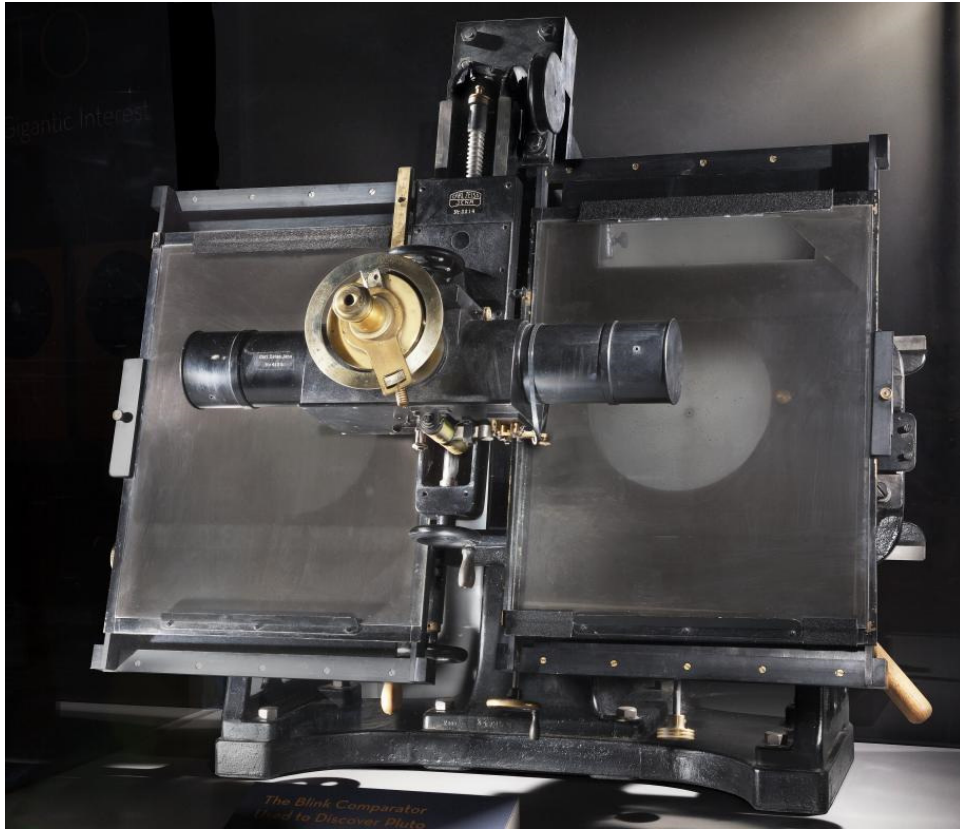
century led astronomers to speculate that Uranus's orbit was being disturbed by another planet besides Neptune.

In 1906, Percival Lowell—a wealthy Bostonian who had founded Lowell Observatory in Flagstaff, Arizona, in 1894—started an extensive project in search of a possible ninth planet, which he termed "Planet X".^[13] By 1909, Lowell and William H. Pickering had suggested several possible celestial coordinates for such a planet. Lowell and his observatory conducted his search until his death in 1916, but to no avail. Unknown to Lowell, his surveys had captured two faint images of Pluto on March 19 and April 7, 1915, but they were not recognized for what they were. There are fourteen other known precovery observations, with the oldest made by the Yerkes Observatory on August 20, 1909.

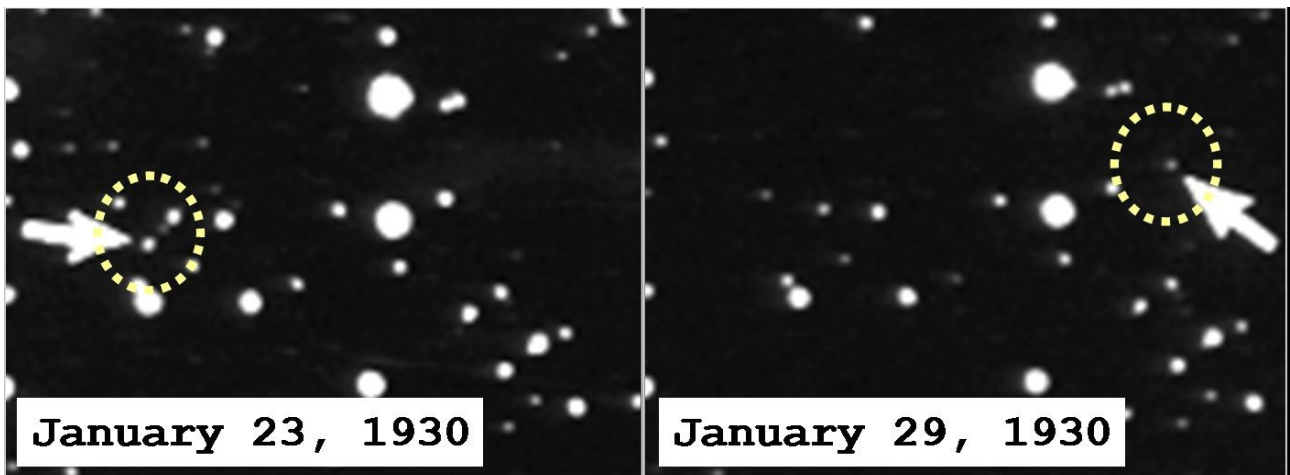
Percival's widow, Constance Lowell, entered into a ten-year legal battle with the Lowell Observatory over her husband's legacy, and the search for Planet X did not resume until 1929. Vesto Melvin Slipher, the observatory director, gave the job of locating Planet X to 23-year-old Clyde Tombaugh, who had just arrived at the observatory after Slipher had been impressed by a sample of his astronomical drawings.

Tombaugh's task was to systematically image the night sky in pairs of photographs, then examine each pair and determine whether any objects had shifted position. Using a blink comparator [see photo bellow], he rapidly shifted back and forth between views of each of the plates to create the illusion of movement of any objects that had changed position or appearance between photographs. On February 18, 1930, after nearly a year of searching, Tombaugh discovered a possible moving object on photographic plates taken on January 23 and 29 [see photo below]. A lesser-quality photograph taken on January 21 helped confirm the movement. After the observatory obtained further confirmatory photographs, news of the discovery was telegraphed to the Harvard College Observatory on March 13, 1930.

Below is a photograph of the blink comparator used to discover Pluto.



Below are two photos taken using the blink comparator showing the difference in position of Pluto over 6 days.



- *Real world:* The real-world phenomenon under investigation is the motion of an unknown object across the background of stars.
- *Model:* The model was the Newtonian model of gravitation which predicted a planet beyond Neptune to account for the discrepancies in Neptune's orbit.
- *Prediction:* The prediction was that there would be a planet in the region of the sky as predicted by Newton's theory of gravity.
- *Data:* A change in the position of an object as seen through the blink comparator.
- *Positive evidence:* A change in the position of an object was found in the region of the sky where one expected a planet.
- *Negative evidence:* None. No other object was seen to change position, and this object that did change position could not have been a star since it was moving too fast.

2.6. Example 6: Astronomy – Halley's comet

The following is taken from *Understanding scientific reasoning* (5th edition), R. Giere, J. Bickle, R. Mauldin, 2006, Thomson Wadsworth Publishing.

HALLEY'S COMET

Newton's work, *The Mathematical Principles of Natural Philosophy*, or *Principia*, an abbreviation of its original Latin title, was first published in 1687. Around 1695, Edmond Halley, an English astronomer and friend of Newton, began applying Newtonian models to the motions of comets. He was probably acting on Newton's own suggestion that comets may be like small planets with very large elliptical orbits. In any case, comets were very interesting objects because they had always been viewed as mysterious, even ominous. Their appearances certainly exhibited no apparent regularity. If Newton's suggestion were correct, however, the behavior of comets would exhibit a great deal of underlying regularity.

Halley began investigating a comet that he had observed in 1682. His observations of 1682 provided a quite precise location for the comet's path relative to the background stars. However, because the comet could be observed only during that small part of its orbit taking it near to the Sun, it was not possible to determine how big the whole orbit might be and, thus, how long it would take for the comet to return. Indeed, it was impossible to determine from those observations whether the orbit was an ellipse, as Newton suggested, or a parabola. Newton's theory allowed the possibility of a parabolic orbit, but such an orbit would mean that the comet would come by only once and then leave the solar system forever. If, however, the orbit were elliptical, the comet should have traveled that same path many times before.

Halley began digging into the records of observations of previous comets. He found 24 recorded observations, going back roughly 150 years, for which the records were precise enough to compare with the observations of 1682. For two of these, one in 1606–1607 and one in 1530–1531, the recorded orbits were very close to that of the 1682 comet. Halley argued that it was extremely unlikely that three different comets should have such similar orbits and concluded that these were three appearances of the same comet in an elliptical orbit with a period of roughly 76 years. He speculated but could not prove that the slight discrepancies in the three orbits were due to gravitational influences from the planets, particularly Jupiter.

But Halley did not stop here. Using the data from all three cases, together with the hypothesis that he was dealing with a system represented by a Newtonian model, Halley calculated the time of the next return. He boldly predicted that the comet should be seen again in the latter part of December 1758. Figure 3.6 will help you to keep in mind the relevant details of this example.

Halley published his work on comets in 1705. It was well received by Newton and the growing band of English Newtonians. It did little, however, to convince the French. Halley died, a respected scientist, in 1743, 15 years before the predicted return of the comet. By this time, even the French were coming around to the Newtonian way of thinking, and Halley's prediction was remembered. In 1756, the French Academy of Science offered a prize for the most accurate calculation of the time the comet would return. The comet reappeared, as predicted, near Christmas 1758 and was officially named "Halley's comet."

Analysis

Step 1. Real World. The real-world object of most direct interest is the path of a comet, particularly the comet observed in 1682.

Step 2. Model. The model was a Newtonian model for two bodies in an elliptical orbit attracting one another by the force of gravity. The two bodies are the Sun and the comet.

