Two

THE CONFIRMATION OF SCIENTIFIC HYPOTHESES

John Earman and Wesley C. Salmon

In Chapter 1 we considered the nature and importance of scientific explanation. If we are to be able to provide an explanation of any fact, particular or general, we must be able to establish the statements that constitute its explanans. We have seen in the Introduction that many of the statements that function as explanans cannot be established in the sense of being conclusively verified. Nevertheless, these statements can be supported or confirmed to some degree that falls short of absolute certainty. Thus, we want to learn what is involved in confirming the kinds of statements used in explanations, and in other scientific contexts as well.

This chapter falls into four parts. Part I (Sections 2.1–2.4) introduces the problem of confirmation and discusses some attempts to explicate the qualitative concept of support. Part II (2.5–2.6) reviews Hume’s problem of induction and some attempted resolutions. Part III (2.7–2.8) develops the mathematical theory of probability and discusses various interpretations of the probability concept. Finally, Part IV (2.9–2.10) shows how the probability apparatus can be used to illuminate various issues in confirmation theory.

Parts I, II, and III can each stand alone as a basic introduction to the topic with which it deals. These three parts, taken together, provide a solid introduction to the basic issues in confirmation, induction, and probability. Part IV covers more advanced topics. Readers who prefer not to bring up Hume’s problem of induction can omit Part II without loss of continuity.
2.1 EMPIRICAL EVIDENCE

The physical, biological, and behavioral sciences are all empirical. This means that their assertions must ultimately face the test of observation. Some scientific statements face the observational evidence directly; for example, “All swans are white,” was supported by many observations of European swans, all of which were white, but it was refuted by the observation of black swans in Australia. Other scientific statements confront the observational evidence in indirect ways; for instance, “Every proton contains three quarks,” can be checked observationally only by looking at the results of exceedingly complex experiments. Innumerable cases, of course, fall between these two extremes.

Human beings are medium-sized objects; we are much larger than atoms and much smaller than galaxies. Our environment is full of other medium-sized things—for example, insects, frisbees, automobiles, and skyscrapers. These can be observed with normal unaided human senses. Other things, such as microbes, are too small to be seen directly; in these cases we can use instruments of observation—microscopes—to extend our powers of observation. Similarly, telescopes are extensions of our senses that enable us to see things that are too far away to be observed directly. Our senses of hearing and touch can also be enhanced by various kinds of instruments. Ordinary eyeglasses—in contrast to microscopes and telescopes—are not extensions of normal human senses; they are devices that provide more normal sight for those whose vision is somewhat impaired.

An observation that correctly reveals the features—such as size, shape, color, and texture—of what we are observing is called veridical. Observations that are not veridical are illusory. Among the illusory observations are hallucinations, afterimages, optical illusions, and experiences that occur in dreams. Philosophical arguments going back to antiquity show that we cannot be absolutely certain that our direct observations are veridical. It is impossible to prove conclusively, for example, that any given observation is not a dream experience. That point must be conceded. We can, however, adopt the attitude that our observations of ordinary middle-sized physical objects are reasonably reliable, and that, even though we cannot achieve certainty, we can take measures to check on the veridicality of our observations and make corrections as required (see Chapter 4 for further discussion of the topics of skepticism and antirealism).

We can make a rough and ready distinction among three kinds of entities: (i) those that can be observed directly with normal unaided human senses; (ii) those that can be observed only indirectly by using some instrument that extends the normal human senses; and (iii) those that cannot be observed either directly or indirectly, whose existence and nature can be established only by some sort of theoretical inference. We do not claim that these distinctions are precise; that will not matter for our subsequent discussion. We say much more about category (iii) and the kinds of inferences that are involved as this chapter develops.

Our scientific languages should also be noted to contain terms of two types. We
have an observational vocabulary that contains expressions referring to entities, properties, and relations that we can observe. "Tree," "airplane," "green," "soft," and "is taller than" are familiar examples. We also have a theoretical vocabulary containing expressions referring to entities, properties, and relations that we cannot observe. "Microbe," "quark," "electrically charged," "ionized," and "contains more protons than" exemplify this category. The terms of the theoretical vocabulary tend to be associated with the unobservable entities of type (iii) of the preceding paragraph, but this relationship is by no means precise. The distinction between observational terms and theoretical terms—like the distinction among the three kinds of entities—is useful, but it is not altogether clear and unambiguous. One further point of terminology. Philosophers often use the expression "theoretical entity," but it would be better to avoid that term and to speak either of theoretical terms or unobservable entities.

At this point a fundamental moral concerning the nature of scientific knowledge can be drawn. It is generally conceded that scientific knowledge is not confined to what we have observed. Science provides predictions of future occurrences—such as the burnout of our sun when all of its hydrogen has been consumed in the synthesis of helium—that have not yet been observed and that may never be observed by any human. Science provides knowledge of events in the remote past—such as the extinction of the dinosaurs—before any human observers existed. Science provides knowledge of other parts of the universe—such as planets orbiting distant stars—that we are unable to observe at present. This means that much of our scientific knowledge depends upon inference as well as observation. Since, however, deductive reasoning is nonampliative (see Chapter 1, Section 1.5), observations plus deduction cannot provide knowledge of the unobserved. Some other mode of inference is required to account for the full scope of our scientific knowledge.

2.2 THE HYPOTHETICO-DEDUCTIVE METHOD

As we have seen, science contains some statements that are reports of direct observation, and others that are not. When we ask how statements of this latter type are to meet the test of experience, the answer often given is the hypothetico-deductive (H-D) method; indeed, the H-D method is sometimes offered as the method of scientific inference. We must examine its logic.

The term hypothesis can appropriately be applied to any statement that is intended for evaluation in terms of its consequences. The idea is to articulate some statement, particular or general, from which observational consequences can be drawn. An observational consequence is a statement—one that might be true or might be false—whose truth or falsity can be established by making observations. These observational consequences are then checked by observation to determine whether they are true or false. If the observational consequence turns out to be true, that is said to confirm the hypothesis to some degree. If it turns out to be false, that is said to disconfirm the hypothesis.

Let us begin by taking a look at the H-D testing of hypotheses having the form of universal generalizations. For a very simple example, consider Boyle's law of
gases, which says that, for any gas kept at a constant temperature $T$, the pressure $P$ is inversely proportional to the volume $V$,\(^1\) that is,

$$ P \times V = \text{constant (at constant } T). $$

This implies, for example, that doubling the pressure on a gas will reduce its volume by a half. Suppose we have a sample of gas in a cylinder with a movable piston, and that the pressure of the gas is equal to the pressure exerted by the atmosphere—about 15 pounds per square inch. It occupies a certain volume, say, 1 cubic foot. We now apply an additional pressure of 1 atmosphere, making the total pressure 2 atmospheres. The volume of the gas decreases to $\frac{1}{2}$ cubic foot. This constitutes a hypothetico-deductive confirmation of Boyle’s law. It can be schematized as follows:

1. At constant temperature, the pressure of a gas is inversely proportional to its volume (Boyle’s law).
   - The initial volume of the gas is 1 cubic ft.
   - The initial pressure is 1 atm.
   - The pressure is increased to 2 atm.
   - The temperature remains constant.
   - The volume decreases to $\frac{1}{2}$ cubic ft.

Argument (1) is a valid deduction. The first premise is the hypothesis that is being tested, namely, Boyle’s law. It should be carefully noted, however, that Boyle’s law is not the only premise of this argument. From the hypothesis alone it is impossible to deduce any observational prediction; other premises are required. The following four premises state the initial conditions under which the test is performed. The conclusion is the observational prediction that is derived from the hypothesis and the initial conditions. Since the temperature, pressure, and volume can be directly measured, let us assume for the moment that we need have no serious doubts about the truth of the statements of initial conditions. The argument can be schematized as follows:

2. $H$ (test hypothesis)
   - $I$ (initial conditions)
   - $O$ (observational prediction)

When the experiment is performed we observe that the observational prediction is true.

As we noted in Chapter 1, it is entirely possible for a valid deductive argument to have one or more false premises and a true conclusion; consequently, the fact that (1) has a true conclusion does not prove that its premises are true. More specifically, we cannot validly conclude that our hypothesis, Boyle’s law, is true just because the observational prediction turned out to be true. In (1) the argument from premises to

---

\(^1\) This relationship does not hold for temperatures and pressures close to the point at which the gas in question condenses into a liquid or solid state.
conclusion is a valid deduction but the argument from the conclusion to the premises is not. If it has any merit at all, it must be as an inductive argument.

Let us reconstruct the argument from the observational prediction to the hypothesis as follows:

(3) The initial volume of the gas is 1 cubic ft.
   The initial pressure is 1 atm.
   The pressure is increased to 2 atm.
   The temperature remains constant.
   The volume decreases to ½ cubic ft.

At constant temperature, the pressure of a gas is inversely proportional to its volume (Boyle’s law).

No one would seriously suppose that (3) establishes Boyle’s law conclusively, or even that it renders the law highly probable. At best, it provides a tiny bit of inductive support. If we want to provide solid inductive support for Boyle’s law it is necessary to make repeated tests of this gas, at the same temperature, for different pressures and volumes, and to make other tests at other temperatures. In addition, other kinds of gases must be tested in a similar manner.

In one respect, at least, our treatment of the test of Boyle’s law has been oversimplified. In carrying out the test we do not directly observe—say by feeling the container—that the initial and final temperatures of the gas are the same. Some type of thermometer is used; what we observe directly is not the temperature of the gas but the reading on the thermometer. We are therefore relying on an auxiliary hypothesis to the effect that the thermometer is a reliable instrument for the measurement of temperature. On the basis of an additional hypothesis of this sort we claim that we can observe the temperature indirectly. Similarly, we do not observe the pressures directly, by feeling the force against our hands; instead, we use some sort of pressure gauge. Again, we need an auxiliary hypothesis stating that the instrument is a reliable indicator.

