Chapter 1: An Introduction to Philosophy of Science

Malcolm Forster, February 24, 2004.

General Philosophy of Science

According to one definition, a general philosophy of science seeks to describe and understand how science works within a wide range of sciences. This does not have to include every kind of science. But it had better not be confined to a single branch of a single science, for such an understanding would add little to what scientists working in that area already know.[1]

Deductive logic is about the validity of arguments. An argument is valid when its conclusion follows deductively from its premises. Here's an example: If Alice is guilty then Bob is guilty, and Alice is guilty. Therefore, Bob is guilty. The validity of the argument has nothing to do with what the argument is about. It has nothing to do with the meaning, or content, of the argument beyond the meaning of logical phrases such as if…then. Thus, any argument of the following *form* (called *modus ponens*) is valid: If P then Q, and P, therefore Q. Any claims substituted for P and Q lead to an argument that is valid. Probability theory is also content-free in the same sense. This is why deductive logic and probability theory have traditionally been the main technical tools in philosophy of science.

If science worked by logic, and logic alone, then this would be valuable to know and understand. For it would mean that someone familiar with one science could immediately understand many other sciences. It would be like having a universal grammar that applies to a wide range of languages.

The question is: How deep and general is the understanding of science that logic and probability provide? One of the conclusions of this book is that there is a tradeoff between generality and depth. More specifically, I aim to show that there is a significant depth of understanding gained by narrowing the focus of philosophy of science to the quantitative sciences.

A Primer on Logic

Logic and probability are the standard tools of philosophy of science. Probability can be seen as an extension of logic, so it is important to understand the basics concepts of logic first.

Logic has many branches. The best known branch of logic is called *deductive* logic. Briefly, deduction is what mathematicians do, except when they use simplifying approximations, which happens a lot in science. Nevertheless, genuine deduction is always an important part of mathematical derivations.
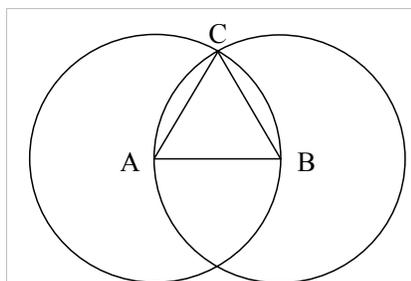

Figure 1.1: Euclid's construction of an equal-sided triangle. Draw a circle centered at A. Mark any point on the circumference as B. Draw another circle of the same radius centered at B. Mark on of the points of intersection of the circles as C, and draw lines connecting A, B, and C.

Example: The first theorem Euclid's *Elements* provides a good example of the kind of deductive reasoning that people admire. Suppose

---

[1] This is intended to be a characterization of *general* philosophy of science. Philosophy of science also includes studies in the foundations of science, which legitimately narrow their focus to particular sciences. From this point on, when I speak of philosophy of science, I mean general philosophy of science.

we construct a triangle in the following way (see Fig. 1.1): 1. Draw a circle centered at point A. Mark a point B on the circumference and draw a line from A to B. Draw a second circle centered at B that passed through A. Mark one of the points at which the circles intersect as C and draw lines from C to A and from C to B.

Theorem: All the sides of the triangle ABC are of equal length.

Proof: Let |AB| denote the length of the line segment AB, and so on.

Step 1: |AB| = |AC| because they are radii of the circle centered at A.

Step 2: |BA| = |BC| because they are radii of the circle centered at B.

Step 3: |AB| = |BA| because AB and BA denote the same line.

Step 4: |AC| = |BC| because they are each equal to the same thing (viz. |AB|).

Step 5: Therefore, |AB| = |AC| = |BC| by steps 1 and 4.

Definition: An *argument* is a set of claims, one of which is the conclusion and the rest of which are the premises.

The *conclusion* states the point being argued for and the *premises* state the reasons being advanced in support the conclusion. They may not be good reasons. There are good and bad arguments.[2]

Remark 1: Each of the five steps in the proof to Euclid's first theorem is an argument. The conclusions in steps 1 to 4 are called *intermediate* conclusions, while the conclusion in step 5 is the *main* conclusion.

Remark 2: All arguments, or sequences of arguments, are examples of reasoning, but is every piece of reasoning an argument? A perceptual judgment such as "I see a blue square", or the conclusions of elementary particle reading in bubble-chamber photographs, or scientists looking through a microscope, may be examples of reasoning that are not arguments. They are derived from what Kuhn (1970) called tacit knowledge, acquired through training and experience (like knowing how to ride a bicycle). It is not easily articulated in any language.

In deductive logic, we assume that the evaluation of an argument involves two questions:

(A) Are the premises true?
(B) Does the conclusion follow from the premises?

As an example, compare the following arguments.

(1) All planets move on ellipses. Pluto is a planet. Therefore, Pluto moves on an ellipse.

(2) Mercury moves on an ellipse. Venus moves on an ellipse. Earth moves on an ellipse. Mars moves on an ellipse. Jupiter moves on an ellipse. Saturn moves on an ellipse. Uranus moves on an ellipse. Therefore, Neptune moves on an ellipse.

(3) Mercury moves on an ellipse. Venus moves on an ellipse. Earth moves on an ellipse. Mars moves on an ellipse. Jupiter moves on an ellipse. Saturn moves on an ellipse. Uranus moves on an ellipse. Therefore, all planets move on ellipses.

Arguments (2) and (3) fare better under question A than question B. Argument (1) has the opposite property—it fares better with respect to question B than question A. In fact, in argument (1), the conclusion actually follows from the premises. This notion is captured by the following definition.

---

[2] To identify arguments look for words that introduce conclusions, like "therefore", "consequently", "it follows that". These are called *conclusion indicators*. Also look for *premise indicators* like "because" and "since".

Definition: An argument is *deductively valid* if and only if it is *impossible* that its conclusion is false while its premises are true.

According to the definition, argument (1) is deductively valid, while arguments (2) and (3) are deductively invalid. Every argument is either valid, or it is invalid. There are no shades of gray.

It is important to understand that the validity is not another name for 'good'. In particular, valid arguments can have false conclusions. This leads to the important distinction between a sound arguments and valid arguments. A *sound argument* is one that is valid *and* has true premises. The conclusions of a sound argument do have to be true.[3] Validity is something weaker than soundness.

The notion of deductively validity is such a ubiquitous concept, which goes by several names. When an argument is deductively valid, we say that the conclusion *follows from* the premises, or the conclusion *is deducible from*, or *proved from* the premises. Or we may say that the premises *imply*, or *entail*, or *prove* the conclusion. We also talk of deductively valid arguments as being *demonstrative*. All these different terms mean exactly the same thing. None of them mean the same thing as 'soundness', which is a far stronger notion.

It is useful to restate the definition of validity in equivalent forms. One restatement is: An argument is *not* deductively valid (that is, deductively *invalid*) if and only if it is *possible* that its conclusion is false while its premises are true. The description of a possible situation in which the premises are true and the conclusion is false is called a *counterexample* to the validity of the argument. So, a third restatement of the definition is: An argument is deductively valid if and only if it has no counterexamples. A fourth restatement, which is useful in probability theory, is in terms of the concept of a possible world. An argument is *deductively valid* if and only if there are no possible world in which the premises are true and the conclusion is false. This is more perspicuously stated as: An argument is *deductively valid* if and only if all the possible worlds in which the premises are true are worlds in which the conclusion is true.

The claim that an argument is (deductively) valid is that same as the claim that the conjunction of all its premises entails the conclusion, where a conjunction of a set of claims is formed by joining the claims with 'and'. Let P denote the conjunction of the premises, and let Q be the conclusion. The argument is valid if and only if P entails Q, which is symbolized as $P \Rightarrow Q$ in this book. Entailment has a useful pictorial representation in terms of the possible worlds. Let the points inside the rounded rectangle represent possible worlds. The points inside the circle labeled P represent possible worlds at which P is true, and those outside the circle are worlds at which P is false. There no possible worlds

Figure 1.2: The fact that P entails Q has two equivalent representations in a possible world diagram. In both cases, the set of P-worlds is a subset of the set of Q-worlds.

represented by the points on the circle itself. A similar story applies to Q-worlds. The fact that $P \Rightarrow Q$ implies that no point inside the shaded region in Fig. 1.2 represents any
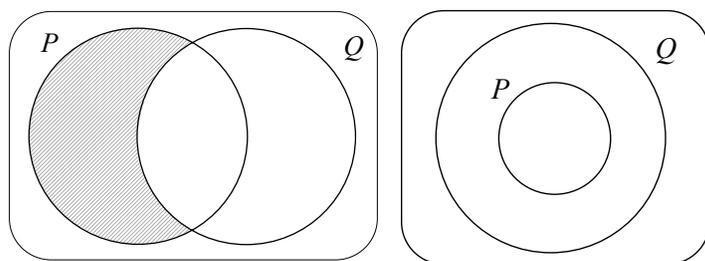
---

[3] Proof: Consider a sound argument. Given that an argument is valid, it is impossible for the premises to be true and the conclusion false. But the premises are true. So, it is impossible for the conclusion to be false. That is, the conclusion must be true.

possible world. That means that all possible worlds at which *P* is true are also *Q*-worlds, which is equivalently represented on the right-hand diagram. Thus, the entailment relation corresponds to the subset relation in the possible worlds representation. Notice that that is all it depends on. It does not matter which world is the actual world.

The sense of "possible" needs some clarification. Consider an example:

(4) Peter is a human being. Peter is 90 years old. Peter has arthritis. Therefore, Peter will not run a sub-four-minute mile tomorrow.

Suppose that the premises are true. It is *physically* impossible that Peter will run a four-minute mile tomorrow even though. But logicians have a far more liberal sense of what is "possible" in mind in their definition of deductive validity. In logic, it *is* possible that Peter will run a sub-four-minute mile tomorrow. So, argument (3) is deductively invalid. The standard for what counts as a valid argument is high, but the standard is met in mathematics (though not always because many mathematical derivations, especially in applied mathematics, rely on methods of approximation).

In any deductively valid argument, there is a sense in which the conclusion is *contained* in premises. Deductive reasoning serves the purpose of *extracting* information from the premises. In a non-deductive argument, the conclusion 'goes beyond' the premises. Inferences in which the conclusion *amplifies* the premises is sometimes called *ampliative* inference.