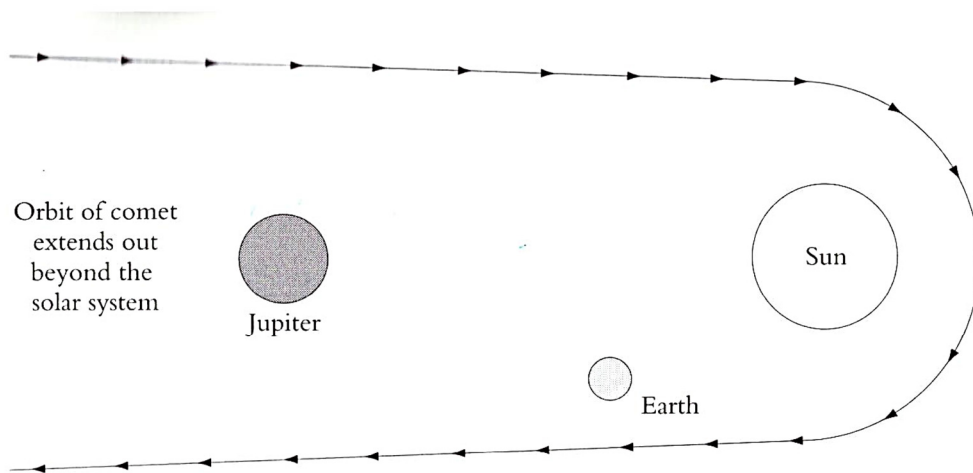


FIGURE 3.6
Halley's comet.

Step 3. Prediction. The prediction was that the comet would return near the end of 1758.

Step 4. Data. A comet with the requisite orbit did appear as predicted.

Step 5. Negative Evidence? No. The data clearly agreed with the prediction.

Step 6. Positive Evidence? The only alternative hypothesis was that another comet with the same orbit just happened to appear right around the predicted time, 76 years later. That seemed to everyone extremely unlikely. So the data provided very good evidence that the Newtonian model fits.

2.7. Example 7: Chemistry – The theory of Phlogiston

The following is taken from *Understanding scientific reasoning* (5th edition), R. Giere, J. Bickle, R. Mauldin, 2006, Thomson Wadsworth Publishing.

Fire, like the motions of the heavens, has always fascinated people. In the Western world, recorded speculation about the nature of fire goes back to the Greeks. We owe to them both the myth of Prometheus and the view that the world is made up of a few separate elements: earth, air, fire, and water. All four are present in the process of combustion. The common-sense view of combustion is that something is driven out of the burning object, leaving only ashes behind.

By the eighteenth century, this something had a well-established name, *phlogiston*—the fire stuff. Assuming that combustible material contains phlogiston explains most of the obvious facts about combustion. Heating drives off the phlogiston into the air; cooling makes it less volatile; smothering holds it in. The well-known fact that a burning candle placed in an enclosed container soon goes out was explained by saying that the enclosed air gets saturated with phlogiston, so that the phlogiston remaining in the wax has nowhere to go.

Phlogiston accounts not only for combustion but also for the very important process of smelting. This is the process by which crude ores are turned into more refined metals. Generally, this is done by carefully heating the ores, together with a measured amount of charcoal, to a controlled temperature. It was claimed that the charcoal contains an excess of phlogiston, which, at moderately high temperatures, leaves the charcoal and combines with the ore to form the metal. This hypothesis was substantiated by the fact that further heating at higher temperatures returns the metal to its original state. The phlogiston is driven out of the metal by the higher temperature. Even rusting was explained as the result of phlogiston slowly escaping from the metal.

These claims may be taken as characterizing *phlogiston models*. Such models lay behind many hypotheses about systems undergoing combustion, rusting, or the process of smelting. The *phlogiston theory* included the general hypothesis that this sort of model fits most cases of combustion, smelting, rusting, and so on. In what follows, we concentrate on combustion.

Combustion is very difficult to study. Most things we commonly burn are made of many different substances and give off many different gases when burned. Moreover, combustion generally is rapid and violent. Progress in such studies required finding some simple, well-controlled subjects for experimentation. In the 1770s, chemists developed a number of techniques for performing such experiments. The leaders were Joseph Priestly in England and Antoine Lavoisier in France. Priestly supported the phlogiston theory; Lavoisier led the revolution that overthrew it.

In the 1770s, using techniques first developed by Priestly, Lavoisier performed several careful experiments with mercury. In one of these experiments, he floated a precisely measured amount of mercury on a liquid and covered it with a glass jar, thus enclosing a known amount of air (Figure 3.7). The mercury was then heated through the glass using the rays of the Sun focused by a powerful magnifying glass (a burning glass). In such circumstances, as Lavoisier well knew, a red powder, or ash, forms on the surface of the mercury. Some of the mercury undergoes a controlled burning.

Applying a phlogiston model to this experiment, we would expect two things. First, the resulting mercury plus red ash should weigh less than the original sample of mercury alone. This is because some phlogiston must be driven off, leaving the ash behind. And the volume of air inside the jar should increase because it now contains the phlogiston that was driven out of the mercury. This means that the level of the liquid inside the jar would drop to make room for the additional “air.” When Lavoisier completed the experiment, the water level had gone *up*, and the mercury/ash combination weighed *more* than the original mercury alone.

Analysis

Step 1. Real World. The real-world phenomenon under investigation is the process of combustion. The specific case investigated is the controlled combustion of mercury.

Step 2. Model. The model is a phlogiston model according to which an invisible substance, phlogiston, is given off by various materials, including mercury, when they are appropriately heated.

Step 3. Predictions. The predictions, based on a phlogiston model and the experimental setup, are (1) that the water level under the bell jar should go down (because of the phlogiston added to the air) and (2) that the mercury/ash combination should weigh less than the original mercury alone (because phlogiston has been driven out of the mercury, which became ash).

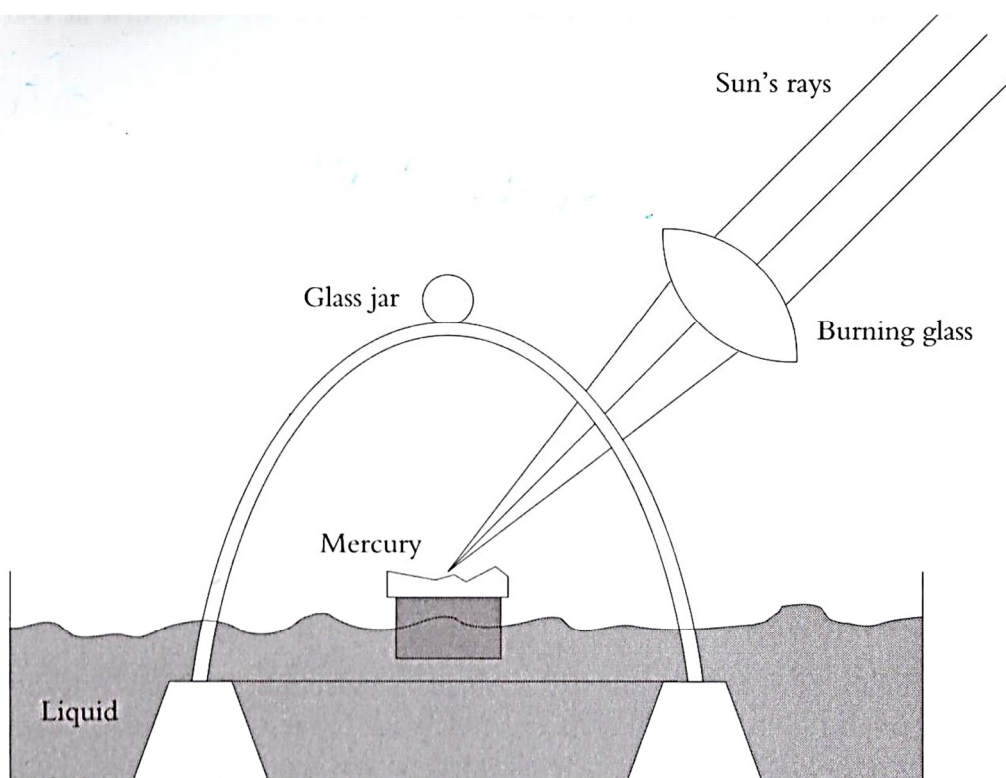


FIGURE 3.7

Lavoisier's experiment.

Step 4. Data. The data are (1) that the water level under the bell jar went up and (2) that the mercury/ash combination weighed more than the original mercury alone.

Step 5. Negative Evidence? The data and the predictions disagree. The data, therefore, provide evidence that the phlogiston model fails to represent the controlled combustion of mercury as carried out in this experiment.

Note that this analysis has only five steps. This happens whenever the analysis ends with evidence against a suggested hypothesis and there is no mention of any alternative supported by the data. Many readers will realize, of course, that there is an obvious alternative to the phlogiston model. According to this alternative model, there is something in the air that combines with the mercury to produce the ash. The water level goes up because the amount of air decreases, and the mercury/ash combination weighs more than the mercury alone because of the added weight of whatever it is in the air that combines with mercury to produce the ash. Lavoisier called this something "oxygen."

Example 8 – Statistics

As previously mentioned stated statistics does not deal with finding the truth of a hypothesis. Instead it deals with whether or not a hypothesis is *significant*. Absolute truth is related to the certainty that an effect is associated with a cause, and the fact that an observed effect will always occur again under the same conditions. Significance is related to the chance or likelihood that an effect is due to a cause, and the effect may or may not occur again under the same conditions but that there is a high probability that it will occur again. So, statistics deals with the *probability* that such hypotheses may, or may not, be likely. In this case evidence is not absolute, but is relative to the trend of large scale data.

As an example of such a situation, consider a company which produces thermometers that are supposed to give readings of 0°C at the freezing point of water. Since the company cannot test every single thermometer it produces it has to test a sample of them (say 50 thermometers). Tests on a large sample of these instruments reveal that, at the freezing point of water some thermometers give readings below 0°C and some give readings above 0°C .

Assuming the mean reading is 0°C and the standard deviation is 0.15°C , and that the readings are normally distributed (i.e. have a bell-shaped curve), we could ask the following questions: what is the probability that one thermometer chosen at random

- i) will read the freezing point of water to be 0.1°C or less?
- ii) will read the freezing point of water to be above -0.2°C ?
- iii) will read the freezing point of water to be between -0.1°C and $+0.2^{\circ}\text{C}$ inclusive?

So, already the idea of what counts as evidence in statistics is going to be different to the idea of evidence in maths or the physical sciences.