The need for auxiliary hypotheses is not peculiar to the example we have chosen. In the vast majority of cases—if not in every case—auxiliary hypotheses are required. In biological and medical experiments, for example, microscopes of various types are employed—from the simple optical type to the tunneling scanning electron microscope, each of which requires a different set of auxiliary hypotheses. Likewise, in astronomical work telescopes—refracting and reflecting optical, infrared, radio, X-ray, as well as cameras are used. The optical theory of the telescope and the chemical theory of photographic emulsions are therefore required as auxiliary hypotheses. In sophisticated physical experiments using particle accelerators, an elaborate set of auxiliary hypotheses concerning the operation of all of the various sorts of equipment is needed. In view of this fact, schema (2) should be expanded:

(4) $H$ (test hypothesis)
    $A$ (auxiliary hypotheses)
    $I$ (initial conditions)
    $O$ (observational prediction)
Up to this point we have considered the case in which the observational prediction turns out to be true. The question arises, what if the observational prediction happens to be false? To deal with this case we need a different example.

At the beginning of the nineteenth century a serious controversy existed about the nature of light. Two major hypotheses were in contention. According to one theory light consists of tiny particles; according to the other, light consists of waves. If the corpuscular theory is true, a circular object such as a coin or ball bearing, if brightly illuminated, will cast a uniformly dark circular shadow. The following H-D test was performed:

(5) Light consists of corpuscles that travel in straight lines.\(^2\)
A circular object is brightly illuminated.
The object casts a uniform circular shadow.

Surprisingly, when the experiment was performed, it turned out that the shadow had a bright spot in its center. Thus, the result of the test was negative; the observational prediction was false.

Argument (5) is a valid deduction; accordingly, if its premises are true its conclusion must also be true. But the conclusion is not true. Hence, at least one of the premises must be false. Since the second premise was known to be true on the basis of direct observation, the first premise—the corpuscular hypothesis—must be false.

We have examined two examples of H-D tests of hypotheses. In the first, Boyle’s law, the outcome was positive—the observational prediction was found to be true. We saw that, even assuming the truth of the other premises in argument (1), the positive outcome could, at best, lend a small bit of support to the hypothesis. In the second, the corpuscular theory of light, the outcome was negative—the observational prediction was found to be false. In that case, assuming the truth of the other premise, the hypothesis was conclusively refuted.

The negative outcome of an H-D test is often less straightforward than the example just discussed. For example, astronomers who used Newtonian mechanics to predict the orbit of the planet Uranus found that their observational predictions were incorrect. In their calculations they had, of course, taken account only of the gravitational influences of the planets that were known at the time. Instead of taking the negative result of the H-D test as a refutation of Newtonian mechanics, they postulated the existence of another planet that had not previously been observed. That planet, Neptune, was observed shortly thereafter. An auxiliary hypothesis concerning the constitution of the solar system was rejected rather than Newtonian mechanics.

It is interesting to compare the Uranus example with that of Mercury. Mercury also moves in a path that differs from the orbit calculated on the basis of Newtonian mechanics. This irregularity, however, could not be successfully explained by postulating another planet, though this strategy was tried. As it turned out, the perturbation of Mercury’s orbit became one of three primary pieces of evidence supporting Einstein’s general theory of relativity—the theory that has replaced Newtonian me-

\(^2\) Except when they pass from one medium (e.g., air) to another medium (e.g., glass or water).
mechanics in the twentieth century. The moral is that negative outcomes of H-D tests sometimes do, and sometimes do not, result in the refutation of the test hypothesis. Since auxiliary hypotheses are almost always present in H-D tests, we must face the possibility that an auxiliary hypothesis, rather than the test hypothesis, is responsible for the negative outcome.

2.3 PROBLEMS WITH THE HYPOTHETICO-DEDUCTIVE METHOD

The H-D method has two serious shortcomings that must be taken into account. The first of these might well be called the problem of alternative hypotheses. Let us reconsider the case of Boyle's law. If we represent that law graphically, it says that a plot of pressures against volumes is a smooth curve, as shown in Figure 2.1.

The result of the test, schematized in argument (1), is that we have two points (indicated by arrows) on this curve—one corresponding to a pressure of 1 atmosphere and a volume of 1 cubic foot, the other corresponding to a pressure of 2 atmospheres and a volume of \( \frac{1}{2} \) cubic foot. While these two points conform to the solid curve shown in the figure, they agree with infinitely many other curves as well—for example, the dashed straight line through those two points. If we perform another test, with a pressure of 3 atmospheres, we will find that it yields a volume of \( \frac{1}{2} \) cubic foot. This is incompatible with the straight line curve, but the three points we now have are still compatible with infinitely many curves, such as the dotted one, that go through these three. Obviously we can make only a finite number of tests; thus, it is clear that, no matter how many tests we make, the results will be compatible with infinitely many different curves.

![Figure 2.1](image)

The Confirmation of Scientific Hypotheses
This fact poses a profound problem for the hypothetico-deductive method. Whenever an observational result of an H-D test confirms a given hypothesis, it also confirms infinitely many other hypotheses that are incompatible with the given one. In that case, how can we maintain that the test confirms our test hypothesis in preference to an infinite number of other possible hypotheses? This is the problem of alternative hypotheses. The answer often given is that we should prefer the simplest hypothesis compatible with the results of the tests. The question then becomes, what has simplicity got to do with this matter? Why are simpler hypotheses preferable to more complex ones? The H-D method, as such, does not address these questions.

The second fundamental problem for the H-D method concerns cases in which observational predictions cannot be deduced. The situation arises typically where statistical hypotheses are concerned. This problem may be called the problem of statistical hypotheses. Suppose, to return to an example cited in Chapter 1, that we want to ascertain whether massive doses of vitamin C tend to shorten the duration of colds. If this hypothesis is correct, the probability of a quick recovery is increased for people who take the drug. (As noted in Chapter 1, this is a fictitious example; the genuine question is whether vitamin C lessens the frequency of colds.) As suggested in that chapter, we can conduct a double-blind controlled experiment. However, we cannot deduce that the average duration of colds among people taking the drug will be smaller than the average for those in the control group. We can only conclude that, if the hypothesis is true, it is probable that the average duration in the experimental group will be smaller than it is in the control group. If we predict that the average duration in the experimental group will be smaller, the inference is inductive. The H-D method leaves no room for arguments of this sort. Because of the pervasiveness of the testing of statistical hypotheses in modern science, this limitation constitutes a severe shortcoming of the H-D method.

2.4 OTHER APPROACHES TO QUALITATIVE CONFIRMATION

The best known alternative to the H-D method is an account of qualitative confirmation developed by Carl G. Hempel (1945). The leading idea of Hempel’s approach is that hypotheses are confirmed by their positive instances. Although seemingly simple and straightforward, this intuitive idea turns out to be difficult to pin down. Consider, for example, Nicod’s attempt to explicate the idea for universal conditionals; for example:

\[ H: (x) (Rx \supset Bx) \text{ (All ravens are black).} \]

(The symbol “\((x)\)” is the so-called universal quantifier, which can be read, “for every object \(x\); ” “\(\supset\)” is the sign of material implication, which can be read very roughly “if . . . then . . . ”.) Although this statement is too simpleminded to qualify as a serious scientific hypothesis, the logical considerations that will be raised apply to all universal generalizations in science, no matter how sophisticated—see Section
2.11 of this chapter. According to Nicod, E Nicod-confirms such an H just in case E
implies that some object is an instance of the hypothesis in the sense that it satisfies
both the antecedent and the consequent, for example, E is Ra.Ba (the dot means
"and"); it is the symbol for conjunction). To see why this intuitive idea runs into
trouble, consider a plausible constraint on qualitative confirmation.

Equivalence condition: If E confirms H and \( \vdash H \equiv H' \), then E confirms H'.

(The triple bar "\( \equiv \)" is the symbol for material equivalence; it can be translated very
roughly as "if and only if," which is often abbreviated "iff." The turnstile "\( \vdash \)"
prefacing a formula means that the formula is a truth of logic.) The failure of this
condition would lead to awkward situations since then confirmation would depend
upon the mode of presentation of the hypothesis. Now consider

\[ H' : (x) (\sim Bx \supset \sim Rx). \]

(The tilde "\( \sim \)" signifies negation; it is read simply as "not.") H' is logically
equivalent to H. But

E: Ra.Ba
does not Nicod-confirm H' although it does Nicod-confirm H. Or consider

\[ H'' : (x) [(Rx \sim Bx) \supset (Px \sim Px)]. \]

Again H'' is logically equivalent to H. But by logic alone, nothing can satisfy the
consequent of H'' and if H is true nothing can satisfy the antecedent. So if H is true
nothing can Nicod-confirm H''.

After rejecting the Nicod account because of these and other shortcomings,
Hempel's next step was to lay down what he regarded as conditions of adequacy for
qualitative confirmation—that is, conditions that should be satisfied by any adequate
definition of qualitative confirmation. In addition to the equivalence condition there
are (among others) the following:

Entailment condition: If E \( \vdash H \), the E confirms H.

(When the turnstile is preceded by a formula ("\( \vdash \)" in "E \( \vdash H' \)"), it means that
whatever comes before the turnstile logically entails that which follows the turnstile—E
logically entails H.)

3 Such examples might lead one to try to build the equivalence condition into the definition of Nicod-
confirmation along the following lines:

(N') E Nicod-confirms H just in case there is an H' such that \( \vdash H \equiv H' \) and such that E implies that
the objects mentioned satisfy both the antecedent and consequent of H'.

But as the following example due to Hempel shows, (N') leads to confirmation where it is not wanted in the case
of multiply quantified hypotheses. Consider

\[ H : (x) (y) Rxy \]
\[ H' : (x) (y)[\sim(Rxy \cdot Ryx) \supset (Rxy \sim Ryx)] \]
E: Rab. \( \sim \) Rba

E implies that the pair a, b satisfies both the antecedent and the consequent of H', and H' is logically equivalent
to H. So by (N') E Nicod-confirms H. But this is an unacceptable result since E contradicts H.

The Confirmation of Scientific Hypotheses
Special consequence condition: If $E$ confirms $H$ and $H \vdash H'$ then $E$ confirms $H'$.

Consistency condition: If $E$ confirms $H$ and also confirms $H'$ then $H$ and $H'$ are logically consistent.

As a result, he rejects

Converse consequence condition: If $E$ confirms $H$ and $H' \vdash H$ then $E$ confirms $H'$.

For to accept the converse consequence condition along with the entailment and special consequence conditions would lead to the disastrous result that any $E$ confirms any $H$. (Proof of this statement is one of the exercises at the end of this chapter.) Note that the H-D account satisfies the converse consequence condition but neither the special consequence condition nor the consistency condition.