Therefore, the property of deductive validity, depends on what claims are included in the list of premises.

Missing premises: We can always add a premise to 'turn' an invalid argument into a valid argument. For example, if we add the premise "No 100-year-old human being with arthritis will run a four-minute mile tomorrow" to argument (3), then the new argument is deductively valid. (The original argument, of course, is still invalid).

Arguments (2) and (3) show that some ampliative inferences are pretty good, even though they are not sound (because they are not valid). Many philosophers of science have tried to say why some ampliative inferences are better than others. A major class of ampliative inferences in science have the fortunate property that all the premises of the argument are statements of the evidence, so that question (A) is answered in the affirmative—the evidence is true. So, the question comes down to how well those premises support the conclusion. This raises what is known as the *problem of confirmation*.

Does Science have a Logic?

The three examples of the previous section were chosen to represent typical kinds of arguments that might arise in scientific inferences. Notice that the conclusion of argument (2) makes a prediction about a particular planet. Since it does not appear to matter which new planet the prediction is about, the problem of understanding the evidential support of generalizations (argument (3)) looks no more difficult than understanding the evidential support of an arbitrary instance of the generalization. So, on the standard view, the fundamental problem is to understand the evidential support of scientific laws, hypotheses, or theories. Or to put is another way, if we could understand the evidential support for the truth of general hypotheses, then the evidential support for predictions made from those hypothesis would fall out as a simple corollary. I reject this assumption in this book. But it is tacitly assumed in all other theories of confirmation.

According to *inductivism*, the problem is to understand how a hypothesis like "All planets move in ellipses" is supported by, or confirmed by, the fact that all the instances

of the generalization observed *so far* have been true. One inductivist strategy is to ask what missing premise must be added to the argument to make it deductively valid, and then to evaluate the truth of the added premise. In our example, we could add the premise "Whatever is true of the observed planets is true of all planets". This is a statement of the *uniformity* of nature. The problem with this statement is that it is plainly false! Not everything that is true of the observed planets is true of all planets. For one thing, the observed planets are closer to the sun than Neptune, or Pluto. The problem is that nature is uniform is some respects, but not in others.

When we try to refine the statement of the uniformity of nature to make it true, there is a danger of ending up with something like "The orbits of the observed planets have the same shape as the orbits of all planets". But now the argument is circular, because whatever reason we have for believing that the added premise is true is that all the planets move on ellipses.

Of course, inductivists are well aware of this issue, and some have tried to find a set of missing premises that are self-evident, or true *a priori*. Immanuel Kant and William Whewell made sophisticated attempts to justify the truth of Newton's theory of motion in this way. Obviously, the post-Newtonian revolution in physics showed that at least one of their axioms is false. This failure does not prove that the same idea cannot work for the new theories of relativity and quantum mechanics. The more telling point was that the strange and unintuitive nature of the new theories convinced many philosophers that our a priori intuitions about the uniformity of nature are unreliable.

Bayesian philosophers of science measure the strength of an ampliative inference in terms of probability. If *E* is a statement of the evidence, and *H* is a statement of the law, hypothesis, or theory, then the strength of evidential support is the probability that *H* is true *given* the evidence *E*. While the theory doesn't tell us what this probability value must be, the theory of probability does impose interesting constraints on the logic of the confirmation relation. I am somewhat ambivalent towards the Bayesian approach. On the one hand, I think that Bayesianism has severe limitations. On the other hand, I plan to show that a restricted version of Bayesianism falls out as a special case in the approach that I develop in this book. I will make this precise in later chapters.

Hypothetico-deductivists approach the problem differently. They do not want to grapple with the problem of evaluating ampliative inferences at all. For them, confirmation is about *testing* hypotheses by *deducing* predictions from them, and seeing whether the predictions are true. Likelihoodists replace deduction with a probabilistic relation: The probability that the evidence is true given that the hypothesis is true. Note that this approach is a different from Bayesianism, which focuses on the probability that the hypothesis is true given the evidence.

Karl Popper's falsificationism is a version of hypothetico-deductivism. It begins with the universally accepted point that no amount of evidence can conclusively prove that any scientific hypothesis is true because there are always untested predictions[4]. Popper's insight is that it is nevertheless possible to conclusively prove that a hypothesis is *false*. Consider some prediction *P* that is entailed by a hypothesis *H*. If *P* is false, then *H* has been proved false. This follows from the definition of entailment, as follows. If *H* entails *P*, then it is impossible for *H* to be true and *P* false. So, if *P* is false, it is impossible for *H* to be true. That is, *H* is false. The remaining problem for Popper is to compare the many hypotheses that are unfalsified by the current evidence. Popper claims

---

[4] Hypothesis *H predicts P* if and only if *H* entails *P* and the true or falsity of *P* can be determined by observation. Note that *P* can refer to a past, present, or future event. This the hypothetico-deductive definition of "prediction", which will be modified in the course of this book.

that we should favor those hypotheses that are highly falsifiable, yet unfalsified, and say that they are best confirmed by the evidence (or best corroborated, as he put it). In other words, science should make bold, falsifiable conjectures, and subject them to severe tests. If they "stick their necks out" without getting their "necks chopped off", then they count as the best scientific hypotheses we have. Again, I plan to argue that this idea has important limitations.

The notion of confirmation investigated in this book differs from all these theories in the following way. It rejects the one assumption that is common to all of them. It is the assumption, already mentioned, that the primary problem is to judge the evidential support of the truth of a hypothesis, and that the evidential support for the *predictions* of a hypothesis rests on the extent to which the hypothesis is confirmed *as true*. I plan to turn this on its head. For me, the primary problem is to evaluate the evidential support for the *predictions* of hypotheses, and the confirmation of the hypotheses themselves derives from that evaluation.   So, the fundamental measure of confirmation is an estimation of the predictive accuracy of a hypothesis. There are many predictive accuracies because there are many predictions that may be considered. As a consequence, the evidential relationship between a hypothesis and its evidence is a multifaceted relationship, which means that the confirmation of a hypothesis is equally multifaceted. There is no such thing as *the* confirmation of a hypothesis measured by a single number—different aspects of the hypothesis receive different evaluations. Or to put it differently, hypotheses are not really the objects of confirmation at all.

This shift in focus addresses a vexing problem in confirmation—the problem of irrelevant conjunctions. Consider the very simple deductive theory of confirmation: Observed evidence E confirms hypothesis H if and only H entails E. The problem is that (H and X) also entails E, where X is hypothesis that is entirely irrelevant to E and H. For instance, let H be Galileo's law of free fall, and X be a theory about bee dancing, where E is data about the motion of projectiles, such as cannonballs. Then, according to the deductive theory of confirmation, E confirms (H and X). By itself it is not an unpalatable consequence of the theory, but if it is coupled to the simple-minded assumption that we should trust the predictions of any confirmed hypotheses, to some degree, then there is a problem. For the predictions of (H and X) include all the predictions of X, which are totally unsupported by E.

For me, the solution is to rely on a more the more detailed relationship between theory and evidence, which includes the irrelevance of X. X has no verified predictive accuracy, and so the predictive accuracy of (H and X) is due to H. The confirmation of a hypothesis should be limited to those aspects of the hypothesis that are responsible for its empirical success.   The "whole truth" of a scientific hypothesis is not what is confirmed. What's confirmed is actively read into a hypothesis, rather than read off of a hypothesis. It's the *whole* of science, including the detailed findings of experimental science, that gives us the best confirmed picture of reality. It's misleading to focus exclusively on our best current *theory* without considering its relationship with the evidence.

As a simple example, Leibniz criticized Newton's view of absolute space, which Newton thought of as an independently existing "vessel" that "contains" the matter in the universe. In Newton's theory, the sun had some absolute velocity relative to this "container space". The puzzling fact was that according to Newton's own theory there is no way of measuring this velocity, because supposing that the velocity of the sun has one value as opposed to another leaves all accelerations the same.   This cut two ways. On the one hand, Newton's assumption was harmless in the sense that the predictions of the theory about relative motions are the same in either case. On the other hand, it postulated the existence of quantities that were unmeasured.

It is conceivable that post-Newtonian physics could have provided the means for measuring Newton's "hidden variables". For example, subsequent developments in electrodynamics during the 1800s led to Maxwell's theory of electromagnetism, according to which electromagnetic waves (such as light) propagated in an invisible ether. The ether could have been Newton's absolute space, and the detection of the motion of the ether would then have been measurements of Newton's hidden variables. But this would not have changed the fact that there was a part of Newton's theory—his conception of absolute space—that was irrelevant to the predictive success of Newton's theory in Newton's time.

Naturally, simple ideas tend to be more complicated that they appear at first sight, which is why I've written a whole book on the subject. In the end, I defend the view that science has a 'logic', albeit one that reaches beyond deductive logic into the murkier realm of statistical inference.

Middle Ground between 'Trivial' and 'False'

Every discipline begins with the simplest ideas and takes them as far as they can go. The philosophy of science is no exception. Here I recount two simple answers to the question: How does science work? Neither answer is adequate, albeit for importantly different reasons. The first answer is precise, but wrong. The second seems right, but only because it is very vague.

The first answer is that science accumulates knowledge by *simple enumerative induction*, which refers to the following pattern of reasoning: All billiard balls observed so far have moved when struck, therefore all billiard balls move when struck. The premise of the argument refers to something that we know by observation alone: All billiard balls observed so far have moved when struck. It is a statement of the empirical evidence. The conclusion extends the observed regularity to all unobserved instances. In particular, it predicts that the next billiard ball will move when struck. Like any scientific theory or hypothesis, the conclusion makes a prediction that extends beyond the evidence. Simple enumerative induction is an instance of what is called *ampliative inference* because the conclusion 'amplifies' the premises.

As Hume pointed out, it is a fact of logic that any form of ampliative inference is deductively invalid. No matter how many times the billiard balls have been observed, and no matter how varied the varied the circumstances, it is *possible* that the next billiard ball will not move when struck (there is such a thing as superglue, and even if there weren't, we could imagine such a thing). Since *any* ampliative inference is fallible and scientific inference is a form of ampliative inference, it follows that science is also fallible. Our very best scientific theories may be false. So, it's not the fallibility of simple enumerative induction that limits its usefulness in philosophy of science. If anything, it provides a weak argument *for* simple enumerative induction, which runs as follows: Science is fallible, simple enumerative induction is fallible, therefore science is simple enumerative induction because this explains why science is fallible.

