Three Levels of Scientific Hypotheses

It is sensible in the philosophy of the quantitative sciences to distinguish between three kinds of hypothesis. The main goal of this chapter is to explain why the distinction is philosophically useful.

The distinction itself is best explained as follows. At the empirical level (at the bottom), there are curves, or functions, or laws, such as $PV = \text{constant}$ the Boyle’s example, or $a = M/r^2$ in Newton’s example. The first point is that such formulae are actually ambiguous as to the hypotheses they represent. They can be understood in two ways. In order to make this point clear, let me first introduce a terminological distinction between variables and parameters. Acceleration and distance ($a$ and $r$) are variables in Newton’s formula because they represent quantities that are more or less directly measured. The distinction between what is directly measured and what it is not is to be understood relative the context. All I mean is that values of acceleration and distance are determined independently of the hypothesis, or theory, under consideration. I do not mean that their determination involves no kind of inference at all. For instance, acceleration is the instantaneous change in velocity per unit time, and this is not something that is directly determined from raw data that records the position of the moon at consecutive points in time. It is consistent with that raw data that the motion of the moon is actually discontinuous, so that the moon has no acceleration. So, there are definitely theoretical assumptions make about the moon’s motion that are used to estimate the moon’s acceleration at a particular time. But these assumptions are not unique to Newton’s theory. The same assumptions are also made by the rival hypotheses under consideration. In fact, the existence of quantities such as instantaneous acceleration is only called into question by the far more recent theory of quantum mechanics. Likewise, in the case of Boyle’s law, there is no controversy in viewing the volume of the trapped air as being determined in a way that does not make use of the theory that Boyle is introducing.

In contrast, the determination of the constant in Boyle’s law does make use of the Boyle’s formula. So, it is treated as a theoretical parameter, and the claim that it represents a real quantity in nature rests on the validity of Boyle’s theory. Similarly, the mass of the earth, $M$, is a theoretical parameter in Newton’s equation. So, the distinction between variables and parameters does not rest on any metaphysical distinction—if the theories are true, then variables and parameters represent quantities that exist in nature. Moreover, parameters represent quantities that vary may vary from one situation to another. For example, the ideal gas law tells us that Boyle’s constant depends on the number of molecules in the gas, and the temperature of the gas.

Pressure is a mixed case because it is the sum of two pressures—the first is that atmospheric pressure, $A$, which is theoretically introduced in Boyle’s theory, and is measured by fitting Boyle’s law to the total evidence (including the results of Torricelli’s experiment). It is therefore a parameter. On the other hand, the weight of the mercury in the $J$-shaped tube, is proportional to the added height of the mercury, $H$, which is a variable. The value of the pressure, $P$, is only determined by using the theoretical assumption that the total pressure is the sum of these component pressures, and this
assumption is introduced by Boyle. In later studies of gases, pressure is treated as a variable because the additivity principle is one the assumptions shared by competing theories. The distinction is relative to the particular epistemological context under discussion.

Now return to the formulae under consideration. If the parameters are assigned particular numbers, determined by applying the formula to a particular empirical application—to the motion of bodies in the earth’s vicinity, or to a particular pocket of gas—then the hypothesis is specific enough to make predictions in new cases. The hypothesis is then represented by a specific curve. If the parameters are viewed as adjustable parameters, then the hypothesis is represented by a family of curves. Any hypothesis that has at least one adjustable parameters is called a model. The difference is that parameters that appear in the formula for a curve are adjusted, while the parameters that appear in the formula, or equation, that represent a model are adjustable.

Again, there can be mixed cases in which some, but not all, of the parameters are adjusted. For example, when Boyle uses Torricelli’s results to assign a particular numerical value to the atmospheric pressure, but treats the constant as adjustable, then the resulting formula is still defines a model.

At the highest level, there are theories. Boyle’s assumption that pressures are additive is a part of his theory. Newton’s theory includes his three laws of motion. It is less clear whether Newton’s inverse square law is a part of his theory. After all, Newton’s strategy is to allow special force laws, such as his inverse square law, or Hooke’s law of springs, \( F = -kx \), where \( k \) is the stiffness of the spring, and \( x \) is the distance that the spring is extended, to be invented for the purpose at hand. So, for me they are a feature of models. Also note that in the very simple model of earth-moon motion we considered, Newton’s inverse square law is not applied in full, because it ignores the gravitational influence of the sun, as well the fact that the center of gravity of the system is not located exactly at the earth’s center because the moon’s mass is not negligible. So if his universal law may well be regarded a part of his theory, we still important to make a distinction between the model and the theory. Also, are we to treat the assumption that atmospheric pressure depends on altitude as a part of Boyle’s theory, or that his constant depends on temperature as a part of his theory? Again, I am happy to include these things as a part of the background theory, so long as it is recognized that particular models may be oversimplified in a way that incorporates these assumptions in as merely approximately true. The definition of what counts as an idealized model depends on the fact that models, if literally construed, are incompatible with the theory. For me this is firstly a reason to make a clear distinction between a theory and its models, and secondly a reason not to construe models literally. I will have more to say on that topic at the end of this chapter.

What is the status of Newton’s assumption that the mass of celestial bodies do not change over time, or his assumption that objects have a single mass value that applies universally to all phenomena. Again, I see nothing wrong with viewing these assumptions as simultaneously a part of the theory and a feature of the models.

The reason that this three-way distinction between curves, models, and theories is valuable relates to a point made in the first chapter. There I pointed out that traditional theories of confirmation in philosophy of science make no essential distinction between curves, models and theories. This is especially true of the Bayesian theory of confirmation, which is most simply formulated as follows: Observed evidence \( E \) confirms a hypothesis \( H \) if and only if \( E \) increases the probability of \( H \); in symbols, \( P(H | E) > P(H) \). Bayesians assume that such probabilities are well defined (at least as
measures of subjective degrees of belief), whether $H$ is the hypothesis represented by a curve, model or theory. It is my intention to develop a concept of confirmation that does not adopt this assumption as a first principle.

Curve Fitting in Three Steps

These three steps in the “colligation of facts” in the quantitative sciences were described by William Whewell 150 years ago.\textsuperscript{24}

**Step 1**: The determination of the dependent and independent variable(s).

**Step 2**: Model selection: The selection of a family of curves. Each curve specifies dependent variable as a function of the independent variable(s).

**Step 3**: The estimate of parameter values: The determination of the curve in the selected family that best fits the data.

It should be clear from this brief description that each step presupposes the completion of the previous steps. The steps form a logical sequence. A minor complication arises from the fact that the selection of a family of curves can be understood in two different ways. Selection can selection for the purpose of carrying out Step 3, or selection could be selection as the best model. Selection as the best model is really Step 4 in the process, any evaluation of rival models is likely to consider their degree of fit as a factor.

The data is only mentioned in Step 3. What counts as relevant data cannot be specified before the variables are determined in Step 1, although the raw data may be available before then.

The simplest form of data consists of a set of $(x, y)$ points, where $y$ is the dependent variable (e.g., gas volume) and $x$ is the independent variable (e.g., pressure). It may be that rival models (families of curves) disagree on the relevant independent variables. But competing models should have the same dependent variable, but need not have the same independent variables.

