

The Assumptions and Methods of Science¹

These notes are a continuation of the first set of notes and they assume knowledge of the same.

Components of scientific methodology

1. Guiding Ideas and the fruits of scientific work

A. Scientific Hypotheses: Hypotheses are tentative explanations of why some pattern in nature occurs. Thus, they are statements that attempt to explain the underlying processes (causes) responsible for patterns in nature (effects). Nobody really knows how people form hypotheses; they result from luck or individual, creative genius. No formula exists for producing them.

The common usage of the term "hypothesis" as a guess or explanation that explains some observation is very close to its meaning in science. However, unlike with common usage, a scientific hypothesis **must have the property that there must be some way (in principle) to show that the hypothesis is false as a result of observation of behavior of the object(s) under study**. Thus, we say that any good scientific hypothesis must be **falsifiable**. The criterion of **falsifiability**, which characterizes scientific hypotheses and provides a basis for distinguishing between scientific and nonscientific hypotheses, was introduced in 1934 by Sir Karl **Popper** (a renowned philosopher of science) in his book, *The Logic of Scientific Discovery*. According to Popper, to be scientific a hypothesis must be tested empirically against nature and accepted only if such testing cannot falsify it. Much of the process of science is finding clever, revealing ways to test hypotheses.

¹ © 2006 by W. R. Healy and K. N. Prestwich, Department of Biology, College of the Holy Cross

Analogy and the extension of knowledge: Popper's views of hypothesis testing trace partly to Darwin and Wallace's notion of natural selection -- their idea that genotypes contained within individuals are tested by nature and those that work the best (survive the test of nature) continue. Although by no means a perfect analogy, Popper saw hypotheses as something like genotypes and experiments and tests of logical consistency as acting like the environment. According to Popper, the best hypotheses continue because they are not falsified by experience with nature but just as with a genotype, none is guaranteed permanence as they are repeatedly tested.

Logical consistency and scientific hypotheses: Scientific hypotheses in fact must meet two tests. First they must be logically consistent both internally and with what is believed to be known about nature. Then, they must be tested further against nature (the material world).
The hypothesis that God is the creator of nature appears to be definitionally true and logically consistent with the philosophical propositions (assumptions) that every effect has a cause and nature is an effect, not a cause. However, it is not a scientific hypothesis because there is no way of falsifying it through testing. The reason is that ultimately this worldview states that God can act outside of material nature in un-seeable and unknowable (to humans) ways.
Take some philosophy and religion while you are here and learn more about this.

B. Scientific Theories: A hypothesis increases in level of certitude if it possesses:

- (i) broad explanatory power and
- (ii) it has survived repeated testing

Eventually, scientists come to refer to such an explanation as a **theory**. The difference between the terms "hypothesis" and "theory" is inexact but refers to the scope of the explanation and the level of certitude attached to it as a consequence of repeated testing. The greater the number of

specific predictions which are verified through experiment, the higher is the level of certitude attached to the hypothesis and the more likely it is to be referred to as a theory.

C. Scientific Laws: Scientific laws, on the other hand, **are generalizations about the behavior or operation of nature**. They are based on repeated observations and they usually do not contain a mechanism or explanation for why these patterns occur. Here are some examples of what are correctly termed scientific laws:

Thermodynamics -- there are three laws of thermodynamics, the best known are the first two (1) that the total amount of matter/energy is constant and that (2) overall entropy (disorder) of the universe as a whole increases in any process (change).

Evolution: members of populations change genetically (and as a result, morphologically) over time.

Again, notice that these laws only describe a pattern -- they do not explain these patterns. The explanations (mechanisms) that explain laws and other broad generalizations are hypotheses and theories. Thus, statistical thermodynamics is a theory that seeks to explain why entropy increases and natural selection and genetic drift are both theories that seek to provide mechanisms to explain evolution.

Use of terms You will see that for various reasons, all of these terms often are misused. For instance, evolution is often termed a theory when it is actually a biological law. Likewise, Darwin and Wallace's theory of natural selection is only a theory in the sense that it predicts that a certain process termed natural selection is responsible for much of evolution. More about all of this later -- and about the rather disingenuous arguments used in popular books such as *Darwin on Trial* that attempt to discredit evolution while seeking to advance a particular brand of vitalism.

Another example is the so-called "Cell Theory" (all living things are composed of cells) -- so named because it reflected the use of the term in the 19th century when the idea developed. The cell theory is more properly biological law (where does that leave viruses? -- we'll discuss that and the usefulness of law later). One other note -- do not take the sequence of hypothesis → theory as an inevitable progression. Some hypotheses are of such limited scope that we would never consider them scientific theories even though they are well established.