Hempel provided a definition of confirmation that satisfies all of his adequacy conditions. The key idea of his definition is that of the development, $dev_f(H)$, of a hypothesis $H$ for a set $I$ of individuals. Intuitively, $dev_f(H)$ is what $H$ says about a domain that contains exactly the individuals of $I$. Formally, universal quantifiers are replaced by conjunctions and existential quantifiers are replaced by disjunctions. For example, let $I = \{a, b\}$, and take

$$H: (x) \, Bx \text{ (Everything is beautiful)}$$

then

$$dev_f(H) = Ba.Bb.$$  

Or take

$$H': (\exists x) \, Rx \text{ (Something is rotten)}$$

then

$$dev_f(H') = Ra \lor Rb.$$  

(The wedge "\lor" symbolizes the inclusive disjunction; it means "and/or"—that is, "one, or the other, or both.") Or take

$$H'': (x) \, (\exists y) \, Lxy \text{ (Everybody loves somebody)};$$

then

$$dev_f(H'') = (Laa \lor Lab).(Lba \lor Lbb).\text{\footnote{In formulas like $H''$ that have mixed quantifiers, we proceed in two steps, working from the inside out. In the first step we replace the existential quantifier by a disjunction, which yields $(x) \,(Lxa \lor Lxb).$ In the next step we replace the universal quantifier with a conjunction, which yields $dev_f(H'')$.}}$$

Using this notion we can now state the main definitions:

The Confirmation of Scientific Hypotheses 51
Def. $E$ directly-Hempel-confirms $H$ just in case $E \vdash \text{dev}(H)$ for the class $I$ of individuals mentioned in $E$.

Def. $E$ Hempel-confirms $H$ just in case $E$ directly confirms every member of a set of sentences $K$ such that $K \vdash H$.

To illustrate the difference between the two definitions, note that $Ra.Ba$ does not directly-Hempel-confirm $Rb \supset Bb$ but it does Hempel-confirm it. Finally, disconfirmation can be handled in the following manner.

Def. $E$ Hempel-disconfirms $H$ just in case $E$ confirms $\neg H$.

Despite its initial attractiveness, there are a number of disquieting features of Hempel’s attempt to explicate the qualitative concept of confirmation. The discussion of these features can be grouped under two queries. First, is Hempel’s definition too stringent in some respects? Second, is it too liberal in other respects? To motivate the first worry consider

$$H: (x) \text{Rxy.}$$

(The expression ‘‘$\text{Rxy}$’’ means ‘‘$x$ bears relation $R$ to $y$.’’) $H$ is Hempel-confirmed by

$$E: \text{Ra}.\text{Rab}.\text{Rbb}.\text{Rba.}$$

But it is not confirmed by

$$E': \text{Ra}.\text{Rab}.\text{Rbb}$$

even though intuitively the latter evidence does support $H$. Or consider the compound hypothesis

$$(x) (\exists y) \text{Rxy.}(x) \sim \text{Rxx}.(x) (y) (z) [(\text{Rxy.Ryz}) \supset \text{Rxz}],$$

which is true, for example, if we take the quantifiers to range over the natural numbers and interpret $\text{Rxy}$ to mean that $y$ is greater than $x$. (Thus interpreted, the formula says that for any number whatever, there exists another number that is larger. Although this statement is true for the whole collection of natural numbers, it is obviously false for any finite set of integers.) This hypothesis cannot be Hempel-confirmed by any consistent evidence statement since its development for any finite $I$ is inconsistent. Finally, if $H$ is formulated in the theoretical vocabulary, then, except in very special and uninteresting cases, $H$ cannot be Hempel-confirmed by evidence $E$ stated entirely in the observational vocabulary. Thus, Hempel’s account is silent on how statements drawn from such sciences as theoretical physics—for example, all protons contain three quarks—can be confirmed by evidence gained by observation and experiment. This silence is a high price to pay for overcoming some of the defects of the more vocal H-D account.

This last problem is the starting point for Glymour’s (1980) so-called bootstrapping account of confirmation. Glymour sought to preserve the Hempelian idea that hypotheses are confirmed by deducing instances of them from evidence statements, but in the case of a theoretical hypothesis he allowed that the deduction of instances can proceed with the help of auxiliary hypotheses. Thus, for Glymour the
basic confirmation relation is three-place—$E$ confirms $H$ relative to $H'$—rather than two-place. In the main intended application we are dealing with a scientific theory $T$ which consists of a network of hypotheses, from which $H$ and $H'$ are both drawn. If $T$ is finitely axiomatizable—that is, if $T$ consists of the set of logical consequences of a finite set of hypotheses, $H_1, H_2, \ldots, H_n$—we can say that $T$ is bootstrap-confirmed if for each $H_i$ there is an $H_j$ such that $E$ confirms $H_i$ relative to $H_j$. These ideas are most easily illustrated for the case of hypotheses consisting of simple linear equations.

Consider a theory consisting of the following four hypotheses (and all of their deductive consequences):

\[
\begin{align*}
H_1: & \quad O_1 = X \\
H_2: & \quad O_2 = Y + Z \\
H_3: & \quad O_3 = Y + X \\
H_4: & \quad O_4 = Z
\end{align*}
\]

The $O$s are supposed to be observable quantities while the $X$s and $Y$s are theoretical.

For purposes of a concrete example, suppose that we have samples of four different gases in separate containers. All of the containers have the same volume, and they are at the same pressure and temperature. According to Avogadro’s law, then, each sample contains the same number of molecules. Observable quantities $O_1$–$O_4$ are simply the weights of the four samples:

\[
O_1 = 28 \text{ g}, \quad O_2 = 44 \text{ g}, \quad O_3 = 44 \text{ g}, \quad O_4 = 28 \text{ g}.
\]

Our hypotheses say

$H_1$: The first sample consists solely of molecular nitrogen—$N_2$—molecular weight 28; $X$ is the weight of a mole of $N_2$ (28 g).

$H_2$: The second sample consists of carbon dioxide—$CO_2$—molecular weight 44; $Y$ is the weight of a mole of atomic oxygen $O$ (16 g), $Z$ is the weight of a mole of carbon monoxide CO (28 g).

$H_3$: The third sample consists of nitrous oxide—$N_2O$—molecular weight 44; $Y$ is the weight of a mole of atomic oxygen $O$ (16 g) and $X$ is the weight of a mole of molecular nitrogen (28 g).

$H_4$: The fourth sample consists of carbon monoxide—$CO$—molecular weight 28; $Z$ is the weight of a mole of CO (28 g).

(The integral values for atomic and molecular weights are not precisely correct, but they furnish a good approximation for this example.)

To show how $H_1$ can be bootstrap-confirmed relative to the other three hypotheses, suppose that an experiment has determined values $O_1$, $O_2$, $O_3$, $O_4$, for the observables. From the values for $O_2$ and $O_4$ we can, using $H_2$ and $H_4$, compute values for $Y + Z$ and for $Z$. Together these determine a value for $Y$. Then from the value for $O_3$ we can, using $H_3$, compute a value for $Y + X$. Then from these latter two values we get a value for $X$. Finally, we compare this computed value for $X$ with the
observed value for \( O_1 \). If they are equal, \( H_1 \) is confirmed. Although this simple example may seem a bit contrived, it is in principle similar to the kinds of measurements and reasoning actually used by chemists in the nineteenth century to establish molecular and atomic weights.

If we want the bootstrap procedure to constitute a test in the sense that it carries with it the potential for falsification, then we should also require that there are possible values for the observables such that, using these values and the very same bootstrap calculations that led to a confirmatory instance, values for the theoretical quantities are produced that contradict the hypothesis in question. This requirement is met in the present example.

In Glymour’s original formalization of the bootstrapping idea, macho bootstrapping was allowed; that is, in deducing instances of \( H \), it was allowed that \( H \) itself could be used as an auxiliary assumption. To illustrate, consider again the earlier example of the perfect gas law \( P(\text{ressure}) \times V(\text{olume}) = K \times T(\text{emperature}) \), and suppose \( P, V, T \) to be observable quantities while the gas constant \( K \) is theoretical. We proceed to bootstrap-test this law relative to itself by measuring the observables on two different occasions and then comparing the values \( k_1 \) and \( k_2 \) for \( K \) deduced from the law itself and the two sets of observation values \( p_1, v_1, t_1 \) and \( p_2, v_2, t_2 \). However, macho bootstrapping can lead to unwanted results, and in any case it may be unnecessary since, for instance, in the gas law example it is possible to analyze the logic of the test without using the very hypothesis being tested as an auxiliary assumption in the bootstrap calculation (see Edidin 1983 and van Fraassen 1983). These and other questions about bootstrap testing are currently under discussion in the philosophy journals. (The original account of bootstrapping, Glymour 1980, is open to various counterexamples discussed in Christensen 1983; see also Glymour 1983.)

Let us now return to Hempel’s account of confirmation to ask whether it is too liberal. Two reasons for giving a positive answer are contained in the following paradoxes.

**Paradox of the ravens.** Consider again the hypothesis that all ravens are black: \( (x) (Rx \supset Bx) \). Which of the following evidence statements Hempel-confirm the ravens hypothesis?

\[
E_1; Ra_1.Ba_1 \\
E_2; \sim Ra_2 \\
E_3; Ba_3 \\
E_4; \sim Ra_4. \sim Ba_4 \\
E_5; \sim Ra_5.Ba_5 \\
E_6; Ra_6. \sim Ba_6
\]

The answer is that \( E_1-E_5 \) all confirm the hypothesis. Only the evidence \( E_6 \) that refutes the hypothesis fails to confirm it. The indoor ornithology of some of these Hempel-confirmation relations—the confirmation of the ravens hypothesis, say, by the evidence that an individual is a piece of white chalk—has seemed to many to be too easy to be true.

**Goodman’s paradox.** If anything seems safe in this area it is that the evidence \( Ra.Ba \) that \( a \) is a black raven confirms the ravens hypothesis \( (x) (Rx \supset Bx) \). But on
Hempel’s approach nothing rides on the interpretation of the predicates $Rx$ and $Bx$. Thus, Hempel confirmation would still obtain if we interpreted $Bx$ to mean that $x$ is blite, where “blite” is so defined that an object is blite if it is examined on or before December 31, 2000, and is black or else is examined afterwards and found to be white. Thus, by the special consequence condition, the evidence that $a$ is a black raven confirms the prediction that if $b$ is a raven examined after 2000, it will be white, which is counterintuitive to say the least.