Given that curve-fitting is the fitting a curve to a set of data points, it is natural to think that one begins with the data. But the raw data is not always in the form required to discover the appropriate regularity. Boyle sought a law determining the volume of a gas, but first he had to understand how the weight of the air combined with the weight of the added mercury combined to produce the total pressure on the gas. A major step in Boyle’s discovery was the determination of the independent variable (the pressure). Similarly in Newton’s argument for his theory of gravitation, the first step was to select the dependent variable to be the change of velocity per unit time rather than the change in speed per unit time.

To overlook the first step is to overlook how much prior knowledge and experience goes into preparing the ground for a successful curve-fitting. Famously, Kepler fitted an ellipse to a set of known positions of Mars in its orbit. But these positions were three-dimensional positions of Mars relative to the sun, which are not observed directly. What’s directly observed are the angular positions of the planets relative to the fixed stars. The determination of the 3-dimensional positions of Mars relative to the sun relied on centuries of intellectual labor. By overlooking this first step, the automated discovery of Kepler’s laws can be made to look easy enough to be performed by a computer (Langley \textit{et al.} 1987).

The second step is the selection the model, or family of curves. This step is known as model selection in the statistics literature. Given that the scientists’ concept of model

\textsuperscript{24} See Chapter 7 of Whewell (1858). The same chapter is reprinted in Butts 1989, 223-237.
differs from the sense of the term commonly used by philosophers of science, it is appropriate to define the term. In the quantitative sciences, a model is an equation, or a set of equations, with at least one adjustable parameter. A model usually applies to a specific kind of situation. In contrast, a theory, such as Newton’s theory of motion, is a general set of laws and principles of broad scope, from which different models are derived using different sets of auxiliary assumptions.

It is worth noting that the quick characterization of a model as a family of curves is not quite right. Curves depend on the choice of a coordinate system, while the regularities or laws that they represent do not. Models are really families of regularities, or equivalently, families of mathematical functions, but the looser way of speaking is harmless so long as this is kept in mind.

Suppose we have a set of data consisting of pairs of \((x, y)\) values. One of the principle goals of curve-fitting is to make predictions of \(y\) for new values of \(x\). It seems deceptively easy—determine a single curve that fits the data and make predictions from that curve. The glitch is that each curve in a family of curves makes a different prediction. For that reason alone, one might not guess that models play any essential role in curve fitting.

Philosophers are like scientists. They try out the simplest ideas first. So perhaps it is unsurprising that two very influential philosophers of science, Reichenbach (1938) and Hempel (1966), did not think of curve fitting as mediated by models. For them, a curve is chosen from the family of all possible curves. Within this grand family, there are many curves that fit the data perfectly, which tacitly assumes that the data are error-free. There are no models derived from theoretical considerations in this barren curve fitting landscape. There is no indirect confirmation of hypotheses by the unification of disparate applications of the theory. There are no independent measurements of theoretical parameters, because parameters are introduced by models. Idealizations are good if they work, but this one is not even close. It has led to important misconceptions about the role of simplicity in curve fitting, as I shall argue soon.

For now, consider a generic example of the family of all straight lines in the \(x\)-\(y\) plane, which is represented by the formula \(y = a + b x\), where \(a\) and \(b\) are adjustable parameters that can take on any value between \(-\infty\) and \(+\infty\). Clearly, any particular set of numerical values assigned to the adjustable parameters will pick out a particular curve in the family. And conversely, any particular curve will determine a set of numerical values for all the adjustable parameters. This simple one-to-one correspondence between curves and parameter values explains how statistical estimation works. If one determines the curve that best fits the data, then one has found estimates of all the adjustable parameters. For that reason, step 3 is referred to as the estimation of the parameters.

There is a sense in which process of fitting a model to data belongs to the context of discovery because a prediction-making curve and estimated values of theoretical quantities are discovered through this procedure. Yet it also belongs to the context of justification because the relationship between the fitted model and the evidence is relevant to its confirmation. The context of discovery and the context of confirmation may be the same, but we should not overstate the conclusion—it does not imply that discovery and confirmation are the same thing.

The third step in curve-fitting uses the data to pick out a particular curve in the family that best fits the data. The definition of ‘fit’ can differ. The best known method of

---

25 Scientists also use the term in two different senses. One is the sense defined here, while the other is in the everyday sense of a model airplane used in wind tunnel experiments, or Watson and Crick’s DNA model, or rats as a model for human beings in testing drug treatments.
determining the curve that best fits the data is called the \textit{method of least squares}. Interestingly, the method of least squares was introduced by Gauss (1777 - 1855) and Legendre (1752 - 1833) in the early 1800s as a way of inferring planetary trajectories from noisy data. If we associate with each curve an error distribution for the $y$ values, we may also define the best curve as the one that makes the data most probable. This alternative definition of fit leads to a statistical method of parameter estimation called the \textit{method of maximum likelihood}. Gauss proved that these two methods are equivalent when the error distribution follows the typical bell-shaped pattern, which now bears his name.

We note in passing that the third step in curve fitting provides us with a non-trivial account of how theoretical quantities are \textit{measured}. If one ignores models, then the measurement of theoretically postulated quantities is shrouded in mystery. There is no need to understand the detailed definition of fit at this stage. There are only two points that should be understood.

1. There is, generally speaking, a unique curve in a family of curves that best fits the data. Exceptions can arise. An example would be the model $y = (a/b)x$, where a set of $(x, y)$ points provides only a unique estimate of the ratio $a/b$ without providing estimates of $a$ and $b$ separately. The important point is that such exceptions do not impede the model’s ability to make predictions because a “reduced” model, in which the ratio is treated as a single parameter, can solve the prediction problem. Moreover, the underdetermination problem may be solved in a positive way if there are indirect ways of estimating the parameters $a$ and $b$. The beam balance is an example of this kind, as will be explained in the next chapter. Other cases may be more problematic, at least at first sight. As an extreme example, it is impossible to estimate all the positions and momenta of molecules in a gas by measuring its thermodynamic properties such as volume, pressure, and temperature. Or a model of learning may postulate the existence of brain states that cannot be determined by behavioral data alone. In each case, there may be a “reduced” model that succeeds in making many predictions, and the search for clever ways of indirectly estimating parameter values in a less “reduced” model is always and important positive development if it succeeds.

2. In general, there will be no curve in the family that fits the data perfectly. This means that the observed value of the dependent variable, $y$, may differ from the value of the $y$ predicted by the best fitting curve. The difference between these two values is often referred to as an error, or a residue. Errors may be due to errors of measurement. Measurements may be rounded off because it’s impossible to tell exactly how high the mercury is in the thermometer, or the barometer—in fact, it may not be well defined because the surface is curved. But errors may also be due to the action of other laws. For example, if the dependent variable is the daytime high temperature in Beijing and the independent variable is the day of the year, then the best fitting curve will predict the average temperature for that time of year. The actual temperature will deviate from the average temperature on any given day. The discrepancy is not due to the inaccuracy of any measurement device. Better predictions may be obtained by defining the residues to be a new dependent variable and selecting other independent variables, such as the presence of cold fronts. This is called the \textit{method of residues}. If the method succeeds, then it provides further confirmation of the original law. It may sound odd to say that a law is confirmed when its errors are confirmed to be \textit{genuine} errors. As always, the right question to ask is: \textit{What} is being confirmed?
In the meantime, we shall return to the idealized situation in which the data is error-free. I want to argue that mediation of models is philosophically necessary even in this situation. The argument shows that that epistemological importance of models is not exhausted by their role in the identification and management of observational errors.