Instead, the main objection is that simple enumerative induction fails to provide any place for *new concepts* in science. The pattern of simple enumerative induction is: All observed *A*'s have been *B*'s. Therefore, all *A*'s are *B*'s. The only terms appearing in the conclusion are *A* and *B*, which are observational terms. Newton's theory of gravitation introduced the concept of gravitation to explain the motion of the planets. Mendel introduced the concept of a gene to explain the inheritance of observable traits in pea plants. Atomic theory introduces the notion of atoms to explain thermodynamic behavior. Psychologists introduced the notion of the intelligence quotient (IQ) to explain

the correlation of test scores. These quantities are not observed, at least not directly. Newton did not see gravity pulling on planets, Mendel did not observe any genes, and Boltzmann did not see molecules in motion. If these scientists were limited to simple enumerative induction, then their theories could not refer to any of those things. Simple enumerative induction is a precise theory of how science works, but it is false.

The same objection applies to any kind of inductive inference in which the conclusion is constructed from the observational vocabulary of the premises. In fact, I shall refer to an ampliative inference with this restriction as an *inductive inference*. Inductive inferences are ampliative, but not all ampliative inference are inductive.

Inference to the best explanation is the name of a non-inductive kind of ampliative inference. Explanations are free to postulate the existence of unobservable things, properties, and quantities. So, this picture of scientific inference allows for conceptual innovation in science. In fact, it can account for almost anything, and herein lies its fault—its vagueness. For if nothing more is said about what counts as an explanation, and what counts as 'best', then practically *any* example of science may be described as an inference to the best explanation. It does little more than replace one mystery by another. This theory of how science works is vaguely right, but until its vagueness is removed, it has very little philosophical merit.[5]

The hard problem is to find firm middle ground between what's trivially true and what's obviously false. Philosophers aim to provide a theory of science that *deepens our understanding* of how science works. I do not want to leave the reader with the false impression that philosophy of science has done no better than the two attempts sketched above. The point is that vague theories may lull their proponents into a false sense of achievement.

At the beginning of the 20th century, there was a movement in philosophy called logical positivism. One of their key philosophical tenets was a rather anti-philosophical doctrine, anti-metaphysical. Their idea was that theoretically postulated quantities in science are either meaningless, or definable in observational terms. Given that the meaningless explanations are not good explanations, their view would be that inference to the *best* explanation can be reduced to some kind an inductive inference.

In contrast, the view that I develop in this book does not assume that theoretical quantities are definable in terms of observational quantities, but it does make a precise distinction between those theoretical quantities that are well grounded in the evidence and those are not. Newton's postulation of absolute velocities is a simple example of theoretical quantities that were not well grounded in the evidence.
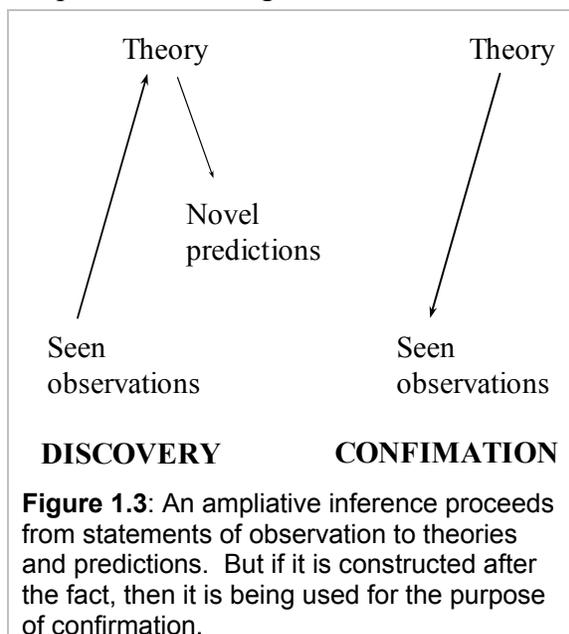
There are other well developed philosophies of science that do a reasonably good job at trading off generality and depth. The first uses deductive logic as a tool, and is called *hypothetico-deductivism*. Earlier versions of Popper's methodology of *falsificationism* were hypothetico-deductive. Bayesianism can be seen as a probabilistic generalization of falsificationism (Earman 1992). Both of these approaches derive some important insights from some simple first principles.

Yet I believe that there is a sense in which they do their job *too* well. These views of science are *so* content-free that they make very little distinction between everyday reasoning and scientific reasoning, which raises the question: Is there anything special about scientific reasoning at all? If these philosophies of science are to be believed, then the answer is "not much". At one level of description, this may be true. There may be

---

[5] Of course, there are sophisticated answers to these questions (see e.g. Lipton ////). My point is only that undeveloped philosophical theories sometimes are too well respected simply because they fit examples better (witness, for example, the literature produced by Harman's (1965)).

features common to all rational forms of reasoning. But at a slightly deeper level of description, I shall argue that there is much to be gained by looking at a methodology that is not a part of everyday reasoning. The methodology of curve fitting is what I have in mind.

The view of curve fitting that I develop in chapter is something more than inductive in nature, for it allows for, in fact mandates, the introduction of theoretical quantities in the form of adjustable parameters. The goal is to introduce as few as possible, so they their measurement is as fully overdetermined by the empirical data as possible. In this way new empirical relations between disparate phenomena are exposed, which provides a more unified understanding of phenomena, which is the role intended for explanations. The difference is that the theoretical quantities play an indispensable role in uncovering the new features of the evidence. Conceptual innovation in science is more than a psychological phenomenon—it play a role in the confirmation of theories.

Theory   Theory

Novel
predictions

Seen   Seen
observations   observations

DISCOVERY   **CONFIMATION**

Figure 1.3: An ampliative inference proceeds from statements of observation to theories and predictions. But if it is constructed after the fact, then it is being used for the purpose of confirmation.

Discovery versus Confirmation

In 1865 Friedrich Kekulé dreamed of a snake biting its tail. Inspired by that vision, Kekulé invented his theory of the six-carbon benzene ring, which helped to explain many of the facts of organic chemistry known at the time.[6] If Kekulé were asked to *justify* his theory, I doubt that he would mention his dream. Discovery is distinct from justification, or confirmation.[7]

The core thesis of hypothetico-deductivism is that the hypothetico-part of the process is a psychological process, *and therefore has nothing to do with philosophy of science*. Discovery, in other words, is irrelevant to the philosophy of science. The deductivism-part refers to the deduction of observational statements from the theory, which are then compared with the observed facts. The deductivism part is about the *justification*, or *confirmation* of hypotheses, which is the central concern of a philosophy of science.

If we think back to the theory of simple enumerative induction, the distinction between discovery and justification is blurred. At first sight, it looks like a method of discovery. From the observation of a number of *A*'s that are *B*'s, and none that aren't, we discover that all *A*'s are *B*'s. Along side this, it provides a theory of confirmation, as well—if a hypothesis is the conclusion of a simple enumerative induction then it is justified by the evidence stated in the premises of the argument, otherwise not. The fact that the discovery of new theories typically involves the invention of new concepts is the

---

[6] See "Kekule von Stradonitz, (Friedrich) August," *Encyclopedia Britannica Online* <http://www.eb.com:180/bol/topic?eu=46047&sctn=1>.

[7] 'Discovery' is a success term, which means that we don't say that someone discovered *X* when *X* doesn't exist, or when *X* is not true. In order to allow for the "discovery" of useful falsehoods in science, such as the now discredited hypothesis that heat is a fluid, we could use the term 'theory invention' or 'theory construction', instead. But the common practice to use 'discovery' as a catch-all term.

reason that induction is uninteresting, both as a theory of discovery and as a theory of confirmation.

Inference to the best explanation is also a kind of inference. Yet it is clear that it does not provide a usable methodology of discovery. You need to have the theories in your mind before you can ask how well they explains the facts. Explanation is a "top-down" process (Fig. 1.3, right). So, inference to the best explanation is introduced as a part of a theory of confirmation, even though it is not always clear what the theory is meant to be.

If the "top down" process of explanation is exhausted by the *deduction* of the facts to be explained from the theory, then the resulting theory of confirmation is a type of the hypothetico-deductivism. If the explanatory relation doesn't have to be deductive, or if other criteria are used such as simplicity or unification, then it is not purely hypothetico-deductive. In either case, inference to the best explanation is still part of a theory of confirmation.

In sum, discovery plays no essential role in philosophy of science to the extent that it is merely a psychological process. Our interest is in justification and confirmation, and it is seen as a kind of genetic fallacy to judge a theory in terms of its origin. Similarly, the fact that we find something surprising is also a fact about our psychology. For example, Poisson pointed out that Fresnel's wave theory of light implied that there should be a bright spot at the center of a small shadow cast by a small circular obstacle of the right size. Poisson saw this as a necessary but absurd consequence of Fresnel's theory. But the consequence is not absurd, for when the experiment was carried out, such a spot was in fact found. The result was clearly surprising to Poisson, but does this psychological surprise factor increase the confirmation of the observation? It may well be a psychological fact that it impressed Poisson. But is this relevant to the confirmation of Fresnel's theory?

Perhaps it's not the element of surprise that it relevant to confirmation but only the fact that Fresnel's theory predicted the phenomenon *in advance*. Evidently, he did not construct his theory in order to make the correct prediction. But if the historical order of events is relevant, then confirmation depends on something more than the logical relationship between a theory and its evidence. It depends on the *order* of discovery, and so we are back to the question whether discovery and confirmation are completely separate issues.

Historical versus Logical Theories of Confirmation

Hempel (1966, 37) claims that "it is highly desirable for a scientific hypothesis to be confirmed… by 'new' evidence—by facts that were not known or not taken into account when the hypothesis was formulated. Many hypotheses and theories in natural science have indeed received support from such 'new' phenomena, with the result that their confirmation was considerably strengthened."