---

**Part II: Hume’s Problem of Induction**

### 2.5 THE PROBLEM OF JUSTIFYING INDUCTION

Puzzles of the sort just mentioned—involving blite ravens and grue emeralds (an object is grue if it is examined on or before December 31, 2000 and is green, or it is examined thereafter and is blue)—were presented in Nelson Goodman (1955) under the rubric of the new riddle of induction. Goodman sought the basis of our apparent willingness to generalize inductively with respect to such predicates as “black,” “white,” “green,” and “blue,” but not with respect to “blite” and “grue.” To mark this distinction he spoke of *projectible predicates* and *unprojectible predicates*, and he supposed that there are predicates of each of these types. The problem is to find grounds for deciding which are which.

There is, however, a difficulty that is both historically and logically prior. In his *Treatise of Human Nature* ([1739–1740] 1978) and his *Enquiry Concerning Human Understanding* (1748) David Hume called into serious question the thesis that we have any logical or rational basis for any inductive generalizations—that is, for considering any predicate to be projectible.

Hume divided all reasoning into two types, reasoning concerning *relations of ideas* and reasoning concerning *matters of fact and existence*. All of the deductive arguments of pure mathematics and logic fall into the first category; they are unproblematic. In modern terminology we say that they are necessarily truth-preserving because they are nonampliative (see Chapter 1, Section 1.5). If the premises of any such argument are true its conclusion must also be true because the conclusion says nothing that was not said, at least implicitly, by the premises.

Not all scientific reasoning belongs to the first category. Whenever we make inferences from observed facts to the unobserved we are clearly reasoning ampliatively—that is, the content of the conclusion goes beyond the content of the premises. When we predict future occurrences, when we retrodict past occurrences, when we make inferences about what is happening elsewhere, and when we establish generalizations that apply to all times and places we are engaged in reasoning concerning matters of fact and existence. In connection with reasoning of the second type Hume directly poses the question: What is the foundation of our inferences from the observed to the unobserved? He readily concludes that such reasoning is based upon *relations of cause and effect*. When we see lightning nearby (cause) we infer that the sound of thunder (effect) will ensue. When we see human footprints in the sand
(effect) we infer that a person recently walked there (cause). When we hear a knock
and a familiar voice saying "Anybody home?" (effect) we infer the presence of a
friend (cause) outside the door.

The next question arises automatically: How can we establish knowledge of the
cause-effect relations to which we appeal in making inferences from the observed to
the unobserved? Hume canvasses several possibilities. Do we have a priori knowl-
edge of causal relations? Can we look at an effect and deduce the nature of the cause?
He answers emphatically in the negative. For a person who has had no experience of
diamonds or of ice—which are very similar in appearance—there is no way to infer
that intense heat and pressure can produce the former but would destroy the latter.
Observing the effect, we have no way to deduce the cause. Likewise, for a person
who has had no experience of fire or snow, there is no way to infer that the former
will feel hot while the latter will feel cold. Observing the cause, we have no way to
deduce the effect. All of our knowledge of causal relations must, Hume argues, be
based upon experience.

When one event causes another event, we might suppose that three factors are
present—namely, the cause, the effect, and the causal connection between them.
However, in scrutinizing such situations Hume fails to find the third item—the causal
connection itself. Suppose that one billiard ball lies at rest on a table while another
moves rapidly toward it. They collide. The ball that was at rest begins to move. What
we observe, Hume notes, is the initial motion of the one ball and its collision with the
other. We observe the subsequent motion of the other. This is, he says, as perfect a
case of cause and effect as we will ever see. We notice three things about the
situation. The first is temporal priority; the cause comes before the effect. The second
is spatiotemporal proximity; the cause and effect are close together in space and time.
The third is constant conjunction; if we repeat the experiment many times we find that
the result is just the same as the first time. The ball that was at rest always moves away
after the collision.

Our great familiarity with situations similar to the case of the billiard balls may
give us the impression that "it stands to reason" that the moving ball will produce
motion in the one at rest, but Hume is careful to point out that a priori reasoning cannot
support any such conclusion. We can, without contradiction, imagine many possibil-
ities: When they collide the two balls might vanish in a puff of smoke; the moving ball
might jump right over the one at rest; or the ball that is initially at rest might remain
fixed while the moving ball returns in the direction from which it came. Moreover, no
matter how closely we examine the situation, the thing we cannot see, Hume maintains,
is the causal connection itself—the "secret power" by which the cause brings about
the effect. If we observe two events in spatiotemporal proximity, one of which follows
right after the other, just once, we cannot tell whether it is a mere coincidence or a
genuine causal connection. Hans Reichenbach reported an incident that occurred in a
theater in California as he was watching a movie. Just as a large explosion was depicted
on the screen the theatre began to tremble. An individual's first instinct was to link them
as cause and effect, but, in fact, by sheer coincidence, a minor earthquake occurred at
precisely that moment. Returning to Hume's billiard ball example, on the first obser-
vation of such a collision we would not know whether the motion of the ball originally
at rest occurred by coincidence or as a result of the collision with the moving ball. It
is only after repeated observations of such events that we are warranted in concluding that a genuine causal relation exists. This fact shows that the causal connection itself is not an observable feature of the situation. If it were an observable feature we would not need to observe repetitions of the sequence of events, for we would be able to observe it in the first instance.  

What, then, is the basis for our judgements about causal relations? Hume answers that it is a matter of custom or habit. We observe, on one occasion, an event of type C and observe that it is followed by an event of type E. On another occasion we observe a similar event of type C followed by a similar event of type E. This happens repeatedly. Thereafter, when we notice an event of type C we expect that it will be followed by an event of type E. This is merely a fact about human psychology; we form a habit, we become conditioned to expect E whenever C occurs. There is no logical necessity in all of this.

Indeed, Hume uncovered a logical circle. We began by asking for the basis on which inferences from the observed to the unobserved are founded. The answer was that all such reasoning is based upon relations of cause and effect. We then asked how we can establish knowledge of cause-effect relations. The answer was that we assume—or psychologically anticipate—that future cases of events of type C will be followed by events of type E, just as in past cases events of type C were followed by events of type E. In other words, we assume that nature is uniform—that the future will be like the past—that regularities that have been observed to hold up to now will continue to hold in the future.

But what reason do we have for supposing that nature is uniform? If you say that nature's uniformity has been established on the basis of past observations, then to suppose that it will continue to be uniform is simply to suppose that the future will be like the past. That is flagrantly circular reasoning. If you say that science proceeds on the presumption that nature is uniform, and that science has been extremely successful in predicting future occurrences, Hume's retort is the same. To assume that future scientific endeavors will succeed because science has a record of past success is, again, to suppose that the future will be like the past. Furthermore, Hume points out, it is entirely possible that nature will not be uniform in the future—that the future need not be like the past—for we can consistently imagine all sorts of other possibilities. There is no contradiction in supposing that, at some future time, a substance resembling snow should fall from the heavens, but that it would feel like fire and taste like salt. There is no contradiction in supposing that the sun will not rise tomorrow morning. One can consistently imagine that a lead ball, released from the hand, would rise rather than fall. We do not expect such outlandish occurrences, but that is a result of our psychological makeup. It is not a matter of logic.

The English philosopher John Locke had claimed that in one sort of situation we do observe the actual power of one event to bring about another, namely, in cases in which a person has a volition or desire to perform some act and does so as a result. We might, for example, wish to raise our arms, and then do so. According to Locke we would be aware of our power to produce motion in a part of our body. Hume gave careful consideration to Locke's claim, and argued that it is incorrect. He points out, among other things, that there is a complex relationship of which we are not directly aware—invoking transmission of impulses along nerves and the contractions of various muscles—between the volition originating in the brain and the actual motion of the arm. Hume's critique effectively cut the ground from under Locke's claim.
What applies to Hume's commonsense examples applies equally to scientific laws, no matter how sophisticated they may be. We have never observed an exception to the law of conservation of angular momentum; nevertheless, tomorrow it may fail. Within our experience, the half-life of $^{14}\text{C}$ has been 5730 years; tomorrow it could be 10 minutes. We have never found a signal that could be propagated faster than light; tomorrow we may find one. There is no guarantee that the chemistry of the DNA molecule will be the same in the future as it has been up to now. The possibilities are endless.

We should be clear about the depth and scope of Hume's arguments. Hume is not merely saying that we cannot be certain about the results of science—about scientific predictions, for example. That point had been recognized by the ancient skeptics many centuries before Hume's time. Hume's point is that we have no logical basis for placing any confidence in any scientific prediction. From this moment on, for all we can know every scientific prediction might fail. We cannot say even that scientific predictions are probable (the concept of probability will be examined in detail in Part III). We have no rational basis for placing more confidence in the predictions of science than in the predictions of fortune tellers or in wild guesses. The basis of our inferences from the observed to the unobserved is, to use Hume's terms, custom and habit.

2.6 ANSWERS TO HUME

Hume's critique of inductive reasoning struck at the very foundations of empirical science. It can be formulated as a dilemma. Science involves ampliative inference in an essential way. If we ask for the warrant or justification of any sort of ampliative inference, two responses seem possible. We could, on the one hand, attempt to offer a deductive argument to show that the conclusion follows from the premises—that the conclusion will be true if the premises are—but if any such argument could be given it would transform induction into deduction, and we would be left without any sort of ampliative inference. We could, on the other hand, try to offer an inductive justification, but any such justification would be circular—it would involve the use of induction itself to justify induction. The result is that, on either alternative, it is impossible to provide a suitable justification for the kinds of reasoning indispensable to science—and to common sense as well. It is this situation that led Broad (1926) to remark that induction is the glory of science and the scandal of philosophy.

It goes almost without saying that philosophers have adopted a variety of strategies to deal with Hume's dilemma. We consider some of the more appealing and/or influential ones (a number of these approaches are discussed in detail in Salmon 1967, Chapter 2).

1. The success of science. In spite of Hume's clear arguments concerning the circularity of justifying induction by using induction, it is difficult to escape the feeling that the most basic reason for relying on the methods of science is the remarkable success they have achieved in enabling us to explain natural phenomena and predict future events. What better basis could there be for judging the worth of a method than its track record up to now? And certainly no method of astrology,
crystal gazing, divination, entrail reading, fortune telling, guessing, palmistry, or prophecy can begin to match the success of science. It would seem absurd to give up a highly successful method in exchange for one whose record is patently inferior.