Recall that our main criticism of simple enumerative induction in chapter 1 was that it did not allow for the introduction of new theoretical concepts. On the other hand, inference to the best explanation did provide for this, but did so in a vague way. We now see how it is that curve fitting, properly understood, does allow for the introduction of theoretical quantities, and the logic of the process is well defined.

**Beam Balance Predictions**

The beam balance example is easy to understand, but also rich enough to illustrate many philosophically interesting features of curve fitting in science.

Suppose we hang an object, labeled $a$, on one side of a beam balance, and find how far a second object $b$ has to hang on the other side to balance $a$. The distance that $a$ is hung from the fulcrum is labeled $x$ while the distance to the right of the fulcrum at which $b$ balances $a$ is labeled $y$. See Fig. 2.1. If $b$ is moved to the left of this point then the beam tips until the object $a$ rests on the ground and if $b$ is moved to the right the beam tips the other way. In the first instance, $x$ is 1 centimeter (cm) and the beam balances when $y$ is 1 cm. This pair of numbers is a *datum*, or a *data point*. We repeat this procedure 2 more times with different values of $x$, and tabulate the resulting data in Table 1. We are now asked to make the following prediction: Given that $a$ is hung at a distance of 4 cm to the left of the fulcrum, predict the distance at which $b$ will balance; *viz.* predict $y$. Once we notice that for all 3 data, $y$ is equal to $x$, it appears obvious that we should predict that $b$ will be 4 cm from the fulcrum when $a$ is hung at 4 cm. But what *general* method of prediction applies to this and other examples?

**The Problem of Many Curves**

The naive answer to the question about *how* to make predictions is roughly this: A regularity is observed in past data, and we should infer that this regularity will apply to future instances. In the beam-balance example, this idea may be spelt out in terms of the following inference: In all the observed instances, object $b$ balanced on the other side of the beam at an equal distance as object $a$. *Therefore*, $b$ will balance at the same distance as $a$ in every instance. We recognize the general pattern, or *form*, of the argument as that of *simple enumerative induction*:

In all the observed instances, system $s$ conforms to regularity $R$

System $s$ will always conform to regularity $R$.

In this example, the system $s$ is the beam balance with objects $a$ and $b$ hung at opposite sides of the fulcrum, and the regularity $R$ is expressed by the mathematical equation:

![Figure 2.1: A beam balance](image)
Philosophers of science, from Reichenbach (1938) to Hempel (1966) have been quick to recognize what’s wrong with this form of ampliative inference: It wrongly presupposes that regularity \( R \) is unique. A quick glance at Fig. 2.2 shows that there are infinitely many ‘regularities’ that fit the same set of data. There is no such thing at the observed regularity. So, any pattern of inference that tells us to extend the observed regularity from observed instances to unobserved instances makes a false presupposition. As previously mentioned, it’s not the fallibility of the inference that is in question. It’s the cogency of the method that has been challenged.

Does Simplicity Save the Day?

Reichenbach (1938) and Hempel (1966) assume that this problem is solved by bringing in simplicity. Their idea is to modify enumerative induction in the following way:

\[
R \text{ is the simplest regularity to which system } s \text{ is observed to conform.}
\]

System \( s \) always conforms to regularity \( R \).

Note that this form of ampliative inference does not conclude that the system conforms to the simplest possible regularity. It only says that at each stage, \( R \) is determined by the simplest curve that fits the data. Thus, the inferred regularity \( R \) will generally become more complex as more data are collected.

However, this solution will only work if the requisite notion of simplicity is coherent. Philosophers have tried to define such a notion, but none of these definitions succeed. For example, one might think of defining the simplicity of a curve by its number of adjustable parameters. But a single curve has zero adjustable parameters—all its parameters are adjusted. The number of adjustable parameters is a property of a family of curves, and any single curve is a member of infinitely many families of curves, each having different numbers of adjustable parameters. Popper (1959) tried to define simplicity in terms of falsifiability. But all curves are equally falsifiable—they are all falsifiable by a single data point.

There is a brilliant argument by Priest (1976) that shows, in my opinion, that no definition will work. Note that the requisite notion of simplicity is one that applies to single curves. It is not argument against the possibility of defining the simplicity of a family of curves.

Priest’s idea applies to the beam balance example in the following way. Instead of using \( y \) and \( x \), let’s label the variables \( y \) and \( x \), respectively. This makes it easier to write down the equation for the curve \( R_1 \) in Fig. 2, which is:

\[
y = f_1(x) = -0.5x^3 + 3x^2 - 4.5x + 3.
\]

In the \( x-y \) plane, this curve looks more complex than the curve \( y = x \). Now suppose we define a new variable \( y’ \) as:
\[ y' \equiv y - f_i(x). \]

The original data set in \( x-y \) coordinates is \{\((1,1),(2,2),(3,3)\)\}. Now the same data set in \( x-y' \) coordinates is \{\((1,0),(2,0),(3,0)\)\} because \( y' \equiv y - f_i(x) = f_i(x) - f_i(x) = 0 \). The simplest curve through the three points is a horizontal line with the equation \( y' = 0 \). We can now solve the prediction problem: When \( x = 4 \), \( y' = 0 \). In fact, the prediction is that \( y' = 0 \) for any value of \( x \). In other words, given any value of \( x \), we predict that \( y = y' + f_i(x) = f_i(x) \). Therefore, the new prediction is given by the curve \( R_1 \) because \( R_1 \) looks simpler than \( R \) or \( R_2 \) in the new coordinate system. The new prediction disagree, so there is a big problem. Simplicity does not save the day!

In the original coordinate system, the prediction based on \( R \) was that \( y = 4 \) cm when \( x = 4 \) cm. But in the new coordinate system, the intuitively simplest curve predicts that \( y = 1 \) cm when \( x = 4 \) cm. If our judgment of simplicity is based on what \textit{looks} simplest, then the judgment depend on the way we \textit{represent} the curve. But a method prediction should not depend on how we describe the problem anymore than what’s true should depend on whether we speak French or English.

The simplest way of expressing the conclusion of Priest’s argument is that all curves are equally simple. Or more exactly, the aim is to judge that simplicity of \textit{functions}, which are defined in mathematics as a set of ordered pairs \((x, y)\) such that there is a unique \( y \) value for any given \( x \) value. Curves are just pictorial representations of functions, and different curves represent the \textit{same} function in different coordinate systems. In fact, it seems that any curve can represent any function if the coordinates are chosen the right way. Paradoxically, it is exactly because our intuitive judgments are of the simplicity of \textit{curves} that these same judgments cannot define the simplicity of functions.

Priest’s point has greater relevance once we realize that scientists commonly seek new representations that make laws \textit{look} simpler. The motion of the planets relative to the fixed stars is complicated because the perceived motion is actual the combination of two simpler motions—the sun’s motion relative to us and the planet’s motion around the sun. Kepler’s discovery that Mars moves on a simple elliptical orbits depended on 1,500 years of intellectual labor, which led to the Copernican idea of looking at the data in a three-dimensional sun-centered coordinate system. Or take the puzzling phenomenon of interference of the “self-interference” of an electron passing through a double slit apparatus. The probabilities for the two single slit experiments do not simply add together. So, physicists invented a complicated representation of the probabilities in terms of the amplitude of wave functions and in this representation the linearity of addition is restored. Our intuitive judgment of the simplicity of curves has all the hallmarks of another kind of simplicity, which has to do with economy of use, ease of calculation, and so on. This kind of simplicity is called \textit{pragmatic virtue} because it is connected with practical issues. This is contrasted with \textit{epistemic} virtues, such as truth, approximate truth, or predictive accuracy. There is no denying that pragmatic simplicity is important to scientists. Unfortunately, this kind of simplicity depends on the mode of representation—it is language dependent, whereas truth and other epistemic virtues such as predictive accuracy are language independent. That is why pragmatic simplicity is no good as an indicator of truth.