The aforementioned hypotheses are statements concerning a "population" ("population" here is used in the statistical sense as a large group of objects which display a trend, pattern or distribution). For example,

1. "this value comes from a set of data having this pattern or distribution"
2. "the average of this sample of data comes from this larger data set."
3. "two samples have been taken at random but we don't know if they come from the same set of data. We can test this by testing to see if their averages are similar enough."

Such statements are framed in a specific manner known as a *statistical hypothesis*. So, for the statements above we have

1. "this sample value comes from a population which is normally distributed."
2. "the mean \bar{x} of this sample comes from a population which is normally distributed with mean μ "
3. "there is no difference (at a statistical level) in the means \bar{x}_1 and \bar{x}_2 between the two samples."

So far there is no distinction between the underlying idea of a statistical hypothesis and a scientific hypothesis: both types of hypotheses are questions about predicting something. The distinction comes in the way we test our data in order to confirm or deny our hypothesis. In statistics we must decide whether it is *likely* that the sample mean *is close to* or *far away from* the population mean. Here the words, "likely", "close to", and "far away from" are deliberately chosen to illustrate the probabilistic nature of the evidence derived from statistical testing.

Statistics tests performed on these hypotheses are based on setting up a null hypothesis and an alternate hypothesis. What do such hypotheses mean? As an example, suppose that the lengths of metal bars produced by a machine follow the pattern of the normal distribution and have mean length $\mu = 420$ cm and standard deviation $\sigma = 12$ cm. The machine is serviced, after which a sample of 100 bars gives a mean length of $\bar{x} = 423$ cm. Has the servicing of the machine affected the mean length of bars produced? In other words, *is there evidence of a difference in the mean length of bars produced?* Two hypotheses are set up:

- the null hypothesis H_0 : There is no significant change/difference between the sample mean and the population mean;
- the alternative hypothesis H_1 : There is a significant change/difference between the sample mean and the population mean.

From the perspective of the hard sciences there is clearly a difference between the mean length of the sample and the original mean length of the population. This difference is 3 cm. So on the basis of our evidence our hypothesis is false.

From the perspective of statistics the question is whether or not the difference of 3cm significant. Why consider significance as a criteria for a hypothesis? Because different samples contain different data values which implies different means between samples between samples and the population mean. So there is no way we can make a direct comparison of the

mean of a sample and the mean of a population since there is always a natural variation of means between samples, whether such samples come from mechanical processes (due to machines wearing out or becoming un-calibrated) or ecological processes involving living organism (since these do not respond in mechanistic ways).

In the case of the length of metal bars any one bar chosen at random after the servicing of the machine won't confirm or deny the hypothesis. For example, if the bar we choose happens to have a reading of 420cm this does not confirm our hypothesis that the servicing had no effect. Maybe we were lucky in choosing a bar which had the same length as the average length. Or maybe the servicing has had an effect on other aspects of the machine which we have not hypothesised. Similarly, if we choose a bar (after the machine has been serviced) which has length 425cm, this does not falsify our hypothesis. This does not automatically confirm that the servicing has had an effect since we have a standard deviation of 12cm, so bar lengths up to 432cm would be common.

So the nature of what constitutes evidence in statistic is not clear cut. The evidence will actually be based on

- *the sample mean being sufficiently close to the population mean* for us to accept that it is likely that the servicing of the machine has not significantly altered the way the machine works,

or

- *the sample mean being sufficiently far away from the population mean* for us to accept that it is likely that the servicing of the machine has significantly altered the way the machine works.

or

- in the case of the example above involving thermometers the evidence will be based on the likelihood that a thermometer chosen at random could read the freezing point of water sufficiently close to 0°C or too far away from 0°C, or could read the freezing point of water *within an acceptable range* of temperatures.

But what do we mean by “sufficiently close”, “sufficiently far away from”, or “within an acceptable range”? How do we determine this? Well, this depends on how sure we want to be as to the whether the servicing has had an effect. This assuredness is expressed as a percentage. So, we would say that we want to be 95% or 99% or 99.9% sure that the servicing

of the machine has had no effect on the mean length of bars it produces. And it is us who choose the percentage assuredness.

Again, this approach is totally different to the nature of evidence in the hard sciences. Here,

- we are looking for definitive evidence that an electron exists, or that a magnetic field has this-or-that effect or property, or that a thermometer reads the freezing point of water as 0° Celsius, or that Halley's comet will return in 76 years from its last appearance, etc.,
- we are not looking to be 95% or 99% confident that the electron exists. We are not looking to be 95% or 99% confident of the shape of the magnetic field around a rectangular bar magnet; We are not looking to be 95% or 99% confident that Halley's comet will return 76 years after it last appeared.

Having said this, when physicist at CERN are trying to find new particles, or confirm the existence of a particle predicted by theory (such as the Higg's Boson), they use probability levels or confidence intervals to confirm the existence of such a particle. However, they set probability level of 0.00003%, or a 3 in 10 million chance that the particle discovered is mistaken for the one they are looking for. In that case it is most likely the particle they have found is indeed for the one they are looking for. They call this a "5 sigma" level of significance.

Another way of explaining the idea of percentage assuredness is to say that there is an event which would be rare if H_0 is true. For example, such an event would occur only 5% of the time or only 1% of the time or even only 0.1% of the time. If this is the case then we cannot accept our alternative hypothesis that the event is due to some deliberate effect or action, and we say that there is no reason to reject the null hypothesis that there is no difference or no effect. If an event has occurred the question is not, "Is this an event which be would be rare if my null hypothesis were true?" but "Is there an alternative hypothesis under which the event would be relatively frequent?". Hence, we don't say that "we accept our null hypothesis" only that "we see no reason to reject our null hypothesis". This is because there may be other hypotheses relating to other causes or multiple causes which could explain the effect observed. We need ask such a question because of the nature of how we analyse large quantities of data, i.e. we analyse large quantities of data statistically in order to determine probabilities, likelihoods, trends, patterns, etc. Hence evidence based on statistical hypothesis testing is of a different nature to evidence gathered from hard sciences.

The null hypothesis is never proved or established but is possibly disproved in the course of experimentation. But evidence (which is not of a statistical nature) from the hard sciences does not take this form. For example, supposing you have evidence that the electron exists. According to the statistical way of framing a hypothesis we say

Here is evidence against the hypothesis that the electron does not exist,

whereas in science we say

Here is evidence that the electron exists.

The hard scientist typically does not try to disprove things. S/he is always looking for appropriate evidence which *confirms* his hypothesis.

In general, no evidence which comes from the statistical analysis of data will ever guarantee the truth or falsehood of a hypothesis. Such evidence can only suggest (to lesser or greater degrees of probability) whether or not the effect (i.e. something having happened, or the outcome) is due to a hypothesised cause (i.e. the thing causing that outcome).

Finally, it comes down to the way in which one interprets the test results in the context of the discipline. It is much easier to confirm or reject a null hypothesis when analysing the manufacture of nuts and bolts than it is to do so when studying ecological phenomena simply because of the wider range of variability of living systems over time compared to the narrower range of variability over time of inanimate objects. So, in statistics one might say that it is not so much that we have evidence not to reject our null hypothesis or to accept our alternative hypothesis, but whether the evidence is of sufficient weight to be convincing.

2.8. Exercises

If you have time, and the interest, analyse one or more of the following texts in the same manner as illustrated in examples 2) - 7) above:

- Analyse each text according to the criteria Real world, Model, Prediction, Data, Negative evidence, Positive evidence. Do not just give a “yes” or “no” answer but explain why you believe what you are saying.
- If necessary, check the internet for evidence which could have supported each model/theory.

These texts are taken from p86 of *Understanding scientific reasoning* (5th edition), R. Giere, J. Bickle, R. Mauldin, 2006, Thomson Wadsworth Publishing.

EXERCISE 3.1

The Tychonic System of the World

Most popular descriptions of the scientific revolution of the seventeenth century, like that in this text, follow Galileo's own presentation in concentrating on the conflict between the Ptolemaic and Copernican systems. Yet at the time there was a third alternative, which had been proposed by the equally famous Danish astronomer, Tycho Brahe (1546–1601). In the Tychonic system, Earth remains at the center of the universe and the Sun revolves around Earth. The other five planets, however, revolve around the Sun, not Earth.

Evaluate the evidence for the Copernican system provided by Galileo's observations of the phases of Venus, this time taking into account Tycho Brahe's alternative model (ignoring the Ptolemaic model). Does your analysis provide any help in understanding why it took so long for the Copernican hypothesis to become widely accepted?

EXERCISE 3.2

The Discovery of Neptune

During the first half of the nineteenth century, astronomers were still working out tables and charts giving the positions of the various planets. In this, they were aided by Newtonian theoretical models. But then the outermost planet, Uranus, caused some difficulties. Its observed orbit differed from what it should have been according to the then best-fitting Newtonian models. And the difference was much too great to be attributed solely to inaccuracies in measurement. They were forced to conclude that their current models were not correct. But they did not give up Newton's theory of celestial mechanics. By that time, there had been so many successful predictions using Newtonian models that they were reluctant to conclude that the general theory could be wrong. Around 1843, the English astronomer J. C. Adams and, somewhat later, the French astronomer Leverrier independently calculated that the observed orbit of Uranus could be explained if there were an additional planet beyond Uranus whose gravitational force produced the deviations from the earlier Newtonian predictions—which, of course, assumed no such planet. Using this more elaborate Newtonian model, Adams and Leverrier were able to calculate just where the new planet should be at any particular time. The planet, named "Neptune" by Leverrier, was observed in 1846, just where it was predicted to be.

EXERCISE 3.3

The Missing Planet: The Story of Vulcan

Like the orbit of Uranus, the observed orbit of the innermost planet, Mercury, failed to fit Newtonian models by amounts that could be reliably measured. Fresh from their discovery of Neptune, many astronomers immediately assumed that there must be yet another planet, closer to the Sun than Mercury. Leverrier even named the new planet Vulcan and calculated just where it should be. Although several people claimed to have seen Vulcan, these reports were never substantiated.