The discovery of the so-called Balmer series in the emission spectrum of hydrogen gas in 1885 is one example Hempel considers. J. J. Balmer constructed a formula that reproduced the values of λ for $n = 3, 4, 5,$ and 6 as follows:

$$\lambda = b \frac{n^2}{n^2 - 2^2},$$

The constant b is an adjustable parameter in Balmer's model, which he found to be approximately 3645.6 Å by fitting his formula to the 4 data points (4 pairs of values of n and λ). Balmer's formula then predicts the value of λ for higher values of n. He was

unaware that 35 consecutive lines in the series had already been measured, and that his predicted values agreed well with the measured values.[8]

It is uncontroversial to say that the agreement of Balmer's predictions with the unseen data confirmed Balmer's hypothesis. Yet, as Hempel notes, a puzzling question arises in this context.[9] What if Balmer's model had been constructed with full knowledge of all 35 lines of the Balmer spectrum? In this fictitious example, the model is the same and total data is the same, and so the logical relationship between them is the same. The only difference is the historical order of events. If confirmation is a logical relationship between theory and evidence, then historical circumstances should make no difference. Yet many people have the intuition that the confirmation is stronger in the actual case than in the fictitious case. Is this intuition correct?

Care must be taken to understand what the logical theory of confirmation implies. It claims only that historical facts are relevant to the assessment of the confirmation *if* the *all* the relevant details of the logical relationship the theory and evidence are specified. If only some of the relevant logical details are specified, then the historical circumstances may be relevant. For instance, suppose that we were only told that Balmer has a formula that fits all 35 data points very well, without being shown the formula, and with being told how many adjustable parameters Balmer used. In this case we have reason to question the significance of Balmer's discovery because it is possible that Balmer used a formula with 30 adjustable parameters. Anyone with enough time and patience can find a formula to fit 35 data points. In fact, there is a very easy recipe for doing so—just assign one adjustable parameter to each point. The problem is the fit between the hypothesis and data is "fudged". Therefore, the number of adjustable parameters used to fit the data is a *relevant* part of the logical relationship between the evidence and the hypothesis. The degree of fit achieved by the formula is an incomplete description of the relevant logical facts.

To change the example a slight but important way, suppose that we don't know how many adjustable parameters Balmer used, but we are told that he knew of only 4 data points. This historical information is relevant, even on the logical theory. But, the example does *not* refute the logical theory of confirmation, because the historical facts tell us something relevant about the logic of the example; namely, that Blamer did not use more than 4 adjustable parameters (for no more than 4 adjustable parameters can be determined from 4 data points). Therefore, the logical theory allows that historical facts are relevant to judging confirmation. For an example to be a genuine counterexample to the logical theory, two conditions must be met: (Logical Completeness) All the relevant logical facts of the example are specified, and (Historical Relevance) the historical circumstances are relevant to confirmation.

The important point is that the logical theorists must not be short-changed on Logical Completeness condition. An opponent might believe that the only relevant logical fact is the degree of fit between the formula and the total evidence. If this were granted, then the logical theory would be easy to refute. The example in the previous paragraph would be a counterexample. The correct response is that the example is not logically complete.

The logical theorist can make this response to any alleged counterexample, although the logical theorist must respond by saying exactly what logical details have been omitted. In the previous example, this challenge was met: The example fails to specify the number of adjustable parameters used to fit the formula.

---

[8] For more detail, see Chapter 4, Holton and Roller, 1958.
[9] For further discussion, see Musgrave, 1974.

13

In the Balmer example, I concede that the Logical Completeness condition is met. We are given the exact formula, and we see that it has only one adjustable parameter. We then move onto the question about historical relevance. My intuitions are that the historical circumstances are irrelevant. The logic of the example tells me that I would be equally impressed if Balmer had known all the data, and then found a formula with only one adjustable parameter that fit the data.

My claim is that I would be *equally* impressed in both the actual and fictitious versions of the example. My claim is not that I am 100% convinced that the formula must fit new data for n greater than 37. It is logically possible that new data would not fit Balmer's formula, and that there the correct formula is a little more complicated than Balmer supposed. It doesn't matter whether we assign high or low credence to this possibility. The fact remains that we would be fooled equally by the data in either circumstance. Balmer's prediction in advance does no more work or no less work in discounting this possibility than if we were to begin with all 35 data points.

The fact that all scientific hypotheses are underdetermined by their evidence is not a problem for any theory of confirmation. In fact it is an advantage, for it explains why Bohr's derivation of Balmer's formula from his quantum theory of the hydrogen atom in 1913 provided *additional* support for the formula. For Bohr's theory was independently supported by diverse evidence other than spectroscopic measurements.[10]

We have seen that a logical theory of confirmation can explain why history is relevant to assessing the confirmation of a theory. But is the opposite true? Can the historical theory explain those cases in which predictions are not made with full knowledge of the facts to be prediction and are still perceived as major successes for the theory?

One famous example is Einstein's prediction of the precession of the perihelion of Mercury, which the Newtonians had failed to explain for centuries.[11] The planet Mercury has the largest observed precession, of 574 seconds of arc per century. In Newtonian mechanics any precession of a planet's perihelion requires that the effective radial dependence of the net force on the planet be slightly different from $1/r^2$, where r is the distance from the sun. This is effectively what happens when the gravitational influence of the other planets in added to that of the sun. However, detailed Newtonian calculations of that effect predict it to be approximately 531 seconds of arc per century, which fails to account for 43 seconds of arc. The discrepancy was outside the bounds of observational error.

Einstein's general theory of relativity *predicted* the residue correctly. Yet the Einsteinian model was constructed with full knowledge of the correct value. The important fact of the case is that the derivation was open to inspection to anyone who could follow the mathematics. The model was the simplest one possible, treating Mercury as a 'test' particle moving in the spherically symmetric gravitational field generated by the sun. Nobody questioned the significance of this famous test of relativity simply just because it was not predicted in advance. The experts could verify that the calculations were not fudged, because the derivation depended on the simplest possible auxiliary assumptions. And it used an exact solution to Einstein's equations called the Schwarzschild solution, so not tenuous approximation assumptions are introduced in the derivation.

If the value of the precession of Mercury had been predicted in advance, then the general public could have been impressed without having to trust the judgment of

---

[10] See Hempel 1966, 39 for further discussion.

[11] See Marion (1965) for a more detailed account of the example.

experts. This is what happened in the case of Clairaut's prediction of the return of Halley's comet in 1759. In that example, it was not the mere return of the comet that impressed. For it's doesn't take rocket science to predict that a comet observed in 1531, 1607, and 1682, will return in 1759, because it is clear the period of motion is approximately constant. But the simple extrapolation actually predicted that Halley's comet would reach its perihelion (the closest point to the sun) in the middle of 1759. The extraordinary fact was that Clairaut predicted that Halley's comet would actually return several months early, near the beginning of 1759. His prediction was based on calculations of the gravitational effects of Jupiter and Saturn on the comet.

This is an interesting example, because there is a case to be made that the prediction in advance was informative even to the experts. For unlike Einstein's prediction of the perihelion of Mercury, Clairaut used a method of mathematical approximation that is not logically airtight. The prediction in advance had some role in confirming the assumption that the terms omitted in calculation were indeed negligible. But again, the relevance of the historical order of events is fully explained by the logical theory.

There are other apparent counterexamples of the logical theory. In this imaginary example, the logic is very simple and therefore completely specified. Imagine that we wish to confirm the hypothesis that all sodium salts burn yellow.[12] We freeze seawater and place it under a hot flame, and see that the flame does *not* burn yellow, contrary to our expectations. Our initial reaction is that we have refuted the hypothesis, because seawater contains sodium salts, and it did not burn yellow. But further investigation reveals that when salt water freezes, the ice is salt-free. So, the apparent refutation now appears to *confirm* the hypothesis. Now compare this with the situation in which we already know in advance that the ice contains no sodium. In this situation, it seems be irrelevant whether the flame is yellow or not. If our intuitions are correct, it appears that the historical order in which we learn the full facts of the case is relevant to our final assessment of the confirmation. So, the counterexample appears to be genuine.

Here I think the logical theory is saved by a combination of two facts. The first is that in the original story, we have every reason to believe that the hypothesis has been refuted, so our *psychological* estimation of the degree of confirmation goes down—way down. Then we learn a psychologically surprising fact about the way that seawater freezes. In light of that fact, our psychological estimate of the degree of confirmation goes back up. It is easy to confuse the *change* in confirmation with the absolute degree of confirmation. This is where the logical theory can be misunderstood. Its principle aim is to assess the degree of confirmation in light of the total evidence available at a particular time. Of course, the logical theory also says that the degree of confirmation changes if the evidence changes. In this example, the assertion that the ice contains salt is taken to be a fact, even though it is false. Any theory of confirmation must recognize that statements of evidence are often not statements of observational fact, but inferred from other theories, and they may be overthrown. While this complicates the example, it also save the logical theory from refutation once the situation is fully understood.

The most important argument of this section that any genuine counterexample to the logical theory of confirmation must specify *all* the logical features that are relevant to confirmation. But what counts as relevant is up to the theory to say. Therefore, there will be no clear resolution of this debate without a clearer understanding of what is relevant to the logic of confirmation.

---

[12] The example is adapted from Hempel (1965, 19).

The Indirect Confirmation of Boyle's Law

Between the time of Galileo and Newton, around 1660, Robert Boyle investigated how much a pocket of air is compressed by the pressure exerted upon it. Boyle's experiment is touted to be one of the great experiments of all time according to the citations it receives in some textbooks (e.g., Shamos 1959 and Harré 1981). Alongside his data, he published his famous gas law, $PV =$ constant, where P is the pressure on the gas and V is the volume of the gas. The law was later superceded by the Boyle-Charles law, which adds the assertion that the constant is proportional to the temperature of the gas. When the Boyle-Charles law was derived from the atomic theory of gases the 1800's, the result became known as the ideal gas law, written $PV = NkT$, where N is the number of molecules in the gas and k is a universal constant called Boltzmann's constant. The ideal gas law predicts much more than the Boyle-Charles law because it entails *Avogadro's law*: "For equal temperature and pressure all gases contain an equal number of molecules per unit volume" (Khinchin 1949, 121).
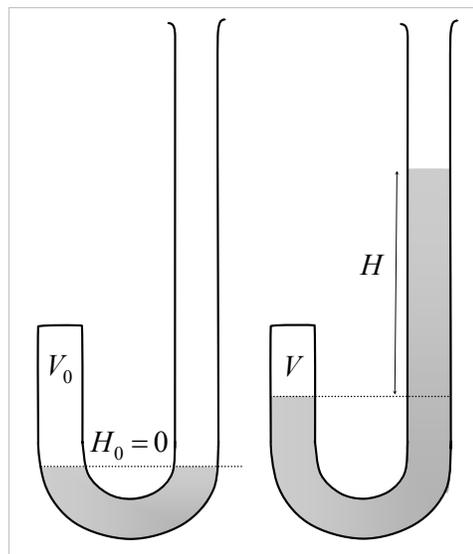

Figure 1.4: Boyle's apparatus. The J-shaped tube is first filled with enough mercury to trap a certain volume of air on the left hand side. Then more mercury is added, slowly so that none of the air escapes. The trapped air is observed to decrease in volume.