Priest’s argument does not show that there is no kind of simplicity that is relevant to the epistemology of science. It is a problem for any \textit{philosophical} account of the relevance that is based on the simplicity of curves. What it shows is that any definition of
the simplicity of a single curve must grapple with the fact that our intuitive judgment of simplicity changes in different coordinate representations.

The good news is that the simplicity of a *family* of curves, defined as the fewness or paucity of its adjustable parameters, is invariant under any transformation of the $x$ and $y$ coordinates. For parameters merely index the functions in a family of functions, and the same parameterization works the same way in any representation. This allows for a positive role for simplicity in *model* selection. Whether the idea pans out has yet to be determined.

**The Mediation of Models**

The advantage of the more complicated description of curve fitting presented at the beginning of this chapter is that removes what appears to be an insurmountable problem—the problem of many curves. The problem disappears as soon as curve fitting is restricted to families of curves because there is generally sufficient random variation in the data to ensure that no curve fits the data perfectly, and only one curve fits it best.

Models are found everywhere in science—they arise in a variety of different ways. In the less mature sciences or in those sciences that study less regular phenomena (such as meteorology), there may be very little by way of background theory. The postulation of models is more a matter of trial and error. Where there are well developed background theories, such as in most, though not all, of physics, an important source of models is by derivations from the theory. As a first approximation, these may be thought of as deductions from the theory with the aid of auxiliary assumptions:

<table>
<thead>
<tr>
<th>Theory</th>
<th>Auxiliary assumptions</th>
<th>Model</th>
</tr>
</thead>
</table>

The illustrations I provide is the derivation of the beam balance model from Newton’s theory of motion. However, this process does not produce a unique model from the theory because there are different choices of auxiliary assumptions. Many models are compatible with a single theory. Hence, the question—How do we choose from among competing models?—arises in either case.

**The Beam Balance Model Deduced**

The derivation of the beam balance model is important for two reasons. First, it deepens the reader’s understanding of the example. But more importantly, it establishes the general point the deduction of a model from a theory does not imply that the theory determines a unique model. It does not obviate the need for model selection criteria.

---

26 Numerous philosophers of science have argued that philosophy of science can ignore the role of models. Regrettably, the sense of ‘model’ assumed by many of these authors is akin to the notion of ‘interpretations’ in predicate logic. For this sense of ‘model’ does not fit the scientific use of the term. Other philosophers, most notably Sober (e.g., 1988), have emphasized the importance of models in science in the sense described here. The deeper examination of the logic of curve fitting is meant to provide a deeper explanation for why models play an indispensable role in science.
Newton’s theory of motion tells us about forces and how they produce motion. In the case of a beam, Newton’s theory tells us that the beam will remain motionless (that is, it will balance) when the forces at every point balance. What are these forces? The idea of leverage is familiar to everyone. For example, everyone knows that it is possible to “magnify” a force using a rigid body such as a crowbar. A long rigid beam may lever a large boulder if the point at which we push down is a long distance from the fulcrum compared with the distance from the point of support (the fulcrum) to the end of the beam applying a force to the boulder (Fig. 2.3). Of course, you have to apply the downward force through a longer distance than the distance that the boulder moves upwards, so the work you do is equal to the work done on the boulder. This is required by the conservation of energy.

The same principle applies to beam balances. The forces applied to the beam arise from the gravitational forces of the two objects. If \( m(a) \) is the mass of \( a \), \( m(b) \) is the mass of \( b \), and \( g \) is the gravitational field strength, then \( a \) exerts a gravitational force of \( m(a)g \) on the beam at the point at which it is hung, and \( b \) exerts a force of \( m(b)g \) at the point at which it is hung. Now focus on the object \( a \). If the beam is to balance then the forces acting on \( a \) must balance. That is, the upward leverage of \( b \) on \( a \) must balance the downward gravitational force \( m(a)g \). By the principle of leverage, \( b \) is exerting an upward force on \( a \) equal its downward force magnified by the ratio of the distance \( y \) to \( x \).

The background theory, viz. Newton’s theory of motion, tells us that these two forces must be equal:

\[
\frac{y}{x} m(b)g = m(a)g .
\]

If we multiply both sides by \( x \) and divide both sides by \( m(b)g \), we derive the equation:

**Beam Balance Model:**

\[
y = \frac{m(a)}{m(b)} x .
\]

This completes the first two steps of curve fitting, for not only have we selected a model (second step) but we have also determined the dependent and independent variables (first step). Note that the mass ratio is playing the role of an adjustable parameter.

The several sections to follow are devoted to the many consequences that the mediation of models has on the normal operation of science. Some are familiar, such as the role of auxiliary assumptions, and some are not-so-familiar, such as the role of simplicity and unification.

**The Problem of Many Models**

In the beam balance example, there were many auxiliary assumptions made in deriving the model, at least implicitly. For instance, we ignored the leverage applied by the mass of the beam itself. This assumption would be true if the beam were massless, or if it were of uniform mass and supported by the fulcrum exactly at the center of the beam. As a second example, we ignored the presence of other forces. We tacitly assumed that there were no electrostatic forces, no puffs of wind, no magnetic forces, and so on. Another tacit assumption is that the gravitational field strength, \( g \), is the same on both sides of the
beam. We know that \( g \) is different at different places on the surface of the earth (e.g., near a large mountain). For a small beam balance the two masses will be at approximately the same place, so \( g \) will be approximately the same, but not exactly the same.

All such simplifying assumptions are called *auxiliary assumptions*. In practice, it is impossible to list all auxiliary assumptions, and often scientists do not make them explicit. Auxiliary assumptions are commonly formulated in the vocabulary of the theory itself, such as the assumption that there were no other forces acting. Yet auxiliary assumptions are not *deduced* from the theory. Newton’s theory of motion does not tell us whether a beam has a uniform density, or whether it has been properly centered. Nor are auxiliary assumptions determined from the data. For example, if \( a \) is subjected to a gravitational field strength of \( g_1 \) while \( b \) is subjected to \( g_2 \), then the model derived from the theory is:

\[
y = \frac{m(a)}{m(b)} g_1 \cdot x.
\]

This model defines exactly the same family of curves (all straight lines passing through the origin, the point at which \( x = 0 \) and \( y = 0 \)). The only difference will be a difference in interpretation: In this model, the slope of the straight line passing through all the data points is interpreted as the value of the ratio of the weights rather than the ratio of the masses.\(^{27}\) Other data *may* help to decide between auxiliary assumptions. For example, we could test the assumption that the gravitational field strength is uniform by seeing whether a single mass stretches a spring different amounts when it is moved from place to place. But note that this test will introduce its own auxiliary assumptions. Eventually, we are going to find auxiliary assumptions that are not derivable from theory even with the aid of the total data. If this is correct, if there are auxiliary assumptions that are not wholly derivable from theory plus data, then there will be many models that are underdetermined by the theory plus data.