EXERCISE 3.4

The Wave Theory of Light: 1818

After a century in the shadow of Newtonian ideas, according to which light is composed of small particles, wave models of light were revived about 1800, first by an Englishman, Thomas Young, and then by a Frenchman, Augustin Fresnel. Fresnel's model was submitted for a prize offered by the French Academy of Sciences. One of the judges for the academy, S. D. Poisson, deduced that according to Fresnel's model, the shadow of a small circular disk produced by a narrow beam of light should exhibit a bright spot right in the center of the shadow. Poisson and the other judges are reputed to have thought that this refuted Fresnel's hypotheses because they had never heard of there being such a phenomenon and regarded it as highly unlikely to exist. No known particle models predicted such a spot. But when the experiment was carried out in carefully controlled circumstances, there was the spot, just as required by Fresnel's model. Fresnel received the prize in 1818.

EXERCISE 3.5

The Wave Theory of Light: 1849

During the first half of the nineteenth century, there was a long controversy over the nature of light. According to Newtonians, light consists of small particles moving at high velocities, so that what we call light rays are really Newtonian particles. The competing theory, advocated mainly by French physicists, was based on the idea that light is really an example of a system of waves, like waves on the surface of a calm lake, in a bowl of jelly, or on a vibrating stretched string. Using standard Newtonian models, it was calculated that light should travel faster in water than in air and by a precisely determined amount. Similarly, using wave models, it was calculated that light should travel slower in water than in air, also by a precisely determined amount. It took until 1849 for anyone to design instruments that could measure the velocity of light accurately enough to detect the predicted differences. When the experiment was finally done, it was found that the velocity of light is lower in water and by the amount claimed by the wave theorists.

EXERCISE 3.6

The Oxygen Theory

Lavoisier was not only interested in refuting the phlogiston theory, he sought also to establish his own oxygen theory. In many respects, oxygen models have just the opposite structure of phlogiston models. According to Lavoisier's oxygen model, when combustion occurs a component of normal air (about one-fifth by his measurements) combines with the combustible material. This component of air came to be called oxygen. Analyze Lavoisier's experiment with mercury as a crucial experiment applied to both phlogiston and oxygen hypotheses. It may help you to know that the red powder that formed on top of the mercury is mercuric oxide.

In-class discussion (if time allows):

- What constitutes evidence in your discipline?
- How do you go about obtaining evidence in your discipline?
- How strong does your discipline's evidence have to be before you accept it as evidence in favour of your hypothesis?
- To what extent does evidence depend on the current theoretical knowledge of one's discipline? For example, is there likely to be evidence of the existence of electrons if our theory about atoms is that they are indivisible (as was first proposed by John Dalton in 1803)? Are we likely to have found evidence of the planet Uranus (discovered officially in 1781) if we didn't have Newton's theory of gravitation (a theory predicts the position of planets but which did not predict the position of Saturn sufficiently accurately, leading to the idea that there must have been a planet beyond Saturn)?
- You may need to do some reading in order to answer this question: True or false
 - i) can there be evidence which supports a false hypothesis?,
 - ii) can there be evidence which goes against a hypothesis which is true?,
 - iii) is it always the case that there is evidence which supports a true hypothesis?

Why? Explain.

3. A commentary on evidence

1) Examples 1) - 7) above highlighted the case of evidence as confirming a theory about the truth of natural phenomena (the structure of the solar system, the discovery of planets, moons, aether, and oxygen). Here we can see that evidence is

evidence in favour of a theory or model designed to represent the Truth.

But what about the things we study which have nothing to do with the truth of nature or the universe (i.e. natural phenomena)? For example, how do we speak of evidence in relation to software design, coding and implementation? How do we speak of evidence with respect to data science? There is no search for truth here. There is only a search for ...? What do these disciplines search for? And how then do you explain or justify that you have found what you are looking for? In this case might we not say that evidence is

*evidence in favour of a theory or model
designed to approach predetermined standards or rules?*

2) Evidence doesn't exist unless we have data. In order to obtain data we need to conduct experiments. But what do we then do with the data gathered from our experiments? We analyse it by identifying, separating, converting, organising and/or modelling it in such a way as to isolate useful information. We then use this information as a basis for accepting (or rejecting) a hypothesis or a theory. Then we can say that on the basis of the data collected, and the analysis performed, we have (or don't have) evidence which supports our hypothesis.

3) Scientific evidence and experimental data are not the same. What constitutes evidence is the way in which we interpret the data. Otherwise such data remains simply as data. But the interpretation of such data can be affected by one's scientific background knowledge as well as one's epistemological beliefs, i.e. the beliefs we hold about what counts as valid or invalid interpretation and knowledge. Scientific data simply acts as the *potential* for evidence for confirming (or denying) a hypothesis.

With respect to calculating the charge on the electron, "Franklin (1987) analyzed Millikan's oil drop experiments and found that not all measured data were used as evidence: In Millikan's work published in 1923, he only used 58 data [points] to support his claim from the 175 data [points] measured since 1911. Among the data that he did not use, two were discarded because the device used to obtain the

data was not appropriate, twelve were discarded because theoretical corrections were necessary to accept these data, *six were discarded because they differed from his expectation, twenty-two were discarded without even being calculated for their charges, and five were discarded simply for being considered in some way insufficient* [emphasis added] (Franklin, 1987)." (from "The nature of scientific evidence and its implications for teaching science", Jongwon Park et al. *Journal of Baltic Science Education*, Vol. 20, No. 5, 2021)

Note that the Truth of a phenomenon or object always provides evidence of the existence of that phenomenon or object. The problem is whether or not we have the right type of evidence, sufficient evidence, or whether or not we interpret the evidence correctly.

3.1. Examples of evidence which supports a false hypothesis

Does evidence in favour of a hypothesis only apply when the hypothesis is true? Can evidence support a false hypothesis? Does a true hypothesis *always* have evidence to support it?

You might think that evidence applies only to hypotheses which are true but this is not the case. There can be evidence for an incorrect hypothesis. A good example of this is the experiment carried out by German physicist Heinrich Hertz in 1883 when he was studying something called cathode rays. Cathode rays had been discovered by another German physicist Julius Plucker in 1859 by passing electricity through a highly rarefied gas (such as air, most of which has been evacuated from a container).

These cathode rays were discovered by passing an electric current through a glass tube containing special equipment, and where the air pressure inside the tube had been significantly reduced (creating a near vacuum). Passing the electricity at one end of the glass tube would then produce a greenish phosphorescent glow on the glass at the opposite end of the glass tube.

Later experiments showed that the position of the glow changed if a magnetic field was applied across the path of the cathode rays, and as time, and more experimentation, progressed two theories came to be developed: cathode rays were either i) atoms or molecules of the gas in the tube that had become negatively charged, or ii) waves forms, just as light was a wave form. The former theory would mean the presence of negatively charged particles (what was later found to be electrons). The reason for this belief was that the rays

were deflected by magnetic fields and this could only occur if the rays were streams of particles. However, Hertz believed in the latter theory (i.e. ii)) because cathode rays passed through thin metallic foils of gold, silver or aluminium without leaving any holes whilst at the same time creating a phosphorescent image on the glass behind the foil. Hence these rays could not be particles.

From this discovery Hertz attempted to find out if these cathode rays were electrically neutral. In other words, would these rays remain undeflected when an electric field was passed across the path of the rays?

- *Real world:* The real-world phenomenon under investigation are cathode rays;
- *Model:* Cathode rays are wave forms;
- *Prediction:* Hertz's prediction was that cathode rays would not be deflected because he believed that cathode rays behaved like waves (not like particles) and hence are electrically neutral;
- *Data:* When an electric field is applied the phosphorescence effect still occurs directly ahead on the glass tube and not to one side of the glass tube;
- *Negative evidence:* None;
- *Positive evidence:* No deflection of the rays was found.

Hertz therefore concluded that cathode rays were electrically neutral, implying the non-discovery of negatively charged particles. And this is the key point: Hertz was one of the best experimental physicists of the 19th century. He knew how to set up and conduct experiments in order to study phenomena. Then

- he had a hypothesis, i.e. "cathode rays are not deflected when an electric field is passed through them". In hindsight his hypothesis was wrong;
- he obtained evidence that this hypothesis was true because his experiment showed no deflection of the rays.

In other words, Hertz obtained valid evidence of an incorrect hypothesis.

Between 1883 and 1897 physics theory and experimentation had improved, and new data had been collected such that in 1897 J. J. Thomson would repeat Hertz' experiment, with better theoretical understanding and experimental set-up, to show that cathode rays were deflectable. This lent weight to the idea of cathode rays being streams of particle (instead of

wave forms) and would ultimately lead to the discovery of the electron.

So the absolute truth was/is that cathode rays can be deflected, leading to the absolute truth that electrons exist whether or not we know they exist. Hertz' "truth" that cathode rays could not be deflected was relative, and based on

- the state of knowledge of physics in 1883,
- scientists' abilities at conducting experiments,
- the type and quality of equipment used at the time,
- the beliefs and cognitive biases scientists had at the time,
- the fact that they were not in a position to know anything else,

etc., So, relative to his situation Hertz was justified in believing that cathode rays were electrically neutral, but in terms of absolute truth Hertz was wrong to conclude that cathode rays were electrically neutral and could not be deflected. Many people replicated Hertz's experiments and found no deflection in the cathode rays. Even J. J. Thomson's initial experiments showed the same results. It was only later in 1897 that he was able to show that cathode rays were deflectable, and this later lead him to discovering the electron.

So, in general, it seems as if evidence in favour of a hypothesis counts as a good reason for believing our hypothesis. However, it should be noted that *absence of evidence is not evidence of absence*. Just because we find no evidence doesn't mean to say that there is none.

For the technically minded: What was the reason for Hertz' "failure" to detect any deflection? Simply put, Hertz' glass tube had not been sufficiently evacuated. There was still too much air in the glass tube. This caused the air to become ionised by the electric field produced between the anode and cathode plates. Hence the positive ions moved towards the negative (cathode) plate and the negative ions moved towards the positive (anode) plate. The net charge on the plates was therefore either zero or so small that any remaining electric field was too weak for the equipment of the day to detect any deflection of the cathode rays (note the deflection would have been expected if the cathode rays contained electrically charged particles).

So, the truth of a phenomenon or object always provides evidence of the existence of that phenomenon or object. However, the outcome of an experiment does not always provide evidence for the truth of the phenomenon or object.