Avogadro's law is what enabled us to discover that water is H_2O at high school. Ignite an enclosed mixture of hydrogen and oxygen molecules so that the ratio of atoms is 2 to 1. After combustion, we are left with water and no gas. So, Boyle's discovery was the first step in a long journey that eventually led the confirmation of the atomic hypothesis and the birth of modern chemistry. Yet Boyle's argued for his law, quite convincingly, without the benefit of hindsight wisdom. What was his argument?

How did Boyle use the known evidence to justify his law? The kind of answer invited by a naive inductive picture of science is that Boyle inferred his law from the results of his experiment, and the strength of the inference is what measures its confirmation. The analysis of the logic of this example suggests that the inductivist's picture is misleading, because the data that is minimally sufficient to suggest the correct form of the law excludes the most important evidence for the law.

Imagine that you build a time machine, and travel back to month before Boyle does his experiment in an attempt to replace his name with yours in the history books. You take a J-shaped glass tube of uniform cross section and trap a pocket of air in the closed end of the tube, separating it from the open air by liquid mercury at the bottom of the tube. By letting enough air escape, the initial volume, V_0, is exactly 12 inches times the area of the cross section of the tube when the level of the mercury is the same on both sides (Fig. 1.4, left). The 12 inches occupied by the trapped air are marked in ¼ inch increments. Now an assistant pours additional mercury into the right side of the tube very slowly, until the volume of the trapped air is decreased by a ¼ inch increment. The level of the mercury on the right side of the tube is seen to be higher than the level of the mercury on the left by an amount that you call the additional *height of the mercury*, which you label H (Fig. 1.4, right). The pair of values for V and H is recorded as the second entry in your table of results. Your first entry is (12, 0) because $H_0 = 0$ when

$V_0 = 12$. The procedure is repeated many times, until you have a table of data similar to the one actually obtained by Boyle (see the Table below). Note that the *observed* quantities are of H and V. Perhaps you think that the pressure is proportional to H because the extra height of the mercury is what is compressing the air. Your mistake is obvious to us because V times H is not even close to being constant. But *we* recognize this as a mistake only because you know that the identification of P with H does not lead to Boyle's law. You know none of this.

So you induce a law relating H to V. After some fiddling around with some simple formulae, you guess that the form of the relationship is $(A + H)V = \text{constant}$, where A is an adjustable parameter. By fitting this law to your to the data, the best value for the parameter a is around 29 inches. So, your final law is
$(29 + H)V = \text{constant}$, or equivalently,
$HV = \text{constant} - 29H$. By publishing this result, you may have set science back by more than 100 years.

Or perhaps not? Critics complain that your law does not *explain* the phenomenon you have discovered—it provides very little by way of insight or understanding. But you are quick to respond to your critics: "My explanation is the *best* explanation",

Volume of trapped air, V (inches)	Height of mercury, H (inches)	(29⅛+H)V = constant.
12	0	349.56
11 ½	1.44	351.44
11	2.81	351.34
10 ½	4.38	351.75
10	6.19	353.10
9 ½	7.88	351.50
9	10.13	353.25
8 ½	12.50	354.88
8	15.13	353.52
7 ½	17.94	352.95
7	21.19	352.66
6 ½	25.19	353.02
6	29.69	352.86
5 ¾	32.19	352.53
5 ½	34.94	352.33
5 ¼	37.94	352.07
5	41.56	353.45
4 ¾	45.00	352.12
4 ½	48.75	350.46
4 ¼	53.69	351.69
4	58.13	351.52
3 ¾	63.94	348.98
3 ½	71.31	351.54
3 ¼	78.69	350.39
3	88.44	352.77

Table: Boyle's data. The J-shaped tube is first filled with enough mercury to trap a certain volume of air on the left hand side. Then mercury is added carefully so that none of the air escapes. The air decreases its volume. Notice that it takes about 29 inches of mercy to half the volume of the gas. *Source*: Shamos 1959, 39.

which is something that you parrot from a philosophy of science class. After all, you have repeated the experiment very carefully, and on each occasion the law turns out to be the same. You are confident that that history will prove you right.

Unfortunately, the critics are not impressed. They claim that your theory is not the "best" explanation because it is not an *explanation* at all! In your defense, you cite a famous philosopher of science (Hempel 1965). Hempel has a precise definition of what an scientific explanation is.[13] This hypothetico-deductive definition says that an explanation has the form of a deductive argument in which the facts to be explained play the role of the conclusion, and there is at least one law of nature stated as a premise of the argument. All the conditions of Hempel's definition are clearly met, except the requirement that your equation is a law of nature. Hempel says that a law of nature must be *universal*, which means that it should not be restricted to any particular place and time.

---

[13] Hempel actually intended his definition to characterize the true but unknown explanation of some fact. His is a metaphysical definition. Hempel never intended that his definition should be used in an epistemological context. Nevertheless, it does well to understand the reasons for this. For a slightly different perspective from the one that follows, see the chapter entitled "Why the Truth Doesn't Explain Much" in Cartwright (1983).

Your critics argue that your law is not universal because of the appearance the number 29 inches in your law is measured solely on the basis of your data, and could be very different in different places (such as at high altitudes where atmospheric pressure is different).

Fortunately, your critics have misunderstood the Hempelian view of laws. You are quick to point out Hempel does not insist that the *quantities* appearing in the law cannot vary. After all, V varies and H varies. And so A can vary too. All that is required is that the mathematical *relationship* among the variables is universal, and your law satisfies this condition.

After pointing this out, a new complaint is that your law is too complicated to count as a genuine law. While you don't see why laws *have* to be simple, you nevertheless respond by proposing that $A + H$ is the sum of two pressures exerted on the gas—the first is the weight of the atmosphere pushing down on the mercury, and the second is the weight of the mercury, which is proportional to H. By defining $P = A + H$, you rewrite your law in the form $PV = $ constant. The critics concede that no law could be simpler than that.

Unfortunately for you, Hempel has another requirement for explanation. He insists that the laws must "cover" at least one other phenomenon besides the one that you explain. That is, the law cannot be ad hoc in the sense of being invented solely for the purpose of explaining the phenomenon from which the law is induced. And apparently it's not enough that it covers repeated instances of the *same kind* of experiment. You meet the critics' demand by investigating the *expansion* of gases, as opposed to their compression.

Figure 1.5: Boyle's second experiment, which investigated the expansion of gases.

In fact, Boyle performed the same experiment. Imagine that you allow some air to enter a straight tube, which has exactly the same cross-section as the J-shaped tube. You adjust the amount of air in the tube so that it occupies a length of 12 inches when the level of the mercury inside the tube is even with the level outside the tube (Fig. 1.5, left). The tube is now marked in ¼ inch increments *below* the current level of the mercury. You raise the tube slowly, so that the volume of the gas *increases* by ½ an inch, and record the height of the mercury, H, above the level in the trough (Fig. 1.5, right). The weight of this column of mercury is "pulling down" on the air trapped inside the tube, thereby expanding the air. After slowly lifting the tube by small increments, you record the results in a table. Your theory is that the pressure on the gas is less than the pressure of the atmosphere by the amount H. That is, $P = A - H$, so your law states that $(A - H)V = $ constant. After comparing this equation with the new set of data, you arrive at the same value of A as before *as well* the same value of the constant. Thus, the single law, $PV = $ constant, explains both sets of data. Your critics are impressed, but they are not entirely satisfied. Nonetheless, you are pleased by the result, and you now view the independent confirmation of your law as being very important.

Yet the evidence you have now presented for your law is *not* as strong as the evidence that Boyle presented for his law. How is this possible? How can there be a greater variety of evidence for the law than you have already presented? What else can the law explain besides the compression and expansion of gases? The answer is 'nothing'! Your mistake is to focus exclusively on what the law can *explain*, for this leads you to overlook important *indirect* evidence for the law.
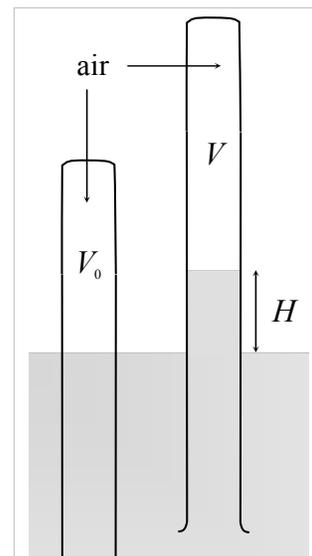
When Boyle introduced his law, he did not mention the expansion of gases until much later in the discussion. Instead, he alludes to the fact that atmospheric pressure was already known to be equivalent to approximately 29 inches of mercury. In Boyle's own words: "…we observed that the air contained in the shorter [tube], which was hermetically sealed at the top, was condensed by 29 inches of mercury into half the space it possessed before; from whence it appears, that if it were able in so compressed a state, by virtue of its spring, to a resist a cylinder of mercury of 29 inches, *besides the atmospheric cylinder incumbent upon that*, it follows that its compression in the open air, being but half as much, it must have but half that weight from the atmosphere that lies upon it in the compressed state."[14] That is, Boyle is pointing to the coincidence that the volume of the trapped air in his experiment is halved by 29 inches of mercury, and this happens to be equal to the pressure of the atmosphere. So, the volume is halved when the pressure is doubled. The interesting fact is that the experiment that measured atmospheric pressure had nothing to do with the compression or expansion of gases.

Figure 1.6: Torricelli's experiment. When the tube is raised less more 29 inches above the surface, then the mercury level inside the tube will remain at about 29 inches.