Suppose, however, that a scientist—either an actual practitioner of science or anyone else who believes in the scientific method—is challenged by a crystal gazer. The scientist disparages crystal gazing as a method for predicting the future on the ground that it has not in the past been a very successful method, while the scientific method has, on the whole, worked well. The crystal gazer might correctly accuse the scientist of using the scientific method to justify the scientific method. The method of science is based, after all, on projecting past regularities into the future. To predict that the scientific method will continue to be successful in the future because it has been successful in the past is flagrantly circular. "If you are going to use your method to judge your method," the crystal gazer might remark, "then I have every right to use my method to judge my method." After looking into the crystal ball, the crystal gazer announces that the method of crystal gazing (in spite of its past lack of success) is about to become a very reliable method of prediction. "Furthermore," the crystal gazer might add, "since you used your method to cast aspersions on my method, I will use my method to judge yours: I see in my crystal ball that the scientific method is going to have a run of really bad luck in its forthcoming use as a method of prediction."

As Hume's argument regarding circularity had clearly shown, it is difficult to see anything wrong with the logic of the crystal gazer. 6

2. Ordinary language dissolution. Perhaps the most widely adopted approach to Hume's problem of induction is the attempt to dissolve it—rather than trying to solve it—by showing that it was not a genuine problem in the first place. One way to state the argument is this. If we ask what it means to be reasonable, the obvious answer is that it means to fashion our beliefs in terms of the available evidence. But what is the meaning of the concept of evidence? There are two forms of evidence, corresponding to two kinds of arguments—namely, deductive and inductive. To say that a proposition is supported by deductive evidence means that it can be deduced from propositions that we are willing to accept. If, for example, we accept the postulates of Euclidean geometry, then we are permitted to accept the Pythagorean theorem, inasmuch as it follows deductively from those postulates. Similarly—so the argument goes—if we accept the vast body of empirical evidence that is available, we should be prepared to accept the law of conservation of angular momentum (recall the figure skater in Chapter 1). This evidence inductively supports the claim that angular momentum is always conserved, and hence, that it will continue to be conserved tomorrow, next week, next month, next year and so on. To ask—in the spirit of Hume—whether we are justified in believing that angular momentum will be conserved tomorrow is to ask whether it is reasonable to base our beliefs on the available evidence, which, in this case, is inductive evidence. But basing our beliefs on evidence is just what it means to be rational. To ask whether we should believe on the basis of inductive evidence is tantamount to asking whether it is reasonable to be

6 Max Black and R. B. Braithwaite both argued that inductive justifications of induction could escape circularity. The arguments of Black are criticized in detail in Salmon (1967, 12–17); Braithwaite's arguments are open to analogous criticism.
reasonable (two classic statements of this view are given by Ayer 1956, 71–75 and Strawson 1952, Chapter 9). The problem vanishes when we achieve a clear understanding of such terms as “evidence” and “rationality.”

The foregoing argument is often reinforced by another consideration. Suppose someone continues to demand a justification for the fundamental principles of induction, for example, that past regularities can be projected into the future. The question then becomes, to what principle may we appeal in order to supply any such justification? Since the basic principles of inductive reasoning, like those of deductive reasoning, are ultimate, it is impossible to find anything more basic in terms of which to formulate a justification. Thus, the demand for justification of our most basic principles is misplaced, for such principles define the concept of justification itself.

In spite of its popular appeal among philosophers, this attempt to dispose of Hume’s problem of justification of induction is open to serious objection. It can be formulated in terms of a useful distinction, drawn by Herbert Feigl (1950), between two kinds of justification—validation and vindication. A validation of a principle consists in a derivation of that principle from other, more basic, principles that we accept. For example, we often try to validate moral and/or legal principles. Some people argue that abortion is wrong, and should be outlawed, because it is wrong to take human life (except in certain extreme circumstances) and human life begins at the time of conception. Others (in America) argue, by appealing to certain rights they take to be guaranteed by the Constitution of the United States, that abortion should be permitted. What counts as a validation for any individual obviously depends upon the fundamental principles that person adopts.

Validation also occurs in mathematics and logic. The derivation of the Pythagorean theorem from the postulates of Euclidean geometry constitutes a good mathematical example. In logic, the inference rule modus tollens

\[
\begin{align*}
(6) \quad & p \supset q \\
\quad & \sim q \\
\quad & \sim p
\end{align*}
\]

can be validated by appealing to modus ponens

\[
\begin{align*}
(7) \quad & p \supset q \\
\quad & p \\
\quad & q
\end{align*}
\]

and contraposition

\[ (p \supset q) \equiv (\sim q \supset \sim p). \]

A less trivial example in deductive logic is the validation of the rule of conditional proof by means of the deduction theorem. The deduction theorem shows that any conclusion that can be established by means of conditional proof can be derived using standard basic deductive rules without appeal to conditional proof. Conditional proof greatly simplifies many derivations, but it does not allow the derivation of any conclusion that cannot be derived without it.
To vindicate a rule or a procedure involves showing that the rule or procedure in question serves some purpose for which it is designed. We vindicate the basic rules of deductive logic by showing that they are truth-preserving—that it is impossible to derive false conclusions from true premises when these rules are followed. This is a vindication because we want to be guaranteed that by using deductive rules we will never introduce a false conclusion by deduction from true premises.

Where induction is concerned we know that—because it is ampliative—truth preservation cannot be guaranteed; we sometimes get false conclusions from true premises. We would like to be able to guarantee that we will usually get true conclusions from true premises, but Hume's arguments show that this goal cannot be guaranteed either. As we will see, Reichenbach tried to give a different kind of vindication, but that is not the issue right now. The point is that, of the two kinds of justification, only one—validation—requires appeal to more basic principles; vindication does not. Vindications appeal to purposes and goals. When it is noted—as in the foregoing argument—that there is no principle more basic in terms of which induction can be justified, that shows that induction cannot be validated; it does not follow that induction cannot be vindicated.

If we keep clearly in mind the distinction between validation and vindication, we can see that the ordinary language dissolution fails. When we pose the question, "Is it reasonable to be reasonable?" it is easy to be fooled by an equivocation. Two senses of the word "reasonable" correspond to the two senses of "justification." One sense of "reasonable" ("reasonable_1") corresponds to vindication; in this sense, to ask whether something is reasonable is to ask whether it is a good means for achieving some desired goal. Where induction is concerned, that goal may be described roughly as getting true conclusions or making correct predictions as often as possible. The other sense of "reasonable" ("reasonable_2") corresponds to validation. In this sense, being reasonable includes adopting the generally accepted basic principles of inductive inference. If we now ask, "Is it reasonable, to be reasonable_2?" the question is far from trivial; it now means, "Does it serve our goal of predicting correctly as often as possible (reasonable_1) to use the accepted rules of inductive inference (reasonable_2)?" This is just another way of phrasing the fundamental question Hume raised concerning the justifiability of induction; the basic problem has not been dissolved, but only reformulated.

3. Inductive intuition. When Goodman posed the new riddle of induction, he made some sweeping claims about the nature of justification of logical principles. These claims applied both to deduction and to induction. In both cases, he said, we must confront the basic principles we hold dear with the kinds of arguments we are prepared to accept as valid or logically correct. (The term "valid" is often defined so that it characterizes logically correct deductive arguments only; if it is so construed, we need another term, such as "logically correct" to characterize inductive arguments that conform to appropriate logical principles.) When an argument that we want to retain conflicts with a principle we do not want to relinquish, some adjustment must be made. Speaking of deduction, Goodman says:

The point is that rules and particular inferences alike are justified by being brought into agreement with each other. A rule is amended if it yields an inference we are unwilling to
accept; an inference is rejected if it violates a rule we are unwilling to amend. The process of justification is the delicate one of making mutual adjustments between rules and accepted inferences; and in the agreement achieved lies the only justification needed for either. (1955, 67, italics in the original)

He continues:

All this applies equally well to induction. An inductive inference, too, is justified by conformity to general rules, and a general rule by conformity to accepted inductive inferences. Predictions are justified if they conform to valid canons of induction; and the canons are valid if they accurately codify accepted inductive practice. (Ibid.)

Rudolf Carnap, whose theory of probability will be examined in item 6 in Section 2.8, seems to have had a similar point in mind when he said that the basic justification for the axioms of inductive logic rests on our inductive intuitions (Schilpp 1963, 978).

Goodman’s claim about deductive logic is difficult to defend. We reject, as fallacious, the form of affirming the consequent

\[
\begin{align*}
(8) & \quad p \supset q \\
\qquad & \quad q \\
\qquad & \quad \underline{p}
\end{align*}
\]

because it is easy to provide a general proof that it is not necessarily truth-preserving. The rejection is not the result of a delicate adjustment between particular arguments and general rules; it is based upon a demonstration that the form lacks one of the main features demanded of deductive rules. Other argument forms, such as modus ponens and modus tollens, are accepted because we can demonstrate generally that they are necessarily truth-preserving. (Going beyond truth-functional logic, there are general proofs of the consistency and completeness of standard first-order logic.)

The situation in inductive logic is complicated. As we will see when we study the various proposed interpretations of the concept of probability, there is an enormous plethora of possible rules of inference. We can illustrate by looking at three simple rules as applied to a highly artificial example. Suppose we have a large urn containing an extremely large number of marbles, all of which are known beforehand to be either red, yellow, or blue. We do not know beforehand what proportion of the marbles is constituted by each color; in fact, we try to learn the color constitution of the population of marbles in the urn by removing samples and observing the colors of the marbles in the samples.

Suppose, now, that the contents of the urn are thoroughly mixed, and that we draw out a sample containing \(n\) marbles, of which \(m\) are red. Consider three possible rules for inferring (or estimating) the percentage of the marbles in the urn that are red:

*Induction by enumeration:* if \(m/n\) of the marbles in the sample are red, infer that approximately \(m/n\) of all marbles in the urn are red.
A priori rule: regardless of the makeup of the observed sample, infer that approximately \( \frac{1}{3} \) of all marbles in the urn are red. (The fraction \( \frac{1}{3} \) is chosen because three colors occur in the total population of marbles in the urn.)