If this is true, then at any point of time, a theory has many viable models. Each model will, in general, make incompatible predictions. When we observe whether some of these predictions are true, we narrow down the set of unfalsified models. But the problem remains. There are still many rival models that make incompatible predictions.

Normal and Revolutionary Science

The picture of science assumed in this chapter is represented in Fig. 2.4. At the top, a theory, such as Newton’s theory of motion, entails many models using different auxiliary assumptions. A model is not protected from refutation by any auxiliary assumptions. Models are vulnerably exposed to the data, and as more data are collected, discrepancies, or anomalies, may become apparent. More complicated models may remove these anomalies. New discrepancies may appear when now or more precise

\(^{27}\) Weight is the *force of gravity*, equal to the mass times the gravitational field strength - the weight of an object is different on earth than it would be on the moon because the gravitational field strength is different, but its mass would not change.
data is gathered, and the process is repeated. This is the process of science that Kuhn (1970) refers to as normal science. Its defining feature is that models change, but the background theory does not. But theories do sometimes change, and Kuhn refers to such events as revolutions in science.

In the first half of the twentieth century, philosophers tended to focus on the revolutionary science because they had just witnessed the overthrow of the greatest theory of all time—Newtonian mechanics. And the theories that produced this revolution—Einstein’s special and general theories of relativity and quantum mechanics, introduced startling new ideas about the nature of the world. It is perhaps unsurprising that normal science was seen as philosophically less interesting.

Popper’s methodology of falsificationism provided a deceptively simple picture of theory change in terms of the Darwinian idea of “survival of the fittest”. Science consists of a repeating cycles of bold conjectures and refutations. Theories that are falsified by the evidence are abandoned, and of those theories that survive scientists favor the boldest, or most falsifiable.

By 1960, two things about this new revolution in physics had become apparent. First, the Newtonian empire had not been vanquished from science in the way that Aristotelian physics had been superceded by Newtonian mechanics. Second, the new theories had not provided a clear, unified, understanding of fundamental nature of reality, even though they surpassed all expectations in terms of the scope and accuracy of their predictions.

The time was ripe for Kuhn’s influential book, Structure of Scientific Revolutions, which argued against theories as the bearers of truth. Kuhn’s negative view of theories is devastating to any philosophy of science built on the premise that theories are the primary bearer of objective scientific truth. This included Popper’s falsificationism. Philosophers noted that the fact that false predictions make from theories only with the aid of auxiliary assumptions could be blamed on the auxiliary assumptions, rather than the theory. Thus, Kuhn’s observation that in normal science the theory is never (successfully) questioned is consistent with the logical analysis of prediction. Yet none of this speaks against falsificationism as a solution of the problem of many models. Whether it plays this diminished role successfully is the main topic of the next chapter. The purpose of this chapter is to further highlight the philosophical importance of distinguishing between theory and models.

Lakatos (1970) responded on behalf of the Popperian school by attempting to defend the objective rationality of theory choice against Kuhn’s attack, while others such as Sneed (1971) bowed to the Kuhnian view by arguing that theories are noting more than definitions, and therefore devoid of any empirical content. Instead, the empirical content of science is achieved at the level “models”, except that for Sneed, a ‘model’ referred to a very formal logical notion of ‘model’, which is a generalization of the logician’s notion of an interpretation of predicate logic. The sense in which I am using the term ‘model’ is closer to the common usage of the term among scientists.

I am seeking middle ground between Sneed and Lakatos. Contrary to Sneed, I want to allow that theories do have empirical content. In addition, I reject Sneed’s notion that the reconstruction of scientific theories and models in artificial logical languages is insightful. In my view, the formal language of science itself—namely mathematics—provides the best formalism for describing the formalism of the quantitative sciences. It’s perhaps unfair to Sneed to make this sound truistic. It could have been that Sneed’s formalism was important to the ‘logic’ of science. In the final analysis, the insights gained never made the complexity of his formalism worthwhile.

Nevertheless, Sneed’s view of the structure of science has some interesting elements. Those familiar with Sneed’s work will recognize the intellectual debt I owe to his clear
The recognition that the empirical content of models is due in a large part to what are commonly referred to as Sneedian constraints. The constraint that the mass of the earth, $M$, has the same value in the context of terrestrial and celestial models is an example of a Sneedian constraint.

The aim of this section is to downplay Kuhn’s attack on theories by shifting the focus to the level of modeling. In my view, the project of understanding the nuts and bolts of modeling is the key to understanding scientific progress in both revolutionary and normal science.

**Normal science in action:** Prior to the discovery of Neptune, in 1846, wiggles in the observed motion of Uranus, the outmost of the seven planets known at the time, were not explained by Newton’s theory of gravitation. Leverrier and Adams investigated the possibility that an eight as-yet-unknown planet caused those wiggles. They did this by assuming that Newton’s theory was true, to obtain a new set of Newtonian equations using the new auxiliary assumption that there were actually eight planets. They then solved the equations of the new model to obtain a family of possible planetary trajectories. Notice that the new model contains an additional adjustable parameter—the mass of the postulated planet. After fitting the model to the anomalous data the model predicted a fairly precise trajectory for the new planet. When telescopes were pointed in that direction, Neptune was sighted and confirmed to be a planet.28 The old Newtonian model was superceded by a better confirmed Newtonian model. The example shows why it is so important to distinguish between theories and models. But it also shows that the entrenched status of the theory did not protect the old model from disconfirmation, when the model is properly individuated.

What happened was that parts of the old model were effectively retained in the new model because the gravitational influence of the new planet had no discernible effect on the motion of the planets near the sun. The confusion arises from the tacit assumption that hypotheses, as a whole, are the objects of confirmation.

**Revolutionary science in progress:** The development of quantum mechanics began in 1900 when Planck introduced his new black-body radiation formula (Fig. 2.6). Every potter knows that the hotter a kiln gets the whiter the light it emits. By viewing the color of the light through a small plug, a potter can judge the temperature of the kiln. The problem is to account for the shape of the wavelength profiles and their dependence on the temperature of the black body. Prior to Planck, there were two laws derived from classical considerations—one that fitted the high wavelength end of the profile and one that fitted the low wavelength end of the profile. Planck’s law immediately superceded both classical models because it fitted all wavelengths simultaneously with a single adjustable parameter (Planck’s constant, $h$). None of the classical models achieved that kind of unified empirical fit.

Spurred by his empirical success, Planck looked for a deeper understanding of his law. A few months later, after a very intense period of work, he published a derivation of

---

28 Planets are recognized by their motion relative to the fixed stars.
his law from the now-famous quantum hypothesis.\textsuperscript{29} It is interesting to note that Planck’s derivation is different from the modern one. As Khinchin (1960, 117-118) says:

“Planck’s derivation of this formula, of course, has nothing in common with the one presented here, because Planck based his work on the tenets of the wave theory of light. (He could not have done otherwise in those times.) Planck could hardly have surmised this his bold idea would prove to the ‘first swallow’ of that tempestuous spring which in this half-century has radically changed the physics of elementary particles and has led mankind to great technical conquests, down to the realization of the ancient dream of mastering atomic energy.”