The experiment was the famous Torricellian experiment, which was already known to Boyle.   Torricelli immersed a long straight tube in a trough of mercury so that it contains no air. Then he raised the closed end slowly out of the mercury, being careful to keep the open end under the mercury so that no air enters the tube. When the tube is raised a little above the surface, the mercury is fully supported (Fig. 1.6, left). This is the same result expected by every child that has played with plastic containers in a bath tub. But when a tube of mercury is raised more than approximately 29 inches above the surface, an unexpected phenomenon takes place (Fig. 1.6, right). A "vacuum" form in the tube, preventing the mercury from rising any higher than 29 inches above the surface.

Naturally, in those times, the exact nature of this "vacuum" was hotly debated, but this does affect Boyle's argument. Boyle's sees Torricelli's experiment as showing that the weight of the column of air, perhaps over a mile high, is sufficient to support a column of 29 inches of mercury, but no more. Torricelli's apparatus measures atmospheric pressure directly, without the use of any theoretical assumptions about how two pressures add together, or about the relationship between the volume and pressure of a gas. Torricelli had invented the first barometer.

The opening argument that Boyle makes for his law, quoted above, mentions only one datum from the table of evidence that he has for his law—namely, that it takes approximately 29 inches of mercury to halve the volume of gas (see the shaded row in the Table above). Boyle is arguing that his law explains why the same number, 29 inches, that appears in Torricelli's experiment, also emerges from his results. However, Boyle's law does not "cover" Torricelli's experiment, so it's not true that Boyle's law explains Torricelli's result.

The fact that the confirmation is indirect does not mean that indirect evidence is weaker than direct evidence. Rather, it seems to show that the whole is greater than the
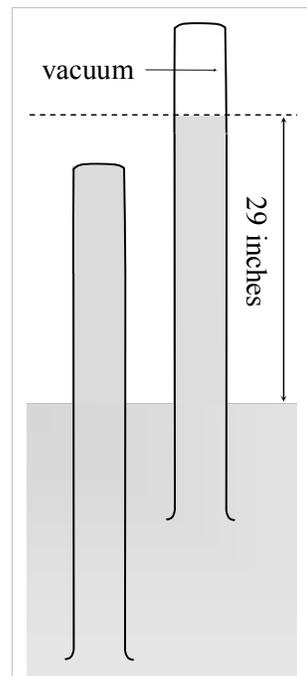
---

[14] Boyle, quoted from Shamos 1959, 39, my emphasis.

sum of the parts. While this may be the *intent* behind Hempel's covering law model of explanation, it violates the letter of the theory.

It seems to me that the point is better explained in terms of the distinction between *prediction and accommodation*. If Boyle is able to add the independently confirmed premise that $A = 29$ inches to his law, then he is able to *predict* the exact form of his law, which is then verified by the experimental results.   But you introduced A as an unknown adjustable parameter. While you are able to fit your law to the data, and infer that $A = 29$ inches, this is done post hoc. You merely accommodate the data, whereas Boyle predicts it.

Admittedly, Boyle used the general form of his law, $PV =$ constant, to make the prediction. But you need that even to accommodate the data. So there is still a clear difference in the arguments. It might be objected that Boyle would have checked his law by fitting it to his data before making the prediction. But this is a point about discovery, which is irrelevant to the logical difference between prediction and accommodation. There are many ways of describing this logical difference, but they all boil down (excuse the pun) to the same thing: The bridge between Boyle's theory and Torricelli's experiment is an important part of the empirical evidence for Boyle's law.

I have no doubt that there is some way of reformulating the point in terms of the notion of explanation. But I see no advantage in making the detour through a controversial explication of the notion of explanation. At its root, explanation is a psychological notion—it has to do with the nature of *understanding*, which is a psychological notion.   Intuitions about explanation differ from one person to the next, and the philosophy of science is not about the psychology of scientists, or philosophers. The dubious assumption that there is such a thing as *the* nature of scientific explanation is best avoided, if possible. And it is possible.

A good illustration of the problem with explanation is provided by Newton's theory of gravitation. Newton postulated a gravitational force that acts instantaneous at a distance across empty space. This notion is absurd to anyone who is convinced by Descartes' mediations about the essential nature of matter and forces as acting locally via the impact of contiguous pieces of matter. To anyone entrenched in the Cartesian dogma, Newton's theory does not explain anything. Newtonians, on the other hand, will claim that Descartes' postulation of an invisible fluid that moves the planets explains nothing because has no independent evidential support. It is the goal of philosophy of science to investigate how the such contradictory intuitions about explanation are *overcome*. For that, one needs an account of confirmation that does not mention explanation.

The account I have given of Boyle's law applies equally well to Newton's theory of gravitation. Part of the evidence for Newton's theory derives from the relational nature of the evidence, the agreement of independent measurements of earth's mass, which indirectly confirms Newton's theory of the moon's motion and his theory of terrestrial motion.

The Unification of Terrestrial and Celestial Phenomena

Ptolemy (100-170 A.D.) developed a surprisingly detailed theory of planetary motion almost 2,000 years ago. Ptolemy's theory assumes that the earth is stationary, and that the sun revolves around the earth. According to Ptolemy's theory, all celestial bodies, the moon, the sun, the planets, and even the stars, move around the earth according to a circle on circle construction, where the biggest circle is called the deferent circle, and the smaller circles are called epicycles. The celebrated Polish astronomer, Nicholas Copernicus (1473 - 1543), was the first to develop a *detailed* heliocentric astronomy, that

placed the sun at the center of the universe, even though the *idea* has been debated by the ancient Greeks. Copernicus did not abandon all the ideas of his predecessors. In fact, he held onto the Aristotelian idea that circular motion was the natural motion of celestial bodies more religiously than Ptolemy did, for he rejected Ptolemy's equant construction, which allows circular motions to deviate from uniform motion on a circle.

The essential difference in Copernicus's theory was that the deferent circles for each planet were centered close to the sun, rather than the earth. Of course, the center of the deferent circle for the earth's moon was still located near the center of the earth. However, the earth-moon system is now embedded on a very large orbit centered at the sun. The second, more radical, departure from Ptolemaic astronomy arose from Copernicus's insistence that the sun and the stars are motionless, and therefore the earth moves. Copernicus's crazy idea is that sun does not rise each morning, but rather, the earth sets. This part of Copernicus's theory appears to conflict with the obvious fact that cannonballs shot straight up in the air land near the cannon. How is that possible if the earth is moving at about 100,000 feet per second (Cohen 1985, 10)?

Tycho Brahe pointed out that all astronomical observations are of the positions of celestial bodies *relative* to the stars *as seen from the earth*. Therefore, the observable predictions of Copernican astronomy depend only the *relative* motions of celestial bodies, and these would be exactly the same if the hand of God were to reach in and hold the earth still while making the sun move around the earth. The planets would still revolve around the sun, except that the sun would be moving. All the predictions of Brahe's theory are the same as Copernicus's theory. Tycho Brahe argument proves that the Copernican thesis that the earth moves is an *inessential* part of Copernican astronomy.

It was Galileo who defended the crazy part of Copernicus's theory. The argument that follows is inspired by Galileo, although it is not intended to be an accurate reconstruction of it. My purpose is to ensure that the reader understands why Newton's theory contradicted the most enlightened and well argued Copernican arguments of the day.

Consider the following thought experiment. Hold a smooth metal ball on the lip of a smooth class bowl and let it go. It rolls down to the bottom of bowl, across the flattened bottom of the bowl, and up the other side; then it rolls back, to and fro, until it eventually lies motionless on the bottom of the bowl. The fact that it eventually stops is due to friction. So, imagine that there is no friction. Then the ball starts at the lip of the bowl, rolls up to the same height on the other side of the bowl, and back again, to and fro, ad infinitum. Now extend this thought experiment by supposing that the bottom of the bowl is much longer and follows the contour of the earth. While it is rolling along the bottom of the bowl it doesn't slow down, for if it did, it could not reach the same height on opposite lip of the bowl. Finally, extend the bottom of this frictionless bowl all the way around the earth, as shown in Fig. 1.7. Then the ball will circle the earth at a constant speed ad infinitum. There is no force required for this motion. This thought experiment is an argument for Galileo's principle of *circular* inertia, which says that any object that is moving in a circle around the earth will continue to move at a constant speed until acted on by an opposing force (such as friction).

The opposing force has to act in the direction of its motion. The force of gravity acting on the ball acts in a direction perpendicular (orthogonal) to its motion, and this force is canceled by the equal and opposite force of the earth's surface on the ball, which explains why the ball remains at the same distance from the earth's center.

Galileo's principle of circular inertia predicts that cannonballs shot straight up in the air will land near the cannon even if the earth is moving. For when it is shot up, it shares the motion of the earth, as does the air surrounding the earth. It does not experience any horizontal force, so it continues to move in circle around the earth until it hits the ground. It hits the ground near the cannon because both share approximately the same circular motion. The same principle predicts that a cannonball fired straight up on a steadily moving ship will land on the ship. Neither of Galileo's predictions depend on how fast the earth is moving, or whether the earth is moving at all.
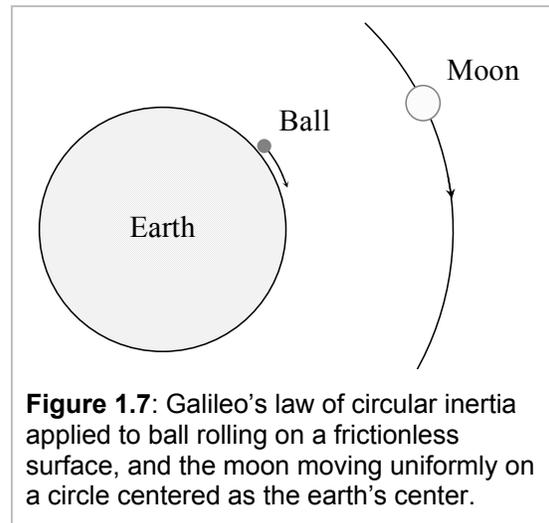

Figure 1.7: Galileo's law of circular inertia applied to ball rolling on a frictionless surface, and the moon moving uniformly on a circle centered as the earth's center.

Given the Aristotelian assumption that the earth is motionless, they also predict that the land-based cannonball will land near the cannon. But in the case of a cannon on a moving ship, they predict something different because the ship is moving relative to the earth. If the cannonball stays in the air long enough, and the ship is moving with sufficient speed, then the Aristotelians predict that the ball will land in the water to the rear of the ship. No Aristotelian was foolish enough to test this prediction because they also knew from everyday experience that if one drops a cannonball from a steadily moving carriage, then it lands near the foot of the carriage.