As another example, consider the belief that the Earth is flat versus the belief that the Earth is spherical. There was always evidence of the Earth being a sphere:

- It is just that people did not have the imagination, insight or awareness to hypothesise the Earth being spherical;
- It is just that we did not have the equipment to perform the experiments which could confirm our insight. As such, we were not able to gather the data which would give us the evidence of a spherical Earth;
- Even if we had the equipment would we have performed the correct experiment which would have allowed us to collect the relevant data from which we could find the evidence of a spherical Earth;
- Even if we had the equipment and performed the correct experiment, then we would have collected the appropriate data to confirm a spherical Earth. But we may have misinterpreted the data and still decided that the Earth was flat.
- Why might we still believe the Earth was flat? Because we might not have the necessary pre-requisite knowledge which would help us come to the right conclusion that the Earth was spherical. We might regard the data suggesting a spherical Earth as anomalous data, and ignore it.

In fact, people of ancient time did conduct the correct experiments: they used their eyes to look to the horizon. They did collect the appropriate data: they saw the horizon to be circular. It is just that they misinterpreted the data to conclude that the Earth was a flat disc. Because the Earth is so big compared to the size of a human being, it looks flat when seen from its surface. If we had been able to fly high above the Earth then we would have had a greater perspective on the shape of the Earth, and seen its shape against the background of space.

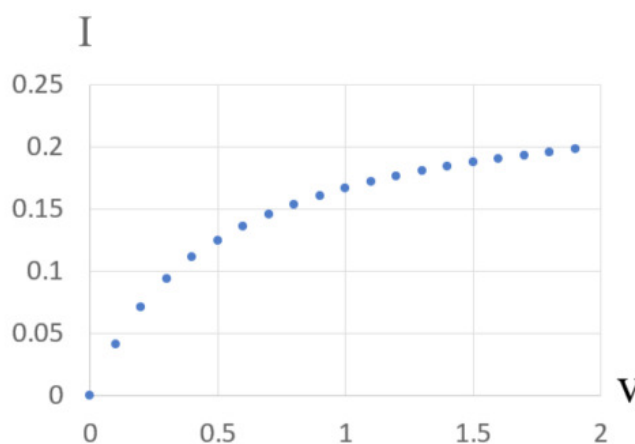
So, we can obtain evidence in favour of a hypothesis even if the hypothesis is incorrect. For example,

- *Hypothesis:* The Earth is flat; *Evidence:* On the basis of repeated observations in our daily lives living and walking on Earth we see the Earth to be flat under our feet; *Conclusion:* Our hypothesis is true.
- *Hypothesis:* The Earth is stationary; *Evidence:* On the basis of repeated observations we see the stars, planets and Moon and Sun revolve around the Earth; *Conclusion:* Our hypothesis is true.

- *Hypothesis*: Cathode rays are electrically neutral; *Evidence*: On the basis of repeated experiments cathode rays are not deflected when an electric current is passed across the rays; *Conclusion*: Our hypothesis is true.
- Ohm's law is $V = IR$. This states the relationship between voltage and current for a given resistance. *Hypothesis*: Ohm's law is true under all conditions; *Evidence*: On the basis of repeated experiments this is seen to be true; *Conclusion*: our hypothesis is true. However, this is true only if the temperature of the resistor is kept constant. If the temperature varies then the linear relationship between V and I fails. This example addresses the distinction between a core hypothesis and conditions or assumptions under which that hypothesis is true.

“Theories and hypotheses consist of core aspect which describe/explain the essence of the phenomenon as well as consisting of conditions, assumption, and other aspects. Therefore, conflicting evidence can disprove these peripheral aspects rather than the core theory. For example, looking at the data on the voltage and current on a small light bulb shown in the figure below, the data do not match Ohm's law. However, this data does not disprove Ohm's law, which is the core theory, because the condition that ‘the temperature of the resistance should be constant’ was not satisfied.

Electric Current and Voltage on a Small Electric Bulb



(from “The nature of scientific evidence and its implications for teaching science”, Jongwon Park et al. *Journal of Baltic Science Education*, Vol. 20, No. 5, 2021.)

In other words, evidence has to be relevant evidence, i.e. evidence gathered according to the conditions, constraints, assumption etc., of our hypothesis.

Note that misleading evidence is still evidence of something. Incorrect/false evidence is still evidence of something. It is just that we may believe it to be evidence in favour of our hypothesis because we have certain cognitive biases or predispositions about our hypothesis (such as a belief that our hypothesis has to be true because it explains what we are observing). But we, and others, should really repeat/replicate our experiments under various conditions in order to gather more data (such as evacuating glass jars to different degrees in order to test for any deflection of cathode rays), or to gather data from alternative experiments (such as testing the shift in Mercury's perihelion, testing the bending of starlight, and testing time dilation (i.e. the slowing down of time due to gravitational fields) for confirming Einstein's theory of relativity). Such collective evidence then acts to strengthen the case that our hypothesis is true (or true at least within certain conditions and assumptions). We therefore need to be careful that we do not make the mistake of drawing the conclusion that our hypothesis is correct based on a single piece of evidence.

Exercise

This exercise may take some time and research, particularly into the history of your discipline. List as much of ideas, beliefs, knowledge, theories, technologies, etc., that were known about in your discipline 20, 30, 50, 100 years ago or more. What did people know, what did they not know, what possible cognitive biases might they have had, what technologies did they have access to in order to test and confirm their hypotheses, etc? Then,

were the people of the day justified in believing
the ideas, knowledge and theories of that time? If so, why? If not, why not?

3.2. Examples of necessary and sufficient evidence: Multiple lines of independent evidence

Evidence which supports a hypothesis does not automatically confirm that the hypothesis is true. Clearly the evidence in favour of the hypothesis has to be *necessary* evidence (otherwise we will never confirm our hypothesis). But is that enough? No. There has to be other supporting evidence in order to confirm our hypothesis, otherwise there can be an alternative hypothesis which is also explained by the current evidence. Consider the following examples:

Example 1

Hypothesis 1	Evidence	Hypothesis 2
Planets move in circular orbits.	We see planets move across the night sky to return every night.	Planets move in elliptical orbits.
The Earth is stationary in space, and the planets move around the Earth.	We see planets move across the night sky to return every night, and the Earth does not move in space.	The Earth moves across the sky and the planets move relative to the Earth.

Therefore, the concept of orbits is a *necessary* condition to explain planetary motion but not a *sufficient* one. We need more data to explain the path of the orbits (circular or elliptical) and whether or not the Earth moves in space (the retrograde motion of Mars becomes an anomaly for a stationary Earth but not for a moving Earth).

Repeated observations over a period of several nights, weeks, and months provides *necessary* evidence that everything revolves. But our hypothesis is that everything revolves around the Earth, so clearly our necessary evidence is not *sufficient* evidence. Even repeating this experiment every night for 12 months happens not to be sufficient because the retrograde motion of Mars (which constitutes anomalous data to our hypothesis) occurs only every 2 years or so, and then only for a few weeks.

Example 2

Hypothesis 1	Evidence	Hypothesis 2
Comets are not planets.	They have elliptical orbits.	Comets are planets because planets also move in elliptical orbits
Comets are not planets.	They have long periods (i.e. the time to make one complete revolution)	Comets are planets because all planets from Jupiter onwards have long periods.


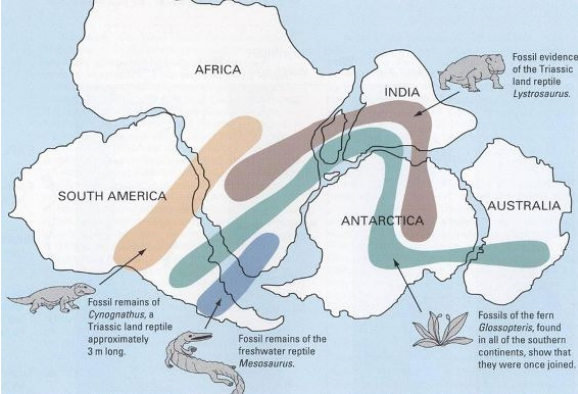
Comets are not planets.	They have a rocky core surrounded by an icy outer layer	Comets are moons because some moons of Jupiter and Saturn also have a rocky core surrounded by an icy outer layer
-------------------------	---	---



Therefore, elliptical orbits, long periods and rocky/icy structure are *necessary* conditions to distinguish between comets and planets but not *sufficient* ones. We need more data in order to identify the distinguishing features of a comet from a planet.

Example 3

Hypothesis 1	Evidence	Hypothesis 2
Millions of years ago the continents were once joined together to form a single land mass.	The distribution of fossils across the world shows that comparable fossils are found on widely separated continents, implying they must all have originally developed on the same land mass.	The species evolved independently on separate continents.
		Hypothesis 3
		The species swam or flew to the other continents to establish a second population.

This is where multiple evidence comes in (see [What Is the Theory of Continental Drift - ScienceAid](#) for four lines of evidence), as illustrated below.

The matching of continents	Fossil correlation
	

Correlation between rocks and mountains	Past climate data
	<p>“He began with glaciers which are ordinarily found in large rivers in the coldest parts of the world such as the North and the South poles. Glacier streams are constantly moving, and while they do, they scratch the rock beneath them, leaving evidence of their trajectory. These scratches are called glacier striation. The interesting finding was that he discovered glacier striations in the tropical rainforest of Africa and South America. His theory was that those regions were not always located on the Tropic, but might have been down closer to the south pole and slowly drifted away.” (Reference: See URL above)</p> 

Example 4

Another example is that Newton's theory of gravitation predicted the return of an object that seemed to behave like a comet. Recognised by Edmund Halley (1656 – 1742) only in 1705 as a comet, its return was predicted to be in 1758. The return of this object on at this time would then be *necessary* but *not sufficient* evidence that it was a comet orbiting the Sun. After all the object could have been a far-away planet on closest approach, or a supernova which had exploded and stayed extremely bright for a few months (as they do).

One would therefore need to wait (according to Newtons' theory) another approximately 76 years to confirm Newton's law of gravitation as correct. Waiting 76 years would have been one way to confirm the law. Another way (which is what Halley actually did) would have been to look at historical records for an object having appeared in the sky at that same location every 76 year or so. If the records showed (which they did) such a reappearing object, along with the fact that this objects produced "fiery" tails, i.e. tails of of steam, as it approached the sun, this would act as the *necessary and sufficient* confirming evidence that the object was a comet with a period of 76 years.