2. Processes of science

a. **Observation (Data collection)**. Many consider the starting point of any scientific investigation to be the collection of observations. These collected observations are termed **data** (plural, datum singular). The random collection of data, however, without any guiding hypothesis is not only tedious, it is most often fruitless and few do it (once again, Popper pointed this out. -- pre-Popper views of science focused on observation as the inevitable, not occasional starting point of the process of science). In fact, data are usually collected to test specific hypotheses, not to formulate them. We will talk more about this later.

b. **Observations and Induction**: This said, let's start with the less common cases when no prior information about a phenomenon exists. What does one do with a large number of observations? For these to be useable, one creates generalizations by the use of induction. Induction is a type of logical reasoning where the premise(s) support the conclusion but do not ensure it. In science this means that induction is the logical process of extracting generalizations (conclusions) from a set of observations (premises). The conclusion is not ensured because it is always possible that a new observation will be contrary to the previous ones.

- In science, induction is used to make a generalized statement based on a limited number of observations. For example, suppose

I trap mice in several different woodland lots and notice that all the mice I have captured are brown. Through induction, I could then argue that all mice are brown.

- Inductive generalizations **describe** broader patterns in nature. Some inductive generalizations are broad enough and well enough established that we call them scientific laws. However, the vast majority of inductive generalizations are not termed laws -- they are simply called generalizations for short. For example, there is no brown mouse "law."
- All inductive generalizations beg for hypotheses to explain them.
- Many think that induction must lead to hypothesis construction. It doesn't have to. It does so only because eventually someone wants to explain the generalization in terms of a mechanism or reason for its existence..

c. **Deduction -- the process of arguing from the general to the particular.** Deduction is most commonly used in science to derive predictions or postdictions (see below) based on hypotheses. Recall that hypotheses are tentative explanations of generalizations. Hypotheses can be tested or can be extended to unknown phenomena using deductive reasoning. Thus, if our working hypothesis is that DNA is the hereditary molecule in all living things, then, if we discover a new type of organism we could predict using deduction that its hereditary molecule is DNA. Note that once we made this deductive prediction, it now allows us to test further our hypothesis that the hereditary molecule of all living things is DNA.

d. **Predictions and Postdictions.** In science the term prediction means the statement of a fact **not** evident or known at the time the statement is made, but one that must exist if the hypothesis is true. It is a logical deduction that follows from the hypothesis. An example of a prediction is the statement that, based on the theory that DNA is the hereditary molecule, then any

newfound form of life would use DNA as the hereditary molecule. The prediction is the logical consequence of the theory. Moreover, It is a scientific prediction because it is testable (see below). Do not simply equate the term "prediction" with the statement of a future event (the usual meaning of the term prediction) because **a scientific prediction can be either:**

- **a literal prediction in the sense of a future event or a**
- **postdiction -- the statement of a past event unknown at the time the statement is made.**

Suppose that, based on observation of a comet's appearance in 1946, 1967 and 1988, an astronomer hypothesizes an orbit for this comet requiring 21 years to complete one full revolution around the sun. This hypothesis would yield two predictions: (a) that the comet will return in the year 2009, which is a literal prediction, and (b) that the comet did appear in the year 1925, which is a postdiction. If both predictions are verified, does this prove the hypothesis? The answer is no; verified predictions only support the hypothesis but don't prove it because (1) future observations could falsify it, e.g., if the comet was recorded in the year 1922 or it next returns in the year 2011, or (2) two or more comets with different orbits might give the appearance of the same 21 year cycle.

e. Tests and Experiments: The central feature of science is that its hypotheses and their predictions can be falsified through testing. It is the role of the scientist is to generate predictions from hypotheses using the logical process of deduction. Most importantly, good science involves generating predictions that can feasibly be shown to be either true or false.

Karl Popper (see above) effectively introduced the concept of falsifiability as being central to hypothesis testing. He stated that it was difficult if not impossible to prove hypotheses. According to Popper, proof exists only in mathematical theorems, not in scientific methodology.

IMPORTANT: Why is it easier to falsify than to prove? For a hypothesis to be proven true, it requires that the predictions of the hypothesis be **always** in accordance with unbiased observation. However, a single or just a few situations where the hypothesis is not borne out suggest that it is incorrect, at least in part. So, a single false prediction disproves and forces either modification or total discarding of the hypothesis. It should be evident that even a thousand correct predictions do not prove for all time that a hypothesis is correct. This is part of the basis for the idea that all scientific work is tentative and not absolute (see earlier discussion on absolute truth).