Newton's principle of inertia, his first law of motion, is a principle of *linear* inertia. It says that a body moving along a *straight line* will continue to move on that line with uniform speed until acted on by a force. Newton's principle contradicts Galileo's principle. Yet the two principles do not differ significantly in their predictions about terrestrial projectiles. Galileo's circles have a very large radius, so the arcs on which cannonballs move during any relatively short flight are approximately straight. So what empirical reason could Newton possibly have for denying Galileo's principle? Perhaps he had no empirical justification at all? Perhaps Newton's law of inertia is merely a definition, or a tautology, or maybe it is adopted on purely conventional grounds?

As a first step towards resolving the puzzle, consider what Galileo's principle implies about the motion of the moon. To keep the argument simple, suppose that the moon travels on a circle around the earth with uniform speed (Fig. 1.7). Galileo's principle of circular inertia predicts that the moon will continue travel with a uniform speed on the same circle unless acted on a by a force, either in the direction the motion, or in a perpendicular direction. If the moon was being pulled by gravity, with no opposing force, it would fall to the earth, just as the cannonball does. The moon is not falling to the earth. Therefore the moon is also outside the sphere of the earth's gravity.

In Galileo's view, terrestrial motion and celestial motions are similar to the extent that both conform to the principle of circular inertia, yet different because the moon is not subject to the earth's gravity. If Newton has any evidence *for* the theory that the earth's gravity affects the moon, then it is also
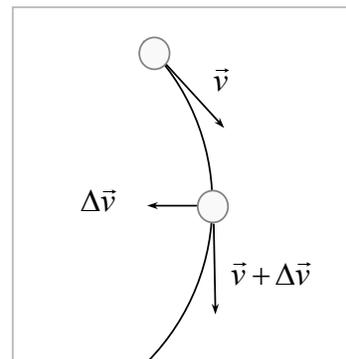

Figure 1.8: The Newtonian concept of acceleration as the change in velocity per unit time, where velocity is a vector quantity having a magnitude and direction.

evidence against Galileo's principle of circular inertia.

An even simpler way of reaching Galileo's conclusion is to argue the moon is not subjected to any force because it is not accelerating. How can Newton possibly deny the soundness of this argument?

The first step in Newton's analysis of the problem was to define the relevant variables. Just as Boyle's law rests on the proper understanding of pressure, Newton's theory begins with a more careful definition of acceleration. As previously noted, the moon is not accelerating if acceleration is defined as the change in speed per unit time. So, Newton's concept of acceleration has to be different. Indeed, for Newton, acceleration is the change in *velocity* per unit time, where velocity has a *direction* as well as a magnitude. In the terminology of contemporary mathematics, velocity is a *vector* quantity. Speed refers to the *magnitude* of the velocity. Since the apple does not change its direction of motion, it is still accelerating towards the center of the earth. But when uniform motion on a circle around the earth is analyzed according to the new definition, a striking fact emerges. The moon is actually accelerating towards the earth!

To see why this is so, look at Fig. 1.8. At two different times, the velocity of the moon is represented by a vector that has the same length, but different directions. The second velocity is the vector sum of the first velocity plus an incremental velocity, labeled $\Delta \vec{v}$. It is easy to see that $\Delta \vec{v}$ points approximately towards the center of the circle. In the limit, when the difference between the two lunar positions is infinitesimal, it is possible to prove that $\Delta \vec{v}$ points exactly towards the center of the circle. Since the acceleration vector is along the same line, the analysis proves that the acceleration is *centripetal*, which means 'center-seeking'.

Although we are not concerned with the psychology of discovery, it is easiest to get the gist of Newton's argument if it looks like an inductive argument.[15] Begin with Kepler's first two laws of planetary motion, and assume that they provide a good empirical description of the moon's motion around the earth.[16] Kepler's first law says that planets move on ellipses with the sun at one focus. The first law says nothing about how fast a planet moves along its orbit. Kepler's second law says that a line drawn from the sun to the planet sweeps out equal areas in equal times. Notice that in the special case in which the ellipse is a circle, the two foci of the ellipse coincide at the center of circle, so the sun is at the center. In that case, Kepler's second law implies that the planet is moving with uniform speed along the circumference of the circle. Kepler's laws say nothing about what causes these motions.[17]

Now consider the possibility that gravity causes the acceleration of the moon towards the earth. According to Newton's theory, any force that causes an acceleration must act in the direction of the acceleration with a magnitude equal to the mass of the body times the magnitude of its acceleration. Thus, Newton's second law, $F = ma$, is actually a vector equation. From the definition of acceleration, Newton proved the following general theorem: If the line from some point O to a body moving along any path sweeps

---

[15] In his major work, the *Principia Mathematica*, Newton presents his theory in a classic hypothetico-deductive format. He begins by stating his three laws of motion, without describing how he discovered them, and then proceeds to deduce theorems from the laws.

[16] The actual motion of the moon does not conform to Kepler's laws very closely, and even the most complex of Newton's models did not explain all of its known idiosyncrasies. In fact, Newton wrote a book on the lunar problem after the *Principia*, which still did not resolve all the anomalies. It was Clairaut who made significant progress on the problem much later. My purpose is provide a very simple account that at least explains why Newton's theory was better confirmed than Galileo's.

[17] Kepler has plenty to say about that, but Newton's argument depends only on the approximate empirical validity of Kepler's law.

out equal areas in equal time, then the body is accelerating towards O.[18] So, if the line from the center of the earth to the moon obeys the area law, then the moon is accelerating towards the center of the earth. Therefore the earth's gravity could well be the cause of the moon's acceleration.

A second relevant application of the same theorem is to uniform motion in a straight line. Consider a body moving in a straight line so that it travels equal distances from A to B, from B to C, from C to D and from D to E in equal times. Then the areas swept out by the line drawn from O will be equal because the areas of the triangles are all equal (They have the same vertical height, and the same base length). Therefore, the area law is satisfied by uniform linear motion, and the body's acceleration is directed towards the point O. But the point O is arbitrary, so the magnitude of all these accelerations must be zero. When combined with Newton's second law, the theorem implies that uniform linear motion is force-free.[19] We can see why Newton rejects Galileo's principle of circular inertia. This does not prove that he's right and Galileo's wrong. It merely shows that Newton's redefinition of acceleration, which is the new quantity to be explained, is directly connected to his *linear* principle of inertia. Moreover, the shift is motivated by Kepler's laws, which were well established as the best predictive descriptions of celestial motions known at the time.
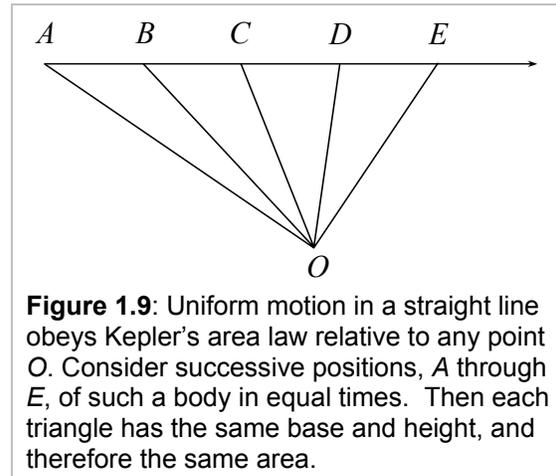

Figure 1.9: Uniform motion in a straight line obeys Kepler's area law relative to any point *O*. Consider successive positions, *A* through *E*, of such a body in equal times. Then each triangle has the same base and height, and therefore the same area.

Continuing in the same vein, Newton proves a theorem that implies that if the line from the moon to the center of the earth obeys the area and the moon is *moving on an ellipse*, then the moon's acceleration towards the earth is inversely proportional to the square of that distance.[20] Thus, if the moon's motion is approximately Keplerian, and the resulting acceleration is caused by the earth's gravity, then the force of gravity must be inversely proportional to the square of the distance between the moon and the earth. Let r denote that distance, and let a be the magnitude of the moon's acceleration towards the earth. Then $a = K/r^2$, where K is some constant.

The argument assumes that earth-moon system is revolving slowly enough around the sun on a sufficiently large orbit, so that at any time, the motion of the earth-moon system is approximately uniform. The fact that it's not exactly uniform is something that Newton takes into account later. My purpose is to show that the inverse square law is not a proverbial rabbit pulled out of a hat. This is not presented as a fact about the psychology of discovery, even if it is. It is presented as a fact about the logic of confirmation. For it ensures that Newton's theory about the earth's gravity, if it succeeds, may also cover a wealth of other empirical information about the motion of the planets around the sun. Just as in the case of Boyle's law, the discovery of Newton's law could have been based on a relatively spare amount of evidence. The confirmation of the theory depends on much more.

[18] Proposition I, Theorem II, Book I, of the *Principia* states that "Every body that moves in any curved line described in a plane, and by a radius drawn to a point either immovable, or moving forwards with an uniform rectilinear motion, describes about that point areas proportional to the times, is urged by a centripetal force to that point."
[19] Or more exactly, the resultant force must be zero.
[20] In Proposition XI, Problem VI, Book I.

Consider the formula $a = K/r^2$ more carefully. If the lunar acceleration is caused by the earth's gravity, then we expect the constant K to be proportional to the earth's mass. Label the earth's mass as M. The constant of proportionality is the universal constant of gravitation, G. To simplify the formulae, suppose that the units of mass are chosen so that $G = 1$. Then, we arrive at the final result, $a = M/r^2$. When we combine this with $F = ma$, then we get Newton's inverse square law of gravitation in its usual form:

$$F = mM/r^2 .$$

The derivation also assumes that the moon and the earth can be treated as point masses. This may seem plausible for the moon and the earth because of the reasonably large distance between them. But if we expect the same law to explain the motion of the apple, then there may be a glitch. For according to the new theory, the apple is subjected to a gravitational pull from every particle of matter in the earth, whether it is on the surface or near the center. Why should it be the case that the sum of all these forces is approximately equal to the force that the earth would exert if all its mass was concentrated at its center? Amazingly, Newton proved a theorem that says this is exactly true, provided that the earth's mass is distributed uniformly with a nested set of spheres. More precisely, the mass density of a point inside the earth must depend only on the distance of the point from the center.
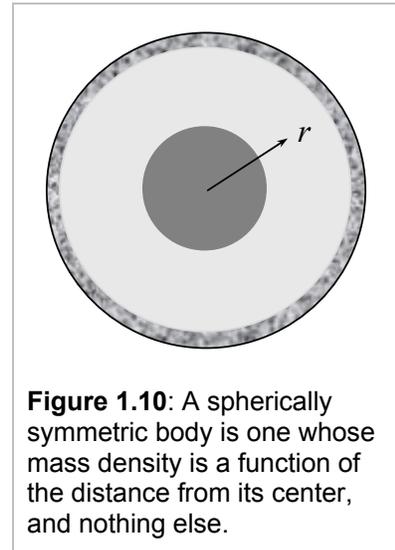

Figure 1.10: A spherically symmetric body is one whose mass density is a function of the distance from its center, and nothing else.