Exercise

From February 1930 (when it was first discovered by Clyde Tombaugh) to August 2006 Pluto was considered to be a planet. In August 2006 the International Astronomical Union (IAU, the world governing body on all matters to do with astronomy) changed the status of Pluto from a planet to a dwarf-planet. By checking the internet and the IAU website find evidence which suggests Pluto is a planet, and find evidence that Pluto is not a planet.

So we can distinguish necessary evidence from sufficient evidence as follows:

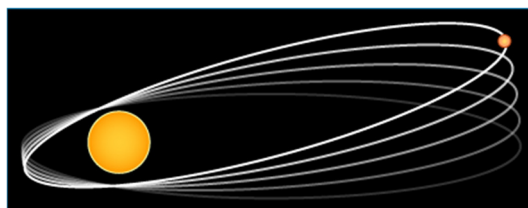
- 1) necessary evidence is evidence we must have in order to confirm the truth of a hypothesis or theory. In this case necessary evidence is only partial confirmation of the truth of the hypothesis or theory;
- 2) sufficient evidence is the minimum total amount of evidence that is needed to confirm the truth of hypothesis or theory. In this case sufficient evidence is confirmation of the total truth of the hypothesis or theory.

For example,

- *Necessary conditions that are not sufficient*
 - For the whole numbers greater than two, being odd is necessary to being prime. However, being odd is not a sufficient condition for being prime since there are odd numbers which are not prime numbers (e.g. 15, 21, etc.);
 - For the geometric figure of a square, it is necessary that the square be four-sided. But this is not a sufficient condition since rectangles and rhombuses have four sides.
So we might say that a square is defined by having four sides and four 90° angles. But even these are not sufficient conditions since a rectangle also has these properties.
So we might say that a square is defined by having four sides of equal length and four 90° angles. But these are not sufficient conditions since a parallelogram also has these properties.
Etc.
- *Sufficient conditions that are not necessary*
 - An integer being divisible by 4 is sufficient for it to be even, but not necessary since divisibility by 2 is sufficient (and necessary) for it to be even.
- *Necessary and sufficient conditions*
 - An integer is even if it is divisible by 2;
 - A prime number is a number which is divisible only by 1 and by itself (not by any other number)
 - A square is defined as (1) a closed polygon, (2) having four sides all of equal length, (3) with two sets of parallel sides, and (4) having internal angles of 90° .
Note that (1) on its own could imply a circle or a rectangle or square, (1) and (2) on their own could imply a parallelogram or a square, (1) and (3) could imply a parallelogram or a square, but only (1) - (4) implies a square.

3.3. Other examples

- Einstein's theory of general relativity has been confirmed by multiple experiments over the last 100 years.
 - *The bending of starlight:* Since massive objects create gravity we needed to study the behaviour of light as it passed a massive object. Would such an object bend starlight? The obvious object to use was the Sun and the opportunity arose in 1919 (4 years after Einstein published his theory) to conduct a suitable experiment. This experiment was conducted by Arthur Eddington, and his aim was to take photographs of the background sky during a total solar eclipse (in Principe, West Africa). With the Sun "blocked out" he would be able to photograph the stars (i.e. collect data) and compare their position recorded during the eclipse with images of the same region of space taken when the Sun was in a different portion of the sky (more data). The before-and-after comparison of the photos confirmed that the stars has "moved", showing this to be evidence of the bending of starlight and evidence in favour of Einstein's theory of relativity.
 - *The precession of the perihelion of Mercury:*
Planets revolve around the Sun in elliptical orbits. This means that there is a point in the orbit when a planet is closest to the Sun and furthest from the Sun. The point of closest approach is called the planet's perihelion. Planetary orbits shift very slightly so that over time a planet's orbit will not be in the exact same location in space as its previous orbit, as illustrated in the diagram below:



So the orbit itself rotates in space. This shifting of the position of the orbit is called precession. As a result of this precession a planet's perihelion will shift. Newton's theory of gravitation had accounted for the precessions of the perihelia of all the planets except for Mercury. But Einstein's general theory of relativity was able to calculate Mercury's position accurately thus providing more evidence that relativity was a valid theory.

- *Time dilation*: Another confirmed prediction of general relativity is that time dilates in a gravitational field, meaning that clocks run slower as they approach the mass that is producing the field. This has been measured directly.
- The Plate Tectonic model of the Earth is supported by multiple independent lines of evidence, for example
 - magnetic stripes in rocks showing sea-floor spreading,
 - the global distribution of earthquakes and volcanoes,
 - comparable fossils found on widely separated continents and satellite measurements.

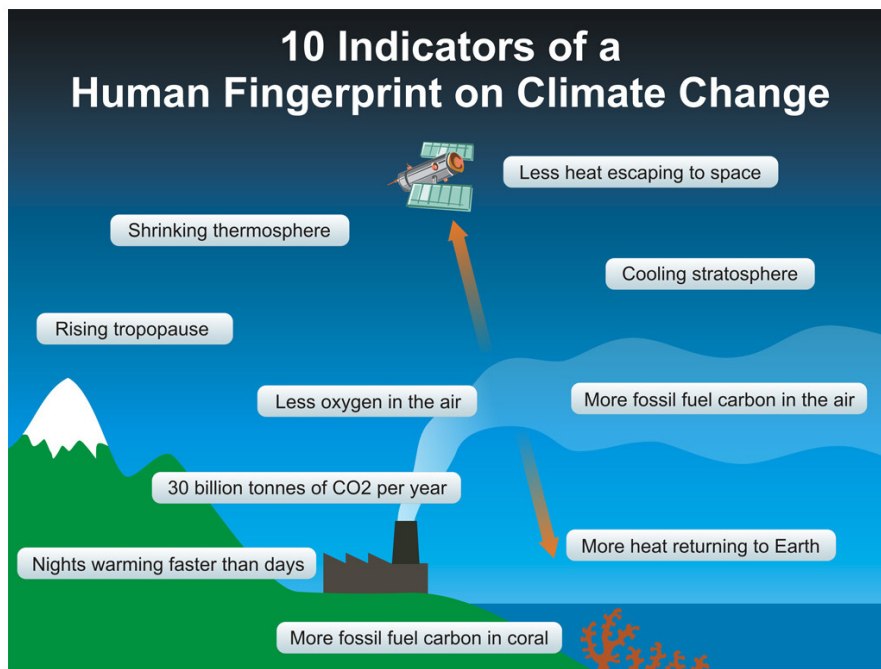
If you are interested in this go to <https://pubs.usgs.gov/gip/dynamic/developing.html> for more info.

- The Big Bang model for the origin of the universe is supported by multiple lines of evidence:
 - the expansion of space as described by Hubble's Law,
 - the existence of the cosmic microwave background radiation,
 - the distribution and evolution of galaxies,
 - the abundance of light elements in the universe,
 - the large-scale structure of the universe,

For more evidence go to <http://www.talkorigins.org/faqs/astronomy/bigbang.html>

- The current model of global warming model proposes that this, and climate change, is caused principally by emissions of greenhouse gases. This is supported by multiple independent lines of evidence such as
 - Satellites measure less heat escaping to space at the precise wavelengths which CO₂ absorbs;
 - Surface measurements find this heat is returning to Earth to warm the surface.
 - Fossil carbon is building up in the atmosphere. (We know this because the two types of carbon have different chemical properties.);
 - Oxygen levels are falling as if carbon is being burned to create carbon dioxide.
 - Corals show that fossil carbon has recently risen sharply;
 - Humans are currently emitting around 30 billion tonnes of CO₂ into the atmosphere;

- If the warming is due to solar activity, then the upper atmosphere (the stratosphere) should warm along with the rest of the atmosphere. But if the warming is due to the greenhouse effect, the stratosphere should cool because of the heat being trapped in the lower atmosphere (the troposphere). Satellite measurements show that the stratosphere is cooling;
- This combination of a warming troposphere and cooling stratosphere should cause the tropopause, which separates them, to rise. This has also been observed;
- It was predicted that the ionosphere would shrink, and it is indeed shrinking.



(Diagram from <https://skepticalscience.com/its-not-us-basic.htm>)

3.4. Exercises

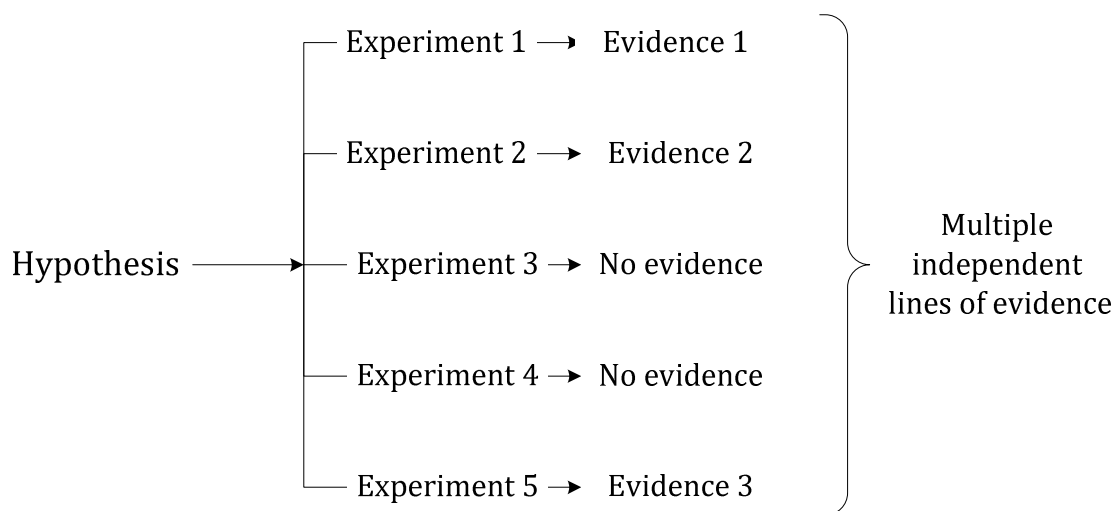
The following comes from [The Concept of Necessary Conditions and Sufficient Conditions \(sfu.ca\)](https://www.sfu.ca/~cepc/sfu.ca):

1. x's being a square is a sufficient condition for x's being a rectangle.
2. x's being a square is a necessary condition for x's being a rectangle.
3. x's being a rectangle is a sufficient condition for x's being a square.
4. x's being a rectangle is a necessary condition for x's being a square.
5. x's being a mother is a sufficient condition for x's being a female.
6. x's being a mother is a necessary condition for x's being a female.
7. x's being a female is a sufficient condition for x's being a mother.
8. x's being a female is a necessary condition for x's being a mother.