Counterinductive rule: if \( m/n \) of the marbles in the sample are red, infer that approximately \((n - m)/n\) of the marbles in the urn are red.

Certain characteristics of these rules can be established by general arguments. The counterinductive rule is so called because it uses observed evidence in a negative way. If we observe the proportion of red marbles in a sample, this rule instructs us to project that the proportion of red in the whole population is approximately equal to the proportion that are not red in the sample. Use of this rule would rapidly land us in an outright contradiction. Suppose, for the sake of simplicity, that our observed sample contains \( \frac{1}{3} \) red, \( \frac{1}{3} \) yellow, and \( \frac{1}{3} \) blue. Using the counterinductive rule for each of the colors would yield the conclusion that \( \frac{2}{3} \) of the marbles in the urn are red, and \( \frac{2}{3} \) of the marbles in the urn are yellow, and \( \frac{2}{3} \) of the marbles in the urn are blue. This is logically impossible; clearly, the counterinductive rule is unsatisfactory.

Suppose we use the a priori rule. Then, even if 98 percent of our observed sample were red, 1 percent yellow, and 1 percent blue, the rule would direct us to ignore that empirical evidence and infer that only about \( \frac{1}{3} \) of the marbles in the urn are red. Because the a priori rule makes observation irrelevant to prediction, it, too, should be rejected.

The rule of induction by enumeration does not have either of the foregoing defects, and it has some virtues. One virtue is that if it is used persistently on larger and larger samples, it must eventually yield inferences that are approximately correct. If we are unlucky, and begin by drawing unrepresentative samples, it will take a long time to start giving accurate results; if we are lucky and draw mainly representative samples, the accurate results will come much sooner. (Some philosophers have derived considerable comfort from the fact that the vast majority of samples that could be drawn are very nearly representative. See Williams 1947.)

Obviously many—indeed, infinitely many—possible rules exist for making inductive inferences. The problem of deciding which of these rules to use is complicated and difficult. We have seen, nevertheless, that general considerations can be brought to bear on the choice. It is not just a matter of consulting our intuitions regarding the acceptability or nonacceptability of particular inductive inferences. This is not to deny, however, that intuitive consideration of particular inferences has a great deal of heuristic value.

Although we have been skeptical about Goodman's success in dismissing the old riddle of induction, we must remark on the importance of his new riddle. First, Hume never explicitly took account of the fact that some forms of constant conjunction do not give rise to habits of expectation. Such Goodmanian predicates as "blite" and "grue" call attention vividly to this point. Second, Goodman's examples provide another way of showing that there can be no noncircular justification of induction by means of a uniformity principle. There are many uniformities, and the question of which ones will extend into the future is the problem of induction all over again.

4. Deductivism. One influential philosopher, Sir Karl Popper, has attacked Hume's problem by denying that science involves any use of induction. He takes
Hume to have proved decisively that induction cannot be justified, and he concludes that science—if it is to be a rational enterprise—must do without it. The only logic of science, he maintains, is deduction.

Popper characterizes the method of science as trial and error, as conjecture and refutation. The scientist formulates bold explanatory hypotheses, and then subjects them to severe testing. This test procedure is very much like hypothetico-deductive testing, but there is an absolutely crucial difference. According to the H-D theory, when the observational prediction turns out to be true, that confirms the hypothesis to some degree. Popper denies that there is any such thing as confirmation. If, however, the observational prediction turns out to be false, modus tollens can be used to conclude deductively that some premise is false. If we are confident of the initial conditions and auxiliary hypotheses, then we reject the hypothesis. The hypothesis was a conjecture; the test provided a refutation. Hypotheses that are refuted must be rejected.

If a bold hypothesis is subjected to severe testing and is not refuted, it is said to be corroborated. Popper emphatically denies that corroboration is any brand of confirmation. H-D theorists regard confirmation as a process that increases to some degree the probability of the hypothesis and, by implication, the probability that the hypothesis will yield correct predictions. Corroboration, in contrast, says nothing whatever about the future predictive success of the hypothesis; it is, instead, a report exclusively on the past performance of the hypothesis. The corroboration-rating is a statement of the past success of the hypothesis as an explanatory theory. The corroboration report is not contaminated with any inductive elements.

Even if we were to grant Popper’s dubious claim that theoretical science is concerned only with explanation, and not with prediction, it would be necessary to recognize that we use scientific knowledge in making practical decisions. If we wish to put an artificial satellite into an orbit around the earth, we use Newtonian mechanics to compute the trajectory, and we confidently expect the satellite to perform as predicted. An inductivist would claim that we base such expectations on the fact that, within certain well-defined limits, Newtonian mechanics is a well-confirmed theory. Popper maintains that, for purposes of practical prediction, using well-corraborated theories is advisable, for nothing could be more rational.

The crucial question is, however, whether anything could be less rational than to use the corroboration-rating of a theory as a basis for choosing it for predictive purposes. Recalling that Popper has emphatically stated that the corroboration-rating refers only to past performance, and not to future performance, the corroboration-rating would seem to be totally irrelevant to the predictive virtues of the theory. The use of highly corroborated theories for prediction has no greater claim to rationality than do the predictions of fortune-tellers or sheer blind guessing. The price for banishing all inductive elements from science is to render science useless for prediction and practical decision making (see Salmon 1981).

5. Pragmatic vindication. Reichenbach fully accepted Hume’s conclusion about the impossibility of proving that nature is uniform. He agreed that we have no way of knowing whether past uniformities will extend into the future. He recognized that, for all we can know, every inductive inference we make in the future may lead
to a false prediction. Nevertheless, he attempted to construct a practical decision-theoretic justification for the use of induction.

Given our inability to know whether nature is uniform, we can consider what happens in either case. Hume showed convincingly that, if nature is uniform, inductive reasoning will work very well, whereas, if nature is not uniform, inductive reasoning will fail. This much is pretty easy to see. Reichenbach suggested, however, that we should consider other options besides the use of induction for purposes of trying to predict the future. Suppose we try consulting a crystal gaze to get our predictions. We cannot say a priori that we will get correct predictions, even if nature turns out to be uniform, but we cannot say a priori that we won’t. We just don’t know. Let us set up a chart:

<table>
<thead>
<tr>
<th></th>
<th>Nature is uniform</th>
<th>Nature is not uniform</th>
</tr>
</thead>
<tbody>
<tr>
<td>We use induction</td>
<td>Success</td>
<td>Failure</td>
</tr>
<tr>
<td>We don’t use induction</td>
<td>Success or Failure</td>
<td>Failure</td>
</tr>
</tbody>
</table>

The crucial entry is in the lower right-hand box. What if nature is not uniform and we do not use induction? One possibility is simply not to make any predictions at all; whether nature is uniform or not, that obviously does not result in successful predictions. Another possibility is that we adopt a noninductive method such as crystal gazing. Any method—including wild guessing—may yield a true prediction once in a while by chance, whether nature is uniform or not. But suppose that crystal gazing were to work consistently. Then, that would be an important uniformity, and it could be established inductively—that is, on the basis of the observed record of the crystal gazer in making successful predictions we could infer inductively that crystal gazing will be successful in making correct predictions in the future. Thus, if crystal gazing can produce consistent successful predictions so can the use of induction. What has just been said about crystal gazing obviously applies to any noninductive method. Reichenbach therefore concluded that if any method will succeed consistently, then induction will succeed consistently. The same conclusion can be reformulated (by contraposition) as follows: If induction does not work, then no other method will work. We therefore have everything to gain and nothing to lose—so far as predicting the future is concerned—by adopting the inductive method. No other method can make an analogous claim. Reichenbach’s argument is an attempt at vindication of induction. He is trying to show that—even acknowledging Hume’s skeptical arguments—induction is better suited to the goal of predicting the future than any other methods that might be adopted.

Although Reichenbach’s pragmatic justification may seem promising at first glance, it does face serious difficulties on closer inspection. The greatest problem with the foregoing formulation is that it suffers from severe vagueness. What do we mean by speaking of the uniformity of nature? Nature is not completely uniform; things do change. At the same time—up to the present at any rate—nature has exhibited certain kinds of uniformity. What degree of uniformity do we need in order for the argument to succeed? We should be much more precise on this point. Likewise, when we spoke about noninductive methods we did not carefully survey all of
the available options. When the argument is tightened sufficiently, it turns out, it does
not vindicate just one rule of inductive inference; instead, it equally justifies an
infinite class of rules. Serious efforts—up to this time—to find a satisfactory basis for
selecting a unique rule have been unsuccessful, (the technical details are discussed in
Salmon 1967, Chapter 6).

Where do things stand now—250 years after the publication of Hume’s Treatise
of Human Nature—with respect to the problem we have inherited from him? Al-
though many ingenious attempts have been made to solve or dissolve it there is still
no consensus. It still stands as an item of “unfinished business” for philosophy of
science (see Salmon 1978a). The problem may, perhaps, best be summarized by a
passage from Hume himself:

Let the course of things be allowed hitherto ever so regular, that alone, without some new
argument or inference, proves not that for the future it will continue so. In vain do you
pretend to have learned the nature of bodies from your past experience. Their secret nature,
and consequently all their effects and influence, may change without any change in their
sensible qualities. This happens sometimes, and with regard to some objects: Why may it not
happen always, and with regard to all objects? What logic, what process or argument secures
you against this supposition? My practice, you say, refutes my doubts. But you mistake the
purport of my question. As an agent, I am quite satisfied in the point; but as a philosopher
... I want to learn the foundation of this inference. (1748, Section 4)

As Hume makes abundantly clear, however, life—and science—go on in spite of
these troubling philosophical doubts.

---

Part III: Probability

2.7 THE MATHEMATICAL THEORY OF PROBABILITY

Our discussion up to this point has been carried on without the aid of a powerful
tool—the calculus of probability. The time has come to invoke it. The defects of the
qualitative approaches to confirmation discussed in Sections 2.3 and 2.4 suggest that
an adequate account of the confirmation of scientific statements must resort to quan-
titative or probabilistic methods. In support of this suggestion, recall that we have
already come across the concept of probability in the discussion of the qualitative
approaches. In our discussion of the H-D method, for instance, we encountered the
concept of probability in at least two ways. First, noting that a positive result of an
H-D test does not conclusively establish a hypothesis, we remarked that it might
render the hypothesis a little more probable than it was before the test. Second, in
dealing with the problem of statistical hypotheses, we saw that only probabilistic
observational predictions can be derived from such test hypotheses. In order to pursue
our investigation of the issues that have been raised we must take a closer look at the
concept or concepts of probability.
The modern theory of probability had its origins in the seventeenth century. Legend has it that a famous gentleman, the Chevalier de Méré, posed some questions about games of chance to the philosopher-mathematician Blaise Pascal. Pascal communicated the problems to the mathematician Pierre de Fermat, and that was how it all began. Be that as it may, the serious study of mathematical probability theory began around 1660, and Pascal and Fermat, along with Christian Huygens, played crucial roles in that development, (for an historical account see Hacking 1975 and Stigler 1986).