Nevertheless, Planck knew that further progress depended on there being some kind of connection between his law and other phenomena, and he hoped his quantum hypothesis would forge such links. His hopes were first realized when Einstein used the quantum hypothesis to explain the photoelectric effect in 1905.\textsuperscript{30} In 1902 Lenard had already documented some qualitative features of the photoelectric effect, but Lenard’s data were not precise enough to allow for an independent determination of Planck’s constant. It was Millikan who collected the necessary data in 1914, for which he won the Nobel prize in 1923. After Millikan’s work demonstrated the agreement of the two independent measurements of Planck’s constant, Einstein was also awarded the Nobel prize for his paper on the photoelectric effect.

If there is any difference between the Neptune example and the Planck example, it lies only in the difference between direct and indirect confirmation. In Planck’s case, the direct confirmation is followed by indirect confirmation of independent measurements of Planck’s constant. In the Neptune example, the new model is directly confirmed by first fitting the anomalous data, and then directly confirmed again by the discovery of Neptune. However, this difference is an accidental difference rather than any systematic difference between normal and revolutionary science. The beam balance model, for example, enjoys both kinds of confirmation.

It is the similarity of the examples that is most striking. In each example, the empirical confirmation or disconfirmation of the models proceeds independently of the status of the theory to which they are attached. The entrenchedness of Newton’s theory does not protect the old Newton model from falsification, and Planck’s model of black body radiation is accepted before he derives it from the quantum hypothesis. Moreover, the additional evidence for Planck’s model comes in the form of an independent measurement of Planck’s constant. As Khinchin points out, Planck’s theory is very different from the contemporary quantum theory. The role of the theory is largely heuristic, for once it prompts the invention of Einstein’s model of the photoelectric effect, which is linked to Planck’s model by a common parameter. The agreement of the independent measurements confirms the model even when the theory from which it is derived is undergoing refinement. However the confirmation of theories is defined, Planck’s quantum theory was insecure.

It is not my intention to claim that the quantum hypothesis is irrelevant to the story. Far from it. The point is that doubts about the quantum theory were not sufficient to undermine the claim that Millikan’s data provided indirect support for Planck’s law of radiation, and that the black body data provided independent support for Einstein’s model of the photoelectric effect. The controversy surrounding the theory was relevant to unraveling the meaning of it all, but the tentative nature of those details did not undermine the strength of the confirmation. My thesis is that these and other empirical

\textsuperscript{29} *Annalen der Physik*, vol. 4, 1901.

\textsuperscript{30} Translations of the original papers by Planck and Einstein are reprinted in Shamos 1959.
successes are what credence to parts of the theory, and the search for connections with an ever widening variety of phenomena is what drove the changes in the theory, and not the other way around.

Hempel’s Confirmation “From Above”

Hempel’s highly influential writings on confirmation included a picture of indirect confirmation as trickling down from theories. One part of my response to Kuhn’s attack on theories is to minimize the role of theories in confirmation. I’m against the trickle down theory of confirmation. It’s not that I buy into Kuhn’s total skepticism about theories. On the contrary, the first step towards defending theories is to show that confirmation comes from the bottom up.

What is the competing view? In Hempel’s own words (1966, 38):

The support that may be claimed for a hypothesis need not all be of the inductive-evidential kin that we have considered so far: it need not consist entirely—or even partly—of the data that bear out the test implications derived from it. Support may also come “from above”; that is, from more inclusive hypotheses or theories that imply the given one and have independent evidential support.

In Hempel’s view, the indirect confirmation of models, or laws, depends on the confirmational status of the background theory—the total confirmation of a model is a function of the direct confirmation of the model plus the confirmation of the theory from which it is derived. On the face of it, this picture is puzzling. First, it implies that in branches of science that have no background theories, there is not such thing as indirect confirmation. Second, it is puzzling how the confirmation of theories could be calculated. For surely, it would be a function of the confirmation of its best models. But then, the confirmation of the models depends on the confirmation of the theory. Hempel’s theory of confirmation gets tied up in a circle.

In Hempel’s view, models inherit the strength or weakness of the background theory. Models of strong theories should be hard to falsify, and the models of weak theories will be difficult to positively confirm. But there is no problem with the view that the old seven-planet model was refuted by the anomalous motion of Uranus. It is only a willingness to confuse the model with the theory, or a failure to clearly distinguish this model with its successor that tempts us to say that it was somehow harder to refute because it had such a venerable pedigree. Certainly, Newtonians would have checked that the data was genuine, and reliable. But this is a consideration in evaluation of any model relative to the evidence. On the other side of the same coin, the fact that Planck’s radiation law could only be derived from an unconfirmed theory did not lead anyone to reject Planck’s law. The doubt was whether it was really inconsistent with classical theory. This did affect the interpretation of the model, and this is an aspect of confirmation that I want to emphasize. But to the extent that Planck’s radiation law was neutral between the two theories, it was well confirmed.

Contrary to the deductive theory, the indirect confirmation of Planck’s model, viewed as a model of a new theory, took place entirely at the level of models, in the form of an independent measurement of Planck’s constant. In this example, indirect confirmation has a purely empirical dimension, which is then lends some support to the overarching theory without the need of some mystical idea about theories being pulled up by their bootstraps.
Nevertheless, Hempel (1966, 38-39) presents an example that suggests persuasively that confirmation can flow down from theories:

[Consider] a hypothetical law for free fall on the moon, \( s = 2.7t^2 \). Although none of its least implications have ever been checked by experiments on the moon [this was written prior to the first landing on the moon], it has strong theoretical support, for it follows deductively from Newton’s theory of gravity and of motion (strongly supported by a highly diversified body of evidence) in conjunction with the information that the radius and mass of the moon are .272 and .0123 of those of the earth and that the gravitational acceleration near the surface of the earth is 32.2 feet per second per second.

The general formula for free fall is \( s = \frac{1}{2}at^2 \), where \( s \) is the distance that the object has fallen, and \( a \) is the acceleration due to gravity. This formula is deduced from the definition of acceleration, \( \dot{s} = a \), where \( \dot{s} \) denotes the second derivative of the distance fallen with respect to time. The only non-definitional assumption required is that the acceleration is constant in time, and does not depend on the mass of the falling object. Near the surface of the earth, the acceleration due to gravity is denoted by \( g \), which is as Hempel notes, approximately equal 32.2 feet per second per second. As proven in chapter 1, \( g \) is related to the earth’s mass by the formula \( g = \frac{M}{D^2} \), where \( M \) is the mass of the earth, and \( D \) is the distance to the center of the earth. On the moon, \( a = \frac{m}{d^2} \), where \( m \) is the mass of the moon, and \( d \) is the radius of the moon. If the ratio of \( d \) to \( D \) and the ratio of \( m \) to \( M \) are known, then we have all the information required to infer that \( s = 2.7t^2 \), or more exactly, \( s = 2.67666t^2 \). The acceleration near the surface of the moon is therefore 5.35 feet per second per second, which is 1/6 of acceleration on the earth. In other words, the deduction can be made using the same simple auxiliary assumptions used in chapter 1—assume that the moon is spherically symmetrical, assume that that only gravitational forces are acting, and so on.

Then again, the earth-moon model discussed in chapter 1 is too simple to provide any means of estimating the mass of the moon. So, in fact the prediction that Hempel refers to is made by a more complicated model—one that measures the motion of the center of gravity of the earth-moon system in its orbit around the sun. That is, it requires information about the monthly variation in the distance of the earth from the sun, for that tells us where the center of gravity is, which then allows us to estimate the ratio of \( m \) to \( M \). This information was available to Newton, and it was from this calculation that Newton predicted that the earth’s core was more dense than its crust. The mass of the moon is estimated independently from tidal phenomena, in particular from comparing the effect of the moon on the tides with the effect of the sun on the tides.