9. x 's being greater than 15 is a sufficient condition for x 's being less than 20.
10. x 's being greater than 15 is a necessary condition for x 's being less than 20.
11. x 's being less than 20 is a sufficient condition for x 's being greater than 15.
12. x 's being less than 20 is a necessary condition for x 's being greater than 15.
13. x 's being less than 12 is a sufficient condition for x 's being less than 20.
14. x 's being less than 12 is a necessary condition for x 's being less than 20.
15. x 's being less than 20 is a sufficient condition for x 's being less than 12.
16. x 's being less than 20 is a necessary condition for x 's being less than 12.
17. x 's having two arms is a sufficient condition for x 's being a human being.
18. x 's having two arms is a necessary condition for x 's being a human being.
19. x 's wanting to do a is a sufficient condition for x 's doing a .
20. x 's wanting to do a is a necessary condition for x 's doing a .
21. x 's being an equilateral rectangle is a sufficient condition for x 's being a square.
22. x 's being an equilateral rectangle is a necessary condition for x 's being a square.

(answers can be found at the end of the notes)

Clearly we need necessary and sufficient evidence. How can we achieve this? The requirement for what is necessary and sufficient evidence in a discipline requires a good understanding of that discipline. But in general we can say that we need multiple independent lines of evidence. We collect as much evidence as possible from as many different sources (say repeated and replicated experiments, or different experiments performed to test the same hypothesis), with each confirming piece of evidence supporting our theory, as illustrated below.



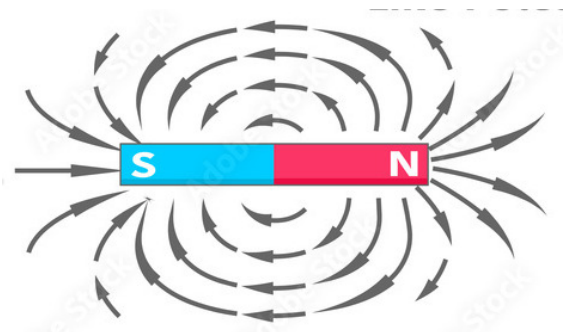
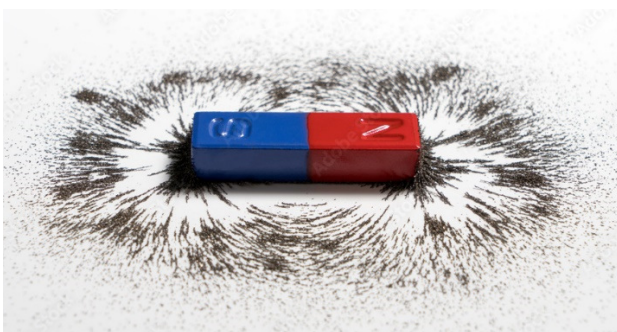
In principle there should be no falsifying evidence. In practice this is not the case, and there is usually anomalous evidence (evidence which our hypothesis/theory cannot explain) or falsifying evidence (evidence which goes against our hypothesis/theory).

Evidence of the truth of a hypothesis or theory relies on the fact that there is a Truth to be found. Our hypothesis may or may not be true, hence the need for multiple evidence, or evidence from multiple sources. But, a physical object or phenomenon is what it is, does what it does, behaves as it behaves. It is the Truth. So, there will always be evidence of the truth of this object, phenomenon, process, behaviours, etc.

4. How do we find evidence for things we cannot see?

How do we find evidence for phenomena we cannot see or touch or hear? For example

- we cannot see or touch wind or heat so how do we know these phenomena exist? Because we can feel them, and we can see their effect on other objects: the wind makes the branches of a tree move and heat melts ice and burns paper;
- we cannot see magnetic fields so how do we know they exist? Because of their effects: magnetic fields attract certain metals and also have a repelling effect. Also, how do we know the shape of the magnetic field? We can't see it directly so we have to conduct a suitable experiment. If we sprinkle some iron filings randomly on a piece of paper and then place a rectangular bar magnet onto the filings these filings will arrange themselves in the way shown below, where the left image is an actual photograph of the effect and the right image is an ideal representation of the effect:



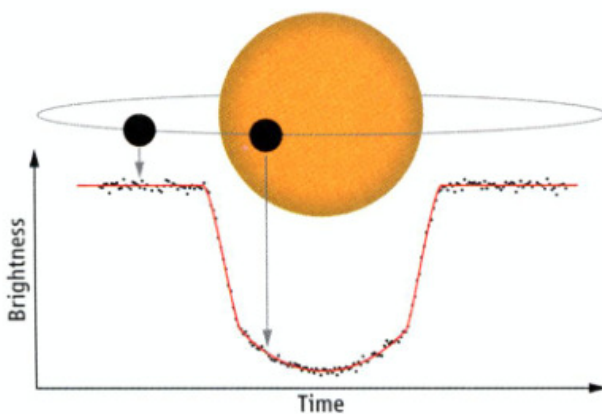
So, we “see” the shape of the magnetic field by the effect it has on the iron filings.

- we cannot see or touch or feel the electron. But we can conduct experiments to see if such a thing exists. This is what happened with the work of Heinrich Hertz and J. J. Thomson (see above). Here it was hypothesised that cathode rays were electrically neutral and therefore a beam of cathode rays would not be deflected. Initially this

hypothesis seemed to be confirmed until better equipment and procedures allowed one to actually deflect cathode rays thus leading to the new hypothesis that this was due to a negatively charged particle, later called the electron.

- if a distant star has a planetary system it is virtually impossible for us to visually see these planets from Earth so we need another means of finding out whether any particular star has planets revolving around them. One way in which this is done is by studying the brightness of a star to see if this decreases or not. If there is a decrease in brightness we then need to know how long this decrease lasts for. If the decrease in brightness last for months or years then this may mean one thing. If the decrease in brightness last for days or weeks then this may mean something else.

In the case of a planet moving across the star the star's brightness will decrease only for a short time. Such a motion is called the transit of a planet across the disc of a star.



Transit of a planet: here is an illustration of a planet (small black circle, shown at two different times and hence two different points in its orbit) and its parent star, which is shown as brighter near its centre and darker at the edge of the star. As the planet moves onto the disk of the star, it first blocks out the light from the edge, then the more intense light from the centre of the star, leading to the cupped shape shown by the red line and data points.

(adapted from “Evidence of Things Not Seen”, Norman W. Murray, *Science*, New Series, Vol. 336, No. 6085 (1 June 2012), pp. 1121-1122, American Association for the Advancement of Science)

More generally, if we cannot see phenomena how do we know they exist? Initially we don't. How do we start experiments which study unknown phenomena? Initially we can't. In the case of magnetism this had been known for a few thousand years because of rocks such as lodestones (and others). Lodestones are naturally magnetized pieces of the mineral magnetite. They are naturally occurring magnets, which can attract iron. It only required for people of the day (i.e. living thousands of years ago) to find them. Similarly for heat. The moment people found out how to make fire they started the scientific process of investigating how to generate it, how it worked, what its effect was, etc.

As for things like x-rays, gamma rays and the electromagnetic spectrum in general this would have to wait until both scientific theory and technology was sufficiently well developed for it to be possible to conceive of such phenomena. Even the accidental discovery of phenomena such as x-rays, penicillin (antibiotics), the cosmic microwave background radiation, etc., required a preparedness of scientific thinking and knowledge, as well as technology for people to discover these. Until this happened we were unable to even guess that such things existed.

Exercises

1) Does your discipline deal with Truth in the way the physical sciences do? If not, there may be no such thing as your hypothesis being true, only that your hypothesis is confirmed or viable. In this context which of the following words best defines the word “evidence” as seen from your discipline?

Fact	Justification	Belief
Opinion	Interpretation	Explanation
Understanding	Valid / validation	Judgement
Significance	Confirmation	Indication
Corroboration	Substantiation	Verification
Manifestation	Indication	Basis for

Explain why your chosen words are more appropriate as ways of describing evidence. For example, if you choose “confirmation”, why is it more appropriate to say in your discipline that a hypothesis is confirmed instead of saying that your hypothesis is true?

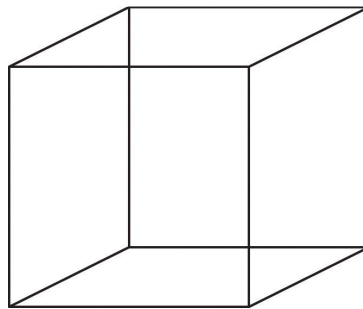
2) Below is a reminder of the types of cognitive biases. Remember that evidence = data + interpretation. How do think cognitive biases affect evidence?

Cognitive bias	Explanation
Anchoring bias	Relying too much on the first piece of information you learn or hear
Blind-spot bias	Recognizing bias in others, but failing to recognize it in yourself

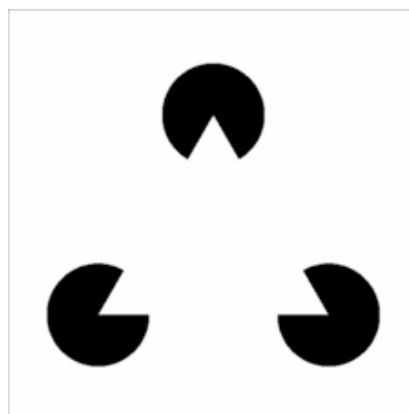
Confirmation bias	Listening to and trusting only information that confirms your beliefs
Negativity bias	Focusing on negative events at the expense of positive or neutral events
Barnum Effect	The tendency to see general personality descriptions as highly accurate descriptions of ourselves

Bias can easily be illustrated by the following examples.