In order to introduce the theory of probability, we take probability to be a relationship between events of two different types—for example, between tossing a standard die and getting a six, or drawing from a standard bridge deck and getting a king. We designate probabilities by means of the following notation:

$$Pr(B \mid A)$$ is the probability of a result of the type B given an event of the type A.

If A is a toss of a standard die and B is getting a three, then “$$Pr(B \mid A)$$” stands for the probability of getting a three if you toss a standard die. As the theory of probability is seen today, all of the elementary rules of probability can be derived from a few simple axioms. The meanings of these axioms and rules can be made intuitively obvious by citing examples from games of chance that use such devices as cards and dice. After some elementary features of the mathematical calculus of probability have been introduced in this section, we look in the following section at a variety of interpretations of probability that have been proposed.

**Axioms (Basic Rules)**

Axiom (rule) 1: Every probability is a unique real number between zero and one inclusive; that is,

$$0 \leq Pr(B \mid A) \leq 1.$$  

Axiom (rule) 2: If A logically entails B, then $$Pr(B \mid A) = 1.$$  

Definition: Events of types B and C are mutually exclusive if it is impossible for both B and C to happen on any given occasion. Thus, for example, on any draw from a standard deck, drawing a heart and drawing a spade are mutually exclusive, for no card is both a heart and a spade.

Axiom (rule) 3: If B and C are mutually exclusive, then

$$Pr(B \lor C \mid A) = Pr(B \mid A) + Pr(C \mid A).$$

This axiom is also known as the special addition rule.

Example: The probability of drawing a heart or a spade equals the probability of drawing a heart plus the probability of drawing a spade.

Axiom (rule) 4: The probability of a joint occurrence—that is, of a conjunction of B and C—is equal to the probability of the first multiplied by the probability of the second given that the first has occurred:

$$Pr(B.C \mid A) = Pr(B \mid A) \times Pr(C \mid A.B).$$
This axiom is also known as the **general multiplication rule**.

Example: If you make two draws *without replacement* from a standard deck, what is the probability of getting two aces? The probability of getting an ace on the first draw is 4/52; the probability of getting an ace on the second draw *if you have already drawn an ace on the first draw* is 3/51, because there are only 51 cards left in the deck and only 3 of them are aces. Thus, the probability of getting two aces is

\[ \frac{4}{52} \times \frac{3}{51} = \frac{12}{2652} = \frac{1}{221}. \]

**Some Derived Rules**

From the four axioms (basic rules) just stated, several other rules are easy to derive that are extremely useful in calculating probabilities. First, we need a definition:

**Definition:** The events \( B \) and \( C \) are **independent** if and only if

\[ \Pr(C|A.B) = \Pr(C|A). \]

When the events \( B \) and \( C \) are independent of one another, the multiplication rule (axiom 4) takes on a very simple form:

**Rule 5:** If \( B \) and \( C \) are independent, given \( A \), then

\[ \Pr(B.C|A) = \Pr(B|A) \times \Pr(C|A). \]

This rule is known as the **special multiplication rule**. (Proofs, sketches of proofs, and other technical items will be placed in boxes. They can be omitted on first reading.)

**Proof of Rule 5:** Substitute \( \Pr(C|A) \) for \( \Pr(B.C|A) \) in Axiom 4.

Example: What is the probability of getting double 6 ("boxcars") when a standard pair of dice is thrown? Since the outcomes on the two dice are independent, and the probability of 6 on each die is 1/6, the probability of double 6 is

\[ \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}. \]

Example: What is the probability of drawing two spades on two consecutive draws when the drawing is done *with replacement*? The probability of getting a spade on the first draw is 13/52 = 1/4. After the first card is drawn, whether it is a spade or not, it is put back in the deck and the deck is reshuffled. Then the second card is drawn. Because of the replacement, the outcome of the second draw is independent of the outcome of the first draw. Therefore, the probability of getting a spade on the second draw is just the same as it was on the first draw. Thus, the probability of getting two spades on two consecutive draws is

\[ \frac{1}{4} \times \frac{1}{4} = \frac{1}{16}. \]
NOTE CAREFULLY. If the drawing is done without replacement, the special multiplication rule cannot be used because the outcomes are not independent. In that case Rule 4 must be used.

Rule 6: \( Pr(\sim B|A) = 1 - Pr(B|A) \).

This simple rule is known as the negation rule. It is very useful.

Example: Suppose you would like to know the probability of getting at least one 6 if you toss a standard die three times.\(^8\) That means you want to know the probability of getting a 6 on the first toss or on the second toss or on the third toss, where this is an inclusive or. Thus, the outcomes are not mutually exclusive, so you cannot use the special addition rule (Axiom 3). We can approach this problem via the negation. To fail to get at least one 6 in three tosses means to get non-6 on the first toss and non-6 on the second toss and non-6 on the third toss. Since the probability of 6 is 1/6, the negation rule tells us that the probability of non-6 is 5/6. Because the outcomes on the three tosses are independent, we can use Rule 5 to obtain the probability of non-6 on all three tosses as

\[
\frac{5}{6} \times \frac{5}{6} \times \frac{5}{6} = \frac{125}{216}.
\]

The probability of getting at least one 6, which is the negation of not getting any 6, is therefore

\[
1 - \frac{125}{216} = \frac{91}{216}.
\]

NOTE CAREFULLY: The probability of getting at least one 6 in three tosses is not 1/2. It is equal to 91/216, which is approximately 0.42.

---

Proof of Rule 6: Obviously every \( A \) is either a \( B \) or not a \( B \). Therefore, by Axiom 2,

\[
Pr(B \lor \sim B|A) = 1.
\]

Since \( B \) and \( \sim B \) are mutually exclusive, Axiom 3 yields

\[
Pr(B|A) + Pr(\sim B|A) = 1.
\]

Rule 6 results from subtracting \( Pr(B|A) \) from both sides.

---

Rule 7: \( Pr(B \lor C|A) = Pr(B|A) + Pr(C|A) - Pr(B.C|A) \).

This is the general addition rule. Unlike Rule 3, this rule applies to outcomes \( B \) and \( C \) even if they are not mutually exclusive.

Example: What is the probability of getting a spade or a face card in a draw from a standard deck? These two alternatives are not mutually exclusive, for there are three

---

\(^8\) This example is closely related to one of the problems posed by the Chevalier de Méré. How many toses of a pair of dice, he asked, are required to have at least a fifty-fifty chance of getting at least one double 6? It seems that a common opinion among gamblers at the time was that 24 toses would be sufficient. The Chevalier doubted that answer, and it turned out that he was right. One needs 25 tosses to have at least a fifty-fifty chance.
cards—king, queen, and jack of spades—that are both face cards and spades. Since there are 12 face cards and 13 spades, the probability of a spade or a face card is

\[
\frac{12}{52} + \frac{13}{52} - \frac{3}{52} = \frac{22}{52}
\]

It is easy to see why this rule has the form that it does. If \( B \) and \( C \) are not mutually exclusive, then some outcomes may be both \( B \) and \( C \). Any such items will be counted twice—once when we count the \( Bs \) and again when we count the \( Cs \). (In the foregoing example, the king of spades is counted once as a face card and again as a spade. The same goes for the queen and jack of spades.) Thus, we must subtract the number of items that are both \( B \) and \( C \), in order that they be counted only once.

How to prove Rule 7. First, we note that the class of things that are \( B \) or \( C \) in the inclusive sense consists of those things that are \( B.C \) or \( \sim B.C \) or \( \sim C \), where these latter three classes are mutually exclusive. Thus, Rule 3 can be applied, giving

\[
Pr(B \lor C | A) = Pr(B.C | A) + Pr(\sim B.C | A) + Pr(\sim C | A).
\]

Rule 4 is applied to each of the three terms on the right-hand side, and then Rule 6 is used to get rid of the negations inside of the parentheses. A bit of simple algebra yields Rule 7.

Rule 8: \( Pr(C | A) = Pr(B | A) \times Pr(C | A.B) + Pr(\sim B | A) \times Pr(C | A.\sim B) \).

This is the rule of total probability. It can be illustrated as follows:

Example: Imagine a factory that produces frisbees. The factory contains just two machines, a new machine \( B \) that produces 800 frisbees each day, and an old machine \( \sim B \) that produces 200 frisbees per day. Among the frisbees produced by the new machine, 1% are defective; among the frisbees produced by the old machine, 2% are defective. Let \( A \) stand for the frisbees produced in a given day at that factory. Let \( B \) stand for the frisbees produced by the new machine; \( \sim B \) then stands for those produced by the old machine. Let \( C \) stand for defective frisbees. Then,

- \( Pr(B | A) = \) the probability that a frisbee is produced by machine \( B = 0.8 \)
- \( Pr(\sim B | A) = \) the probability that a frisbee is produced by machine \( \sim B = 0.2 \)
- \( Pr(C | A.B) = \) the probability that a frisbee produced by machine \( B \) is defective \( = 0.01 \)
- \( Pr(C | A.\sim B) = \) the probability that a frisbee produced by machine \( \sim B \) is defective \( = 0.02 \)

Therefore, the probability that a frisbee is defective =

\[
0.8 \times 0.01 + 0.2 \times 0.02 = 0.012
\]

As can be seen from this artificial example, the rule of total probability can be used to calculate the probability of an outcome that can occur in either of two ways,
either by the occurrence of some intermediate event \( B \) or by the nonoccurrence of \( B \). The situation can be shown in a diagram:

\[
\begin{align*}
A & \quad \sim B \\
\text{ } & \quad \sim B \\
\text{ } & \quad \text{ } \quad C
\end{align*}
\]

Proof of Rule 8: Since every \( C \) is either a \( B \) or not a \( B \), the class \( C \) is identical to the class \( B.C \lor \sim B.C \); moreover, since nothing is both a \( B \) and not a \( B \), the classes \( B.C \) and \( \sim B.C \) are mutually exclusive. Hence,

\[
Pr(C|A) = Pr(C,[B \lor \sim B] |A) \\
= Pr([B.C \lor \sim B.C]|A) \\
= Pr(B.C|A) + Pr(\sim B.C|A) \quad \text{by Rule 3} \\
= Pr(B|A) \times Pr(C|A,B) + Pr(\sim B|A) \times Pr(C|A,\sim B) \\
\text{by Rule 4 applied twice}
\]

We now come to the rule of probability that has special application to the problem of confirmation of hypotheses.