Now that the example is well understood, we can see that the confirmation of the hypothesis is indirect in exactly the same sense that the moon’s motion around the earth provides indirect confirmation of Galileo’s law of free fall. The main difference is that there is no direct confirmation. Hempel’s example is good example of how the independent measurements of parameters lend support to further prediction that new independent measurements will yield the same result.

Once again, the example raises questions about the object of confirmation. There is a clear logical distinction to be made between predicting that the value of the moon’s mass and predicting that the law of free fall on the moon’s surface will be true. For the first prediction could be true while the second is false. Perhaps there is a strange kind of force that acts on bodies very near the moon’s surface. If this were the case, then the law
of free fall would be false, but a corrected form of the law could still provide an estimate of the moon’s mass that agrees with the predicted value. The logical gap between the two hypotheses is bridged by auxiliary assumptions, and it is our confidence in those assumptions explains why we tend to view the two predictions as equally secure.

Nevertheless, Hempel would insist also that each prediction depends on our confidence in Newton’s theory as well. For instance, if predicted law of free fall is wrong, but the predicted value of the moon’s mass is right, then we would show that by showing that the same gravitational force is acting as a component force, and that it combines with other forces by simple addition. Such assumptions, are central to the theory itself. Surely our confidence in these principles rests, in part, on the fact that they are borne out in other applications of the theory that are quite unconnected with gravitational phenomena (for example, the motion of an ice hockey puck connected to two springs).

To tighten our grip on the issue, consider the following thought experiment. Suppose we are in radio communication with alien scientists who live in a distant solar system that cannot be seen from earth. Would be justified in predicting that the motion of the planets in their solar system are governed by Newton’s theory of gravitation? I agree with Hempel that this answer is “yes”, there is some justification for this prediction, even though the justification is never 100% (the gravitational forces might be strong enough to produce relativistic effects). This scenario is close to the real cases in which we predict confidently that twin stars will obey Kepler’s laws of motion to a good approximation (having verified that they are sufficiently far apart to neglect relativistic effects).

Hempel’s example supports the hunch that there are other forms of indirect confirmation besides that provide via the use of shared adjustable parameters. The key question is whether this other kind of indirect confirmation derives solely from the fact that two models, sharing no parameters in common, are deduced from the same theory. There is good reason to think that this cannot be the whole story. For any two unconnected models can be deduced from a single theory. For a start, the conjunction of the two models will do the job. But there are even non-gerrymandered theories that will work. Suppose, for example, that the alien solar system obeys an inverse cube law of gravitation. Then all we need to do it to introduce a new parameter, $\chi$, such that $2^\chi$ if the solar system is ours, and $3^\chi$ if the solar system is alien. Then a generalized law of gravitation could be written:

$$F = GmM/r^\chi.$$

This presents a problem for the deductive account. For if we make the prediction from Newton’s inverse square law, then we shall predict that the alien planets will obey that law, but if we use the generalized law, we will tell them to expect their planets to obey the inverse cube law. If deduction is the only thing that matters, then the two predictions are equally justified. But they are not, so something more is going on than mere deduction.

I don’t have a general resolution of this puzzle (which is reminiscent of the infamous grue problem). The only obvious thing to say is that the law introduces one additional adjustable parameter, and is therefore more complex. But sometimes complex laws are true, so we can’t reject the generalized law on a priori grounds (nor did Newton).

Secondly, $\chi$ is measurable. If we entertain the law, then we have many independent determinations of the parameter, which all agree that it’s value is indiscernibly close to 2. The question is whether it is universally equal to 2. The obvious argument is that we have no evidence that it differs from 2, even after examining the motions of systems
outside our solar system such as twin stars. A prediction based on the auxiliary assumption the $\chi = 3$ has no empirical basis, even though it is logically consistent with the empirical evidence.

The problem is that we have no direct empirical grounds for extending this value to an unseen solar system. But isn’t that the point? The worry arises from a psychological tendency to downplay the role of indirect confirmation, combined with a tendency to view confirmation in black and white: If we view indirect confirmation as having any weight, then it is conclusive. We predict that the inverse square law will hold. But if the prediction must be true, then no direct confirmation is needed. Therefore direct confirmation has no weight. The conclusion is clearly wrong. To avoid the absurd conclusion, we need to reject the idea that indirect confirmation is conclusive. In fact, we need to reject the view that any confirmation is conclusive.

The insight I have to offer is that even abstract features of our models, such as the form of the equations or the additivity of forces, should be viewed as something that is independently “measured” or verified, and that the extrapolation of these structural features new applications of the theory is similar to the extrapolation of parameter values. Extrapolation of any kind is always risky, and the evaluation of the risks presents a difficult problem. Nevertheless, it seems to be universally true in science that the risk depends, to some extent, on whether the feature or structure in question is independently verified. Certain parts of the formalism of Newtonian mechanics, such as some features of absolute space, are not empirically determined at all, as Leibniz showed. On the other hand, it may be that features of strange formalisms, such the commutation relations of quantum mechanical operators, are independently determined by different applications of the theory, in which case they should be extended to new applications. It would be interesting to know whether the development of quantum theory was driven by such considerations.

The recommended epistemological principle boils down to something like this: It is an important to the evaluation of a theory if it can be shown that a more liberal version of the theory reduces to the more specific version if we are able to determine which elements of the more liberal theory have no empirical justification. It seems clear to me that it is impossible to eliminate all unjustified aspects of a theory, partly because it may be too difficult from a mathematical point of view, and partly because it may be impossible to clearly decompose a theory into parts in a principled way. Or else the evidence may be too weak to draw a clear conclusion. Nevertheless, it seems that much could be learnt about the empirical support for a theory by carrying out such an analysis as diligently as possible.

The discussion has led to the same conclusion as before: Careful attention should be paid to the detailed relationship between a theory and its evidence. More particularly, the relationship is not as simple as the mere deductibility of models from some more inclusive theory. The aim of this book is to say why much more than deductive logic is relevant, at least with respect to narrower form of indirect confirmation about by the independent measurement of adjustable parameters.

The broader suggestion of this section is that, while the deductive theory of indirect confirmation is wrong, there is more to indirect confirmation the agreement of independent measurements of theoretical parameters. I plan to show this explicitly in the case of quantum mechanics. This would also motivate the idea that the notion of indirect

---

31 For example, Friedman (1983) defended a realist view of Newton’s theory by reformulating the theory in a co-ordinate free representation that was immune from Leibniz’s attack.
confirmation, as it applies to physics, could extend to the philosophy of the non-quantitative sciences.

Kuhn and the Quine-Duhem Thesis

The Quine-Duhem thesis states that it is always logically possible for a discrepancy between a model and data to be removed by a model derived from a new set of auxiliary assumptions. It is a strong thesis. If it is true, then any theory is logically compatible with any set of observations—a theory can always be made to “save the phenomena” if scientists are sufficiently resourceful. It says that the function of theories is purely heuristic—only role is to generate models. If theories have no empirical content, then there is no problem about the empirical confirmation of theories because they have no empirical content to confirm. If theories are true, then they are true by convention, or by definition—in either case they are mere tautologies. For Kuhn (1970), this would explain why the acceptance of theories appears to be based on a ‘religious’ kind of faith, or psychological indoctrination, or a political kind of persuasion.