In such a case, we say that the earth's mass is *spherically symmetric*. According to Newton's theorem, it doesn't matter that the earth's core is more dense than its crust—the gravitational effect of the entire mass on the apple, or on the moon, is the same as if it were all its mass were concentrated at it center. So, the assumption that justifies the application of $F = mM/r^2$ to the apple is *not* the obviously false assumption that the earth is a point mass, but the far more plausible assumption that the earth's mass is spherically symmetrical.

Another assumption of the argument is that the mass of the moon is small enough relative to the mass of the earth and sufficiently distant from it so that the gravitational effect of the moon on the apple is small. This seems reasonable if we assume that the earth and the moon have approximately the same densities.

Finally, we are in a position to figure out what empirical evidence Newton can cite in support of his theory. From $F = ma$ and $F = mM/r^2$, it follows that $ma = mM/r^2$. The m's cancel, so we get $a = M/r^2$, as before. This implies that only the mass of the earth is needed to explain the acceleration of any body under the influence of the earth's gravity. Thus, the gravitational mass of the earth is independently measured by terrestrial and lunar motions. To make this point explicit, we can shift all observable quantities to the right hand side of the equation to obtain:

$$M = ar^2 .$$

The theory implies that all estimates of M obtained from this equation must agree, at least approximately. Note that such an agreement of measurements would provide the same kind of indirect evidence for Newton's model of lunar motion that Boyle obtained for his law from Torricelli's experiment.

When Newton first did the calculations, he found that the measurements did *not* agree. There is some speculation that this delayed the publication of the *Principia*, even though the evidence for his theory did not merely rest on this one test. For example, the formula $M = ar^2$ implies that the values ar^2 must *always* agree, even when they are calculated from the moon's motion at different times. The fact that these independent measurements do agree is already strong evidence in favor of Newton's account of the moon's motion. Nevertheless, the idea that the earth has two different gravitational masses, one explaining lunar accelerations, and one explaining terrestrial accelerations, would have been perceived as a peculiar fact, even though it is perfectly possible. And certainly, it would have meant that terrestrial and celestial phenomena would not have been unified, even if the two causes would have been similar in kind.

Eventually, the anomaly was resolved when it was discovered that current estimate of the earth's radius was wrong. Once Newton re-calculated the earth's mass using the corrected value, the measurements did agree. And so the published version of the *Principia* proclaimed that gravity acts universally between any two bodies according to the inverse square law.

Recall that the Newtonians were competing with the Cartesians in opposing the received Aristotelian physics. Descartes common sense intuitions about the origin of forces led to the Cartesian view that the planets are propelled by vortices in an invisible fluid that filled the void between the planets. To those who were impressed Descartes intuitions, Newton's theory was absurd. To them, it was not the best explanation. It was no explanation at all. So how did Newton's theory prevail? How did it manage to negate the force of these common sense intuitions acting against it? The answer I have given is that Newton's theory was better confirmed by the observational data.

If view applies to a wider range of examples, then it might explain how radically new ideas—such as Copernicus's crazy idea that the earth moves, or Newton's action-at-a-distance, or Einstein's weird notion that space-time is curved, or the quantum mechanical idea that particles don't have precise positions, or Darwin's dangerous idea that the biblical story of creation is false—can prevail in spite of the psychological entrenchment of contradicting views.[21]

As partial evidence for the wider application of the same ideas, notice how similar Newton's argument is to Boyle's argument. In both cases, the independent measurement of theoretical parameters can be viewed as playing an important role.

There are many issues that have not been discussed in this chapter. What does the confirmation of a theory entitle us to conclude? Does the empirical evidence show that the theory is true, or just that parts of the theory are true? Or does it show that the quantities it postulates exist, or merely that the theory's predictions are reliable to some degree? These questions will be addressed in later chapters.

The Objectivity of Subjective Judgments

The previous two sections have emphasized the important role that the *numerical* agreement of independent measurements can play in overcoming prior dogma. Does this mean that the quantitative sciences, which cannot appeal to such evidence, can never break the shackles of opinions born from a priori judgments and passed down by educational indoctrination? The purpose of this section is to clearly state that this is not my view.

---

[21] It may be that these theories are psychologically satisfying to scientists brought up with the theory from an early age, but this does not explain why the theories ought to be accepted in the first place.

To the contrary, the best quantitative sciences, such as physics, rely on the subjective judgments of experimenters. These judgment are not entirely dissimilar to the ability of biological taxonomists to recognize the salience of a taxonomic trait. It is plausible to me that such judgments are based on the role such traits have played in past predictions (roughly, if this new species has *these* traits, then I bet they have these other traits as well—let's look).[22] It's not plausible to me that expert human judgments can ever be eliminated from science, and I see no reason why they should be. Perhaps a really trivial example concerns the essential role of human perception, which empirically minded philosophers take to be the bedrock of any objective science. When observations of planet's motion are used to confirm Newton's theory, it is essential that the observations are of the same planet. Mars is recognized by its reddish color, and so forth. It is not something that is reduced to some kind of numerical calculation. It could be, I suppose, but it is not done unless the subjective judgments prove to be unreliable.

Expert judgments are expert because they have a proven track record in making accurate predictions. A simple example is chicken sexing, in which the sex of chickens is judged on the basis traits that are not well articulated by those trained in this task. Nevertheless, the reliability of those judgments are constantly tested by later examining the adult birds, or they could be verified by DNA analysis. So why use such an expensive method when there is no need. Another example is the examination of chest x-rays photographs by trained radiographers, whose judgments are verified by the surgeon, and/or by the general practitioner who cares for the patient. Subjective judgments are constantly evaluated for their predictive accuracy. And there is no reason to regard these tests as very different from the kinds of tests apply to quantitative predictions.

As an analogy, robots that roam the Martian landscape are designed to *extend* human capabilities, not to replace them. At every stage of development of robotic technology, robots that interact with humans in real time are preferred over those that don't, other things being equal. If the analogy is a good one, then the tools of quantitative science should always be seen as extending the reach of natural human intelligence. Just as engineers test the performance of robots, quantitative methods are evaluated in terms of their ability to *improve* the scope and accuracy of subjective judgments, without entirely replacing them.

On the same time, the quantitative theories of the so-called exact sciences often contain untested assumptions passed down by tradition. Such biases should be exposed and questioned as much as possible, even if they don't have any immediate effect on the empirical success of the science. This point was emphasized earlier by the problem of irrelevant conjunctions. There is nothing special about quantitative science that immunizes it from the foibles that we more often associate with non-quantitative science.

The proper scope of quantitative methods in sciences as biology, sociology, economics, phylogenetic inference, evolutionary theory, archaeology, and psychology, is hotly contested. Indeed, the use of quantitative modeling should be questioned in every instance, and each case should be judged as best as possible in terms of predictive tests. It seems to me that the same standard applies to all science across the board.

Popper (1959) introduced the problem of how to demarcate between science and pseudoscience; for example, between astronomy and astrology, and between evolutionary theory and the biblical story of creation (creationism). In each of these pairs of theories, the one that we judge to be genuine science is to some extent a quantitative science. Astronomy is a quantitative science, and evolutionary theory is making an ever increasing appeal to quantitative methods. A hasty way of resolving the demarcation

---

[22] See Forster (1986b) for a description of phylogenetic inference along these lines.

dispute would be to claim that astrology and creationism are unscientific because they are non-quantitative. In my view, the reason that astrology and creationism are pseudosciences has to do with their lack of predictive success, where prediction is understood in a logical, rather than historical, way. Either the theory fails to make testable predictions, or their predictions prove to be unreliable, or a bit of both.[23]

   While this book focuses on quantitative methods, it is important to understand that quantitative modeling has a wider scope than some might think. Consider a machine that makes a yes-no response determined by a yes-no input plus an internal state that has a finite number of values. It is a trivial exercise to model the machine mathematically by encoding the inputs and outputs by the numbers 0 and 1, for example, and assigning numbers to each possible internal state. It doesn't matter than the numerical assignments are arbitrary. The actual numbers used in any mathematical model are always arbitrary to some extent.

   It is also possible that the formation of qualitative judgments are based on criteria that are quantitative in form. For example, most, if not all, perceptual judgments are qualitative in form—the book in front of me is blue, the monitor is closer to me than the door, and so. Yet from what we know about the brain, these judgments are formed on the basis of neural processing, which is commonly believed to be describable in terms of numerical quantities, such as the frequency of neural spikes. So, it may be that all perceptual judgments are based on the agreement of *quantitative* "measurements". It would be a mistake to suppose that numerical evidence can only support numerical conclusions. In fact, the same issue arises within the Boyle and Newton examples. The claim that a theoretically postulated quantity actually *exists* is a qualitative claim. It seems clear to me that such claims are supported by numerical relations in the data.

---

[23] A favorite creationist argument is that if creationism is not a science, then neither is evolutionary theory. Evolutionary theory is a science. Therefore, creationism is a science. This argument appeals to the common view that evolutionary theory can only *explain* phenomena "after the fact". It cannot predict future evolution. True, it cannot predict the course of evolution very far into the future. First, evolutionary theory makes plenty of reliable predictions about evolution "in the small". Second, I do not limit the term 'prediction' to the prediction of future unseen facts. The prediction of past unseen facts also counts. Nor do that have to be seen by no one.   What counts is the theory's ability to predict some facts on the basis of other facts. See Kochanski (1973) for a more detailed discussion of prediction in evolutionary theory.