So, modern science generally follows Popper and uses the idea of falsifiability, not the impossible "provability", as the best criterion for deciding whether or not a hypothesis is scientifically useful.

Note that in lab we will see that due to uncertainty in observation, single instances of "false" prediction may not be enough for us to disprove a hypothesis. This is the equivalent of saying that sometimes, when you don't get a good look at something, what appears to be happening is not what is actually happening. We use statistics to estimate the chance that we have come to the correct conclusion.

To summarize, **hypotheses that have not been falsified through testing are provisionally, (but not absolutely), accepted**. Such tentative acceptance is the norm because scientific hypotheses still are always capable of falsification and might be falsified in the future. Remember that the discarded theories of the past were accepted as scientific truths in their day!

Falsification and experimentation. An experiment is simply a way of testing hypotheses by finding out if specific predictions are true or false. In a more restricted sense an **experiment is a manipulation of the conditions that result in some phenomenon**, whereas an observation is simply a witnessing of the phenomenon. The most effective type of experiment is a

controlled experiment wherein all factors that can influence a phenomenon are kept constant except one which is allowed to vary. In this way the effect on the phenomenon under study of the variable can be determined precisely and accurately. The controlled experiment is erroneously considered by many to be the hallmark of science, but in actuality it is hypothesis testing (through either **observation or experimentation**) which distinguishes science from non-science.

Sham treatments: in efforts to make controls and experimental groups as alike as possible (except for the actual experimental manipulation), living organisms that act as controls are often given fake treatments. You are almost certainly with one type of sham treatment -- the administration of a **placebo** drug to controls. Other examples of sham treatments include actual surgery on controls that mimics the surgery given to experimental animals except in the crucial aspect being tested by the experiment. Needless to say, although these are commonly done on animal subjects, there may not necessarily be done when humans are used.

The Role of Statistics in Hypothesis Testing

Now is a good time to consider the **criteria we use to trust experimental results**. Certainly, knowing that an experiment was competently carried out with appropriate, properly functioning equipment is a central requirement. But beyond this, we know that biologist must contend to varying degrees with individual differences (from which they hope to generalize to groups) and with observational errors. How are these handled in deciding whether or not experimental results support or reject a hypothesis? The answer is that we enlist statistical procedures to help us make this judgment. You will be introduced to statistics in the lab and I encourage all of you to take a course in statistics sometime while you are here (the sooner the better). Here is an overview.

In any experiment, **there is always variation** in results between controls and experimental subjects. This happens even when the subjects are the same in both cases. For our purposes there are three sources to this variation:

1. Variation induced by measurement errors: Measurements, whether made by a machine or a human observer always have at least some error. An observer may not see or hear something as clearly one time as the next or the power to an instrument might fluctuate slightly and alter the measurement. Moreover, anytime a **continuous variable**² is being measured and converted to an exact (**discrete**) value, there is always some type of rounding error.
2. Variation due to individual differences not germane to the experiment: biological variation occurs within any group (even with cloned or inbred individuals). Thus, when dealing with organisms, we expect to see differences simply because the subjects are different individual or because individuals vary over the time needed to complete an experiment. Normally we try to control for this in an experiment by comparing groups of individuals which have similar genetic make-ups and which were raised in similar environments. Thus, the power of using individuals as their own controls (but there are also many problems with this approach). In any case, the degree to which subjects vary among individuals and over time induces a bias into our experiments -- we usually say that this bias is due to a **sampling error** where the term comes from the idea that the experimental and control groups were not identical in all ways except for the experimental manipulation.
3. Variation due to experimental manipulation: these are the differences that interest us. These differences are predicted from the original hypothesis to be the result of the manipulation; they are what are being tested. In a controlled experiment we need to be reasonably sure that a predicted outcome occurred or did not occur (i.e., did we falsify or support our original main hypothesis).

² A continuous variable potentially has a near infinite number of values. Height and weight are both, in theory, continuous variables.

Since all of the factors just mentioned contribute to variation, what **we need is a means to estimate the size of any differences caused by the experiment (#3 above) as compared to differences we expect from reasons 1 and 2.** We need this because it is possible that differences between a control and experimental group could be due items #1 and 2 instead of the manipulation. This might lead us to falsely accept or reject a hypothesis that predicted a certain effect from the experimental manipulation.

Inferential statistics are procedures that **allow us to determine the likelihood that some of the difference between the control and experimental groups is due to the experimental manipulation and not to measurement and sampling errors.**

Specifically, inferential tests are normally used to choose between what are called **statistical hypothesis**. These are always part of any statistical test and they represent two opposite outcomes.