It may be surprising to learn that there is a simple argument for Quine-Duhem thesis. For if there are no constraints on what can count as auxiliary assumptions, then one can take any model and any theory, and find an auxiliary assumption that allows the model to be deduced from the theory. In fact, I can give you the general recipe: Add the auxiliary assumption ‘if Theory then Model’. Since modus ponens is a valid argument form, Model follows from Theory in conjunction with this auxiliary assumption. Of course, no-one is impressed by this proof that an Einsteinian model of planetary motion can be deduced from Newton’s theory of gravitation. Clearly, there are extra-logical considerations that determine what can count as a genuine auxiliary assumption. It is a problem for philosophers, not scientists, to say what these additional constraints are. But if such constraints can be found, then it would reduce the plausibility of the Quine-Duhem thesis.

One requirement is that auxiliary assumptions should be couched solely within the vocabulary of the theory. This rules out the use of the assumption ‘if Theory then Model’ whenever the Einsteinian model makes use of non-Newtonian language. Moreover, the requirement is sensible in light of what was said the previous section. For the confirmation of a distinctly Einsteinian model will not provide independent measurements of Newtonian parameters, nor of other features of the Newtonian formalism.

A second consideration is that many auxiliary assumptions are supported by independent evidence, and are therefore not so easily abandoned. For example, an auxiliary assumption may concern the workings of a measuring device such as an optical telescope. Or it may concern a well understood source of observational error such as stellar aberration. Stellar aberration causes the apparent direction of a celestial body to differ slightly from its true direction. The effect arises from the fact that the vector sum of the earth’s velocity and the velocity of light is slightly different from the direction of the incoming light. The size of the aberration is greatest when the earth is moving perpendicular to the line of sight to the planet. However, the error can calculated precisely from known velocities and the angle between them, so the error is easily removed. The anomalies in the motion of Uranus were not blamed on stellar aberration, or the workings of the telescope. More generally, the rejection of any independently confirmed auxiliary assumptions will not solve the problem if the anomalous bubble pops up somewhere else under the proverbial linoleum.
If the Leverrier-Adams eight-planet model had not resolved the puzzle, it is not clear that something other Newtonian model would have. An example that comes to mind is the anomalous precession of Mercury’s perihelion. The usual remedies did not work, even though that fact does not prove that it is logically impossible.

While the truth of the Quine-Duhem thesis may be unclear from any direct analysis, Kuhn argued independently against theories as truth-bearers. He looked at examples of revolutions and asked whether scientists at the time provided clear-cut reasons the theory change. Kuhn (1970) draws a negative conclusion. I shall only explain Kuhn’s conclusion—the interested reader is referred to his book for the detailed evidence.

Kuhn attributes the same idea to another historian of science. As Butterfield (1962, 1-7) expressed it, “of all the forms of mental activity” in scientific revolutions, “the most difficult to induce…is the art of handling the same data as before, placing them in a new system of relations with one another by giving them a different framework.” Kuhn (1970, 85) carries the conclusion a step further: “Others who have noted this aspect of scientific advance have emphasized its similarity to a change in visual gestalt: the marks on paper that were first seen as a bird are now seen as an antelope, or vice versa.” To use the duck-rabbit visual gestalt instead (Fig. 2.6), the marks on the paper represent the data, which are intrinsically the same whether or not the drawing is seen as a duck or as a rabbit. However, the different ways of perceiving the figure shows that there are different ways of judging the significance or the salience of some of its features. For example, the kink at the back of the duck’s head is ‘noise’ when it is seen as a duck, while it is an essential feature when it is seen as the mouth of the rabbit.

Kuhn (1970, 94) uses this analogy to explain why the argument that scientists give for preferring their theory are very often entrenched within the language of the theory itself. As a result, they only succeed in preaching to the converted. There appears to be surprisingly little effort, or success, in the articulation the reasons in terms of a language that is neutral between the rival theories. From this historical evidence, Kuhn leaps to the philosophical conclusion that there is no sense in which theories progress towards the truth. Or as he puts it (Kuhn 1970, 94) “As in political revolutions, so in paradigm choice—there is no standard higher than the assent of the relevant community”. For Kuhn, the word ‘revolution’ is much more than a metaphor.

When one thinks about Kuhn’s metaphor more critically, we see that Kuhn just assumes that the process of placing the same data in a new system of relations in a different framework is something that a theory imposes on the data. After all, how could there be empirical support for new relations from the data themselves? If there were empirical facts that supported new relations, then these facts would be new theory-dependent features of the data, which would imply that the data is not a neutral arbiter between theories. This is exactly the incoherent view that Kuhn endorses when he says that there is no theory-neutral language of observation. It’s a seductive view, but Kuhn has go it wrong. New theories introduce new parameters, such as Planck’s constant, or Newton’s gravitational mass, that predict a new relations in the data. They do no impose new relations on the data. Nature has the final word in saying whether the new relations hold, or not. That is what indirect confirmation is all about. Even in comparing models of different theories, there is a standard higher standard than the assent of the relevant community.
The problem is that Kuhn, like Popper, has no clear distinction between models and theories. Indeed, in other places, he says that science does make progress in puzzle solving, where he is now referring to the progress in modeling. But his sense of progress is limited in its nature: “Compared to the notion of progress most prevalent among both philosophers of science and laymen, however, this position lacks an essential element.” In his view, progress is merely progress in finding better fitting models:

A scientific theory is usually felt to be better than its predecessors not only in the sense that it is a better instrument for discovering and solving puzzles but also because it is somehow a better representation of what nature is really like. One often hears that successive theories grow ever closer to, or approximate more and more closely to, the truth. Apparently generalizations like that refer not to the puzzle-solutions and the concrete predictions derived from a theory but rather to its ontology, to the match, that is, between the entities with which the theory populates nature and what is “really there.” (Kuhn 1970, 206)

Kuhn is right to be skeptical of some of the ontological claims make by theories. Crystalline spheres in ancient astronomy is one example, and Newton’s claims about absolute velocities is another example. He is also right that theories can succeed in performing their heuristic function of producing models even though they carry this excess ontological baggage. The point is a logical one. If \( X \) represents a statement about the existence crystalline spheres in the heavens, or Newton’s assumption that the sun has an absolute velocity, and theory \( T \) implies \( X \), then model \( M \) follows from \( T \) just as well as it follows from \( T \) minus \( X \). Therefore, a theory’s success in puzzle solving is not reason for believing in \( X \). This much is clearly correct. However, there is a logical gap between taking a skeptical view of \( T \) and taking a skeptical view of \( T \) minus \( X \).

The issue comes back to what is confirmed when a model is confirmed. And this is where the details of how confirmation works become relevant. The agreement of independent measurements pinpoints which parts of the theory’s ontology are not redundant to the derivation of the confirmed model. There is no need to accept all the whole theory, lock, stock and barrel. The kind of realism endorsed here is against reading the ontology passively from the best known theory. Rather, we need to read an ontology into a theory, on the basis of its best confirmed model, in order to make the best sense of the model’s detailed relationship with the evidence.

The argument I am offering is based on many promissory notes. Foremost among these is the assumption that there is a solution to the problem of many models. It may seem that the solution is easy—the best confirmed model is the one that fits the data the best. In fact, this solution does not work because it ignores relational facts in the data. The mistake made by the naïve empiricist picture of confirmation is the same mistake made in the interpretation of Kuhn’s duck-rabbit metaphor.