Rule 9: \( Pr(B|A,C) = \frac{Pr(B|A) \times Pr(C|A,B)}{Pr(C|A)} \)

\[
= \frac{Pr(B|A) \times Pr(C|A,B)}{Pr(B|A) \times Pr(C|A,B) + Pr(\sim B|A) \times Pr(C|A,\sim B)}
\]

provided that \( Pr(C|A) \neq 0 \). The fact that these two forms are equivalent follows immediately from the rule of total probability (Rule 8), which shows that the denominators of the right-hand sides are equal to one another.

Rule 9 is known as Bayes's rule; it has extremely important applications. For purposes of illustration, however, let us go back to the trivial example of the frisbee factory that was used to illustrate the rule of total probability.

Example: Suppose we have chosen a frisbee at random from the day's production and it turns out to be defective. We did not see which machine produced it. What is the probability—\( Pr(B|A,C) \)—that it was produced by the new machine? Bayes's rule gives the answer:

\[
\frac{0.8 \times 0.01}{0.8 \times 0.01 + 0.2 \times 0.02} = \frac{0.008}{0.012} = \frac{2}{3}
\]

_The really important fact about Bayes's rule is that it tells us a great deal about the confirmation of hypotheses._ The frisbee example illustrates this point. We have a frisbee produced at this factory (\( A \)) that turns out, on inspection, to be defective (\( C \)), and we wonder whether it was produced (caused) by the new machine (\( B \)). In other
Proof of Bayes’s rule: Bayes’s rule has two forms as given above; we show how to prove both. We begin by writing Rule 4 twice; in the second case we interchange B and C.

\[ Pr(B \mid C \mid A) = Pr(B \mid A) \times Pr(C \mid A \mid B) \]
\[ Pr(C \mid B \mid A) = Pr(C \mid A) \times Pr(B \mid A \mid C) \]

Since the class B.C is obviously identical to the class C.B we can equate the right-hand sides of the two equations:

\[ Pr(C \mid A) \times Pr(B \mid A \mid C) = Pr(B \mid A) \times Pr(C \mid A \mid B) \]

Assuming that \( P(C \mid A) \neq 0 \), we divide both sides by that quantity:

\[ Pr(B \mid A \mid C) = \frac{Pr(B \mid A) \times Pr(C \mid A \mid B)}{Pr(C \mid A)} \]

This is the first form. Using Rule 8, the rule of total probability, we replace the denominator, yielding the second form:

\[ Pr(B \mid A \mid C) = \frac{Pr(B \mid A) \times Pr(C \mid A \mid B)}{Pr(B \mid A) \times Pr(C \mid A \mid B) + Pr(\sim B \mid A) \times Pr(C \mid A \mid \sim B)} \]

Words, we are evaluating the hypothesis that the new machine produced this defective frisbee. As we have just seen, the probability is 2/3.

Inasmuch as we are changing our viewpoint from talking about types of objects and events A, B, C, . . . to talking about hypotheses, let us make a small change in notation to help in the transition. Instead of using “A” to stand for the day’s production of frisbees, we shall use “K” to stand for our background knowledge about the situation in that factory. Instead of using “B” to stand for the frisbees produced by the new machine B, we shall use “H” to stand for the hypothesis that a given frisbee was produced by machine B. And instead of using “C” to stand for defective frisbees, we shall use “E” to stand for the evidence that the given frisbee is defective. Now Bayes’s rule reads as follows:

\[ Rule 9: Pr(H \mid K, E) = \frac{Pr(H \mid K) \times Pr(E \mid K, H)}{Pr(E \mid K)} \]
\[ = \frac{Pr(H \mid K) \times Pr(E \mid K, H)}{Pr(H \mid K) \times Pr(E \mid K, H) + Pr(\sim H \mid K) \times Pr(E \mid K, \sim H)} \]

Changing the letters in the formula (always replacing the same old letter for the same new letter) obviously makes no difference to the significance of the rule. If the axioms...
are rewritten making the same changes in variables, Rule 9 would follow from them in exactly the same way. And inasmuch as we are still talking about probabilities—albeit the probabilities of hypotheses instead of the probabilities of events—we still need the same rules.

We can now think of the probability expressions that occur in Bayes’s rule in the following terms:

\[ Pr(H|K) \] is the prior probability of hypothesis \( H \) just on the basis of our background knowledge \( K \) without taking into account the specific new evidence \( E \). (In our example, it is the probability that a given frisbee was produced by machine \( B \).) \( Pr(\sim H|K) \) is the prior probability that our hypothesis \( H \) is false. (In our example, it is the probability that a given frisbee was produced by machine \( \sim B \).) Notice that \( H \) and \( \sim H \) must exhaust all of the possibilities.

By the negation rule (Rule 6), these two prior probabilities must add up to 1; hence, if one of them is known the other can immediately be calculated.

\[ Pr(E|K,H) \] is the probability that evidence \( E \) would obtain given the truth of hypothesis \( H \) in addition to our background knowledge \( K \). (In our example, it is the probability that a particular frisbee is defective, given that it was produced by machine \( B \).) This probability is known as a likelihood.
\[ Pr(E|K,\sim H) \] is the probability that evidence \( E \) would obtain if our hypothesis \( H \) is false. (In our example, it is the probability that a particular frisbee is defective if it was not produced by machine \( B \).) This probability is also a likelihood.

The two likelihoods—in sharp contrast to the prior probabilities—are independent of one another. Given only the value of one of them, it is \textit{impossible} to calculate the value of the other.

\[ Pr(E|K) \] is the probability that our evidence \( E \) would obtain, regardless of whether hypothesis \( H \) is true or false. (In our example, it is the probability that a given frisbee is defective, regardless of which machine produced it.) This probability is often called the \textit{expectedness} of the evidence.\footnote{Expectedness is the opposite of \\textit{surprisingness}. If the expectedness of the evidence is small the evidence is surprising. Since the expectedness occurs in the denominator of the fraction, the smaller the expectedness, the greater the value of the fraction. Surprising evidence confirms hypotheses more than evidence that is to be expected regardless of the hypothesis.}
\[ Pr(H|K,E) \] is the probability of our hypothesis, judged in terms of our background knowledge \( K \) and the specific evidence \( E \). It is known as the \textit{posterior probability}. This is the probability we are trying to ascertain. (In our example, it is the probability that the frisbee was produced by the new machine. Since the posterior probability of \( H \) is different from the prior probability of \( H \), the fact that the frisbee is defective is evidence relevant to that hypothesis.)
Notice that, although the likelihood of a defective product is twice as great for the old machine (0.02) as for the new (0.01), the posterior probability that a defective frisbee was produced by the new machine (2/3) is twice as great as the probability that it was produced by the old one (1/3).

In Section 2.9 we return to the problem of assigning probabilities to hypotheses, which is the main subject of this chapter.

2.8 THE MEANING OF PROBABILITY

In the preceding section we discussed the notion of probability in a formal manner. That is, we introduced a symbol, "\( Pr() \)," to stand for probability, and we laid down some formal rules governing the use of that symbol. We illustrated the rules with concrete examples, to give an intuitive feel for them, but we never tried to say what the word "probability" or the symbol "\( Pr \)" means. That is the task of this section.

As we discuss various suggested meanings of this term, it is important to recall that we laid down certain basic rules (axioms). If a proposed definition of "probability" satisfies the basic rules—and, consequently, the derived rules, since they are deduced from the basic rules—we say that the suggested definition provides an admissible interpretation of the probability concept. If a proposed interpretation violates those rules, we consider it a serious drawback.

1. The classical interpretation. One famous attempt to define the concept of probability was given by the philosopher-scientist Pierre Simon de Laplace ([1814] 1951). It is known as the classical interpretation. According to this definition, the probability of an outcome is the ratio of favorable cases to the number of equally possible cases. Consider a simple example. A standard die (singular of "dice") has six faces numbered 1–6. When it is tossed in the standard way there are six possible outcomes. If we want to know the probability of getting a 6, the answer is 1/6, for only one possible outcome is favorable. The probability of getting an even number is 3/6, for three of the possible outcomes (2, 4, 6) are favorable.

Laplace was fully aware of a fundamental problem with this definition. The definition refers not just to possible outcomes, but to equally possible outcomes. Consider another example. Suppose two standard coins are flipped simultaneously. What is the probability of getting two heads? Someone might say it is 1/3, for there are three possible outcomes, two heads, one head and one tail, or two tails. We see immediately that this answer is incorrect, for these possible outcomes are not equally possible. That is because one head and one tail can occur in two different ways—head on coin #1 and tail on coin #2, or tail on coin #1 and head on coin #2. Hence, we should say that there are four equally possible cases, so the probability of two heads is 1/4.

In order to clarify his definition Laplace needed to say what is meant by "equally possible," and he endeavored to do so by offering the famous principle of indifference. According to this principle, two outcomes are equally possible—we might as well say "equally probable"—if we have no reason to prefer one to the other.
Compare the coin example with the following from modern physics. Suppose you have two helium-4 atoms in a box. Each one has a fifty-fifty chance of being in the left-hand side of the box at any given time. What is the probability of both atoms being in the left-hand side at a particular time? The answer is 1/3. Since the two atoms are in principle indistinguishable—unlike the coins, which are obviously distinguishable—we cannot regard atom #1 in the left-hand side and atom #2 in the right-hand side as a case distinct from atom #1 in the right-hand side and atom #2 in the left-hand side. Indeed, it does not even make sense to talk about atom #1 and atom #2 since we have no way, even in principle, of telling which is which.

Suppose, for example, that we examine a coin very carefully and find that