- **the Null hypothesis (H_0)** posits that **none of the differences between the control and treatment (experimental) group were the result of the experimental treatment (*i.e.*, there was no discernable experimental effect).** The term null means none -- no effect. By contrast:
- **the Alternative Hypothesis (H_A)** posits that **at least some of any differences between control and experimental groups were due to the experimental manipulation itself** and not solely to measurement and sampling errors.

Examples of inferential statistical tests include **Chi Square, Analysis of Variance, T-test.. All give the probability that the null hypothesis is correct.**

We use the probability that the null hypothesis is true in conjunction with explicit rules (which you will learn in lab) in order to decide whether or not to reject the null hypothesis. **The most commonly used rule is that we accept the null hypothesis if the chance of it being incorrect is greater than 5%.** In

other words, we will accept the notion that the experimental manipulation did nothing and any differences are due to factors such as biological variation and sampling and measurement errors) even if there is only a slightly greater than 5% chance that this hypotheses is true.

- So, if the statistical test we do on our results shows that there is a 6% chance that the results of an experiment were explainable entirely by error (i.e., a 94% that they were explainable by error **plus** treatment effect) we still accept the null hypothesis and assume that the alternative (that there was an experimental effect) is false!
- This gives us a fairly high standard of proof.
- But, if for instance the chance the null hypothesis was correct was only 4% (0.04) we reject it and accept the alternative that there was some experimental treatment effect. Notice that this means that there was a 96% chance that we did the correct thing and a 4% chance that we were incorrect.
- So much for absolute certainty in science (see last set of notes).

If this is all a bit overwhelming, don't fret. We'll do a concrete example in class and you'll also discuss it in lab.

A Bit More About Hypotheses, Theories and Falsification: What happens when one specific prediction of a theory is falsified? Does this mean that a long accepted explanation of how nature operates must be rejected? The answer is no! The theory will be retained as long as it functions in providing a reasonable interpretation of nature and continues to generate new information by being the guiding light behind research projects. Theories, therefore, are paradigms which:

- explain a significant portion of reality,
- interpret new data and

- generate new information by directing research effort through questions raised by the theory itself.

A theory is not just the explanation of a pattern, it is a paradigm or way of looking at the world; as such, it is exempt from naive falsification, i.e., rejection based on the empirical falsification of one or two of its predictions. Of course, if more and more predictions are falsified as time goes on, the usefulness of the theory diminishes and it may be replaced by a new theory that provides a better paradigm for interpreting nature.

Theory replacement, according to the philosopher of science **Thomas Kuhn** is the most exciting event in science and the one that accounts for the success of science over the long run. Note again that the replacement of one theory by another *does not necessarily mean gradual progression towards some form of absolute truth. In many instances the new theory doesn't build on the one it replaces, it actually destroys it.* Accepted theories represent the body of current scientific truth and they are the means whereby we make sense out of the world, interpret new information and expand the frontiers of knowledge through research effort.

"Beauty" and Simplicity in Scientific Theories: The Law of Parsimony (Occam's Razor): This "law" is really more of an assumption. It states that **when two or more explanations exist for observations, the simplest is always to be preferred** (at least until it shown to be incorrect -- then we go for something more complex). It is based on many observations that suggest that nature can be explained by a series of relatively simple rules. A common use of parsimony in biology is working out evolutionary lineages. It is generally assumed that hypotheses that involve the fewest possible steps are most likely to be correct, but they need not be so. A bit more about this later in the course.

Study Questions:

1. Explain why scientists attempt to disprove rather than prove their hypotheses. Discuss the differences between science and non-science.

2. Discuss the importance of theories to the scientific mode of inquiry and comment on the role of absolute truth and progress in science.

3. What are the sources of variation in experimental results and why are statistics used in experiments?

4. Suppose that someone carefully measures the frequency of cricket sounds and finds that they are all below 6 kHz and above 2 kHz. She then concludes that any insects with calls in this frequency range are crickets. What type of logical process did she follow (beyond the term "flawed")? What should happen next to advance the science of cricket calls? Explain.

5. How do world views influence: observation? Scientific theory?

6. Distinguish between each of the following:

- scientific and nonscientific hypotheses
- hypothesis and theory
- pattern and process
- induction and pattern recognition
- deduction and prediction
- hypothesis and fact/information
- theorist and empiricist
- induction and deduction
- Null and alternative hypothesis
- experimental and statistical hypothesis
- continuous vs. discrete variable

7. Identify each of the following:

- Popper
- Kuhn
- falsifiable
- prediction
- postdiction
- hypothesis testing
- controlled experiment
- parsimony; Occam's razor
- sham treatment
- placebo
- experimental and control treatments
- Inferential statistics (and names of examples of typical tests)