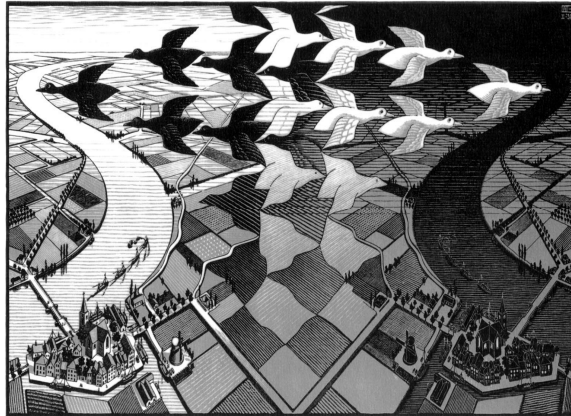
- In the diagram below is there evidence that the cube is pointing towards the upper right or towards the bottom left? How do you justify your answer?



- In the diagram below is there evidence of a triangle, or are you reading a triangle into the diagram, or is it just three circles with sectors cut out? How do you justify your answer?



- In the diagram below is there evidence of white birds flying to the right or black birds flying to the left? How do you justify your answer?



In the three examples above we can say that our interpretation amounts to “reading into” what we are looking at. Does this count as evidence?

5. A complete example

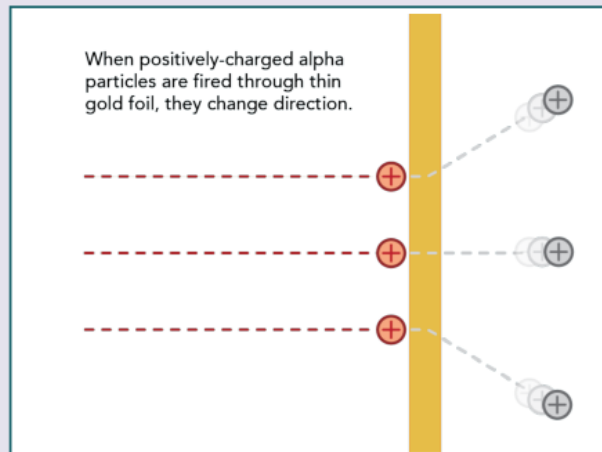
Below is an example taken from <https://undsci.berkeley.edu/understanding-science-101/what-is-science/science-works-with-testable-ideas/>

Using Ernest Rutherford’s investigations into the structure of the atom as an example, it illustrates the process of

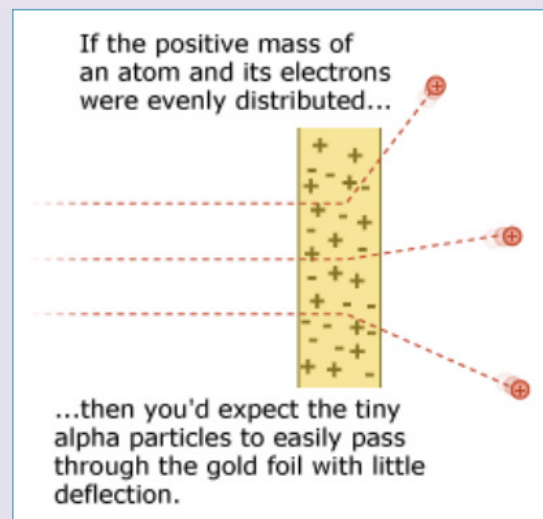
hypothesis --> experiment --> evidence --> theory --> anomalies --> new hypothesis
--> new experiment --> new evidence --> updated theory --> new anomalies, ...

A SCIENCE PROTOTYPE: RUTHERFORD AND THE ATOM

Ernest Rutherford's investigations were aimed at understanding a small, but illuminating, corner of the natural world: the atom. He investigated this world using alpha particles, which are helium atoms stripped of their electrons. Rutherford had found that when a beam of these tiny, positively-charged alpha particles is fired through gold foil, the particles don't stay on their beeline course, but change direction when passing through the foil. Rutherford wanted to figure out what this might tell him about the layout of an atom.



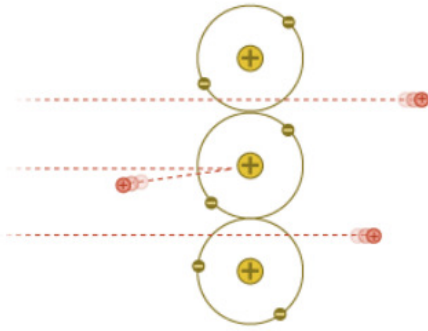
Before 1910, Ernest Rutherford and many other scientists had the idea that the positive charge and the mass of an atom were evenly distributed throughout the whole atom, with electrons scattered throughout. You can imagine this model of the atom as a loosely packed snowball (the positive mass of the atom) with a few tiny grains of sand (the electrons) scattered throughout. The idea that atoms are arranged in this way can be tested by firing an alpha particle beam through a piece of gold foil. If the idea were correct, then the positive mass in the gold foil would be relatively diffuse (the loosely packed snow) and would allow the alpha particles to pass through the foil with only minor scattering.



Ernest Rutherford's lab tested the idea that an atom's positive mass is spread out diffusely by firing an alpha particle beam through a piece of gold foil. The evidence resulting from that experiment was a complete surprise: most of the alpha particles passed through the gold foil without changing direction much as expected, but some of the alpha particles came bouncing back in the opposite direction, as though they had struck something dense and solid in the gold foil. If the gold atoms were really like loosely packed snowballs, all of the alpha particles should have passed through the foil – but they did not!

From this evidence, Rutherford concluded that their snowball model of the atom was incorrect, even though it was popular with many other scientists. Instead, the evidence suggested that an atom is mostly empty space and that its positive charge is concentrated in a dense mass at its core, forming a nucleus. When the positively charged alpha particles were fired at the gold foil, most of them passed through the empty space of the gold atoms with little deflection, but a few of them ran smack into the dense, positively charged nucleus of a gold atom and were repelled straight back (like what would happen if you tried to make the north poles of two strong magnets touch). The idea that atoms have positively charged nuclei was also testable. Many independent experiments were performed by other researchers to see if the idea fit with other experimental results.

The observed deflection of alpha particles shows that atoms have their positive charge concentrated in a dense mass.

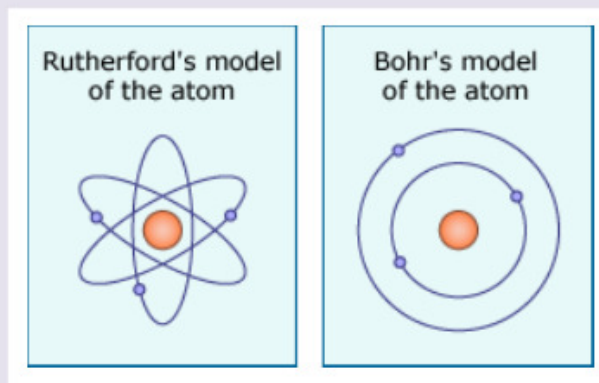


Though Ernest Rutherford came up with the idea that atoms have positively charged nuclei, the research that led to this idea was a collaborative effort: Rutherford was assisted by Hans Geiger, and the critical alpha-scattering experiment was actually carried out by Ernest Marsden, an undergraduate student working in Rutherford's lab.

Furthermore, after his discovery of the layout of the atom, Rutherford published a description of the idea and the relevant evidence, releasing it to the scientific community for scrutiny and evaluation. And scrutinize they did. Niels Bohr noticed a problem with Rutherford's idea: there was nothing keeping the orbiting electrons from spiraling into the nucleus of the atom, causing the whole thing to collapse! Bohr modified Rutherford's basic model by proposing that electrons had set energy levels, which helped solve the problem and earned Bohr a Nobel Prize. Since then, many other scientists have built on and modified Bohr's model.



Ernest Rutherford (right) and Hans Geiger in the physics laboratory at Manchester University, England, circa 1912. Photo credit: Permission of the Alexander Turnbull Library, Wellington, New Zealand, must be obtained before any re-use of this image. Reference number: PAColl-0091-1-011.



Lithium atoms, diagrammed in the Rutherford and Bohr models. Rutherford's model does not differentiate between any of the electrons, while Bohr's places electrons into orbits with set energy levels.



Niels Bohr. Photo credit: Wikimedia

Niels Bohr built upon Ernest Rutherford's work to develop the model of the atom most commonly portrayed in textbooks: a nucleus orbited by electrons at different levels. Despite the new questions it raised (e.g., why do orbiting, negatively-charged electrons not spiral into the positively-charged nucleus?), this model was powerful and, with further modification, led to a wide range of accurate predictions and new discoveries, including predicting the outcome of chemical reactions, determining the composition of distant stars, and conceiving of the atomic bomb.

Ernest Rutherford and his colleagues acted in ways that moved science forward:

- They understood the relevant knowledge in their field. Rutherford had studied physics for more than 20 years when he proposed the idea of the nucleus.
- They exposed their ideas to testing. Even though his original view of the atom suggested that no backscattering should occur, Rutherford decided to look for backscattered alpha particles anyway, just to be thorough.
- They assimilated the evidence. When their experimental results did not support the "snowball" model of the atom, instead of writing those results off as an anomaly, they modified their original ideas in light of the new evidence.
- They openly communicated their ideas so that other physicists could test them as well. Rutherford published the experimental results, a description of his reasoning, and the idea of the nucleus in 1911 in a scientific journal.
- They acted with scientific integrity. In his paper on the topic, Rutherford assigned credit fairly (citing the contributions of his colleagues, Geiger and Marsden) and reported his results honestly — even when experimental results and his theoretical calculations did not match up perfectly.

Answer to the “necessary and/or sufficient conditions” exercises 1 - 22

- | | |
|----------|-----------|
| 1. true | 9. false |
| 2. false | 10. false |
| 3. false | 11. false |
| 4. true | 12. false |
| 5. true | 13. true |
| 6. false | 14. false |
| 7. false | 15. false |
| 8. true | 16. true |

17. false (A chimpanzee, for example, may have two arms, but it is not a human being.)
18. false (A war veteran may have lost both his arms, but he is a human being even so.)
19. false (A person may want to win a foot race, but his/her wanting to do so, does not guarantee he/she can or will do so.)
20. false (A person may do something without wanting to do so, e.g. without wanting to hurt another's feelings, we sometimes do.)
21. true*
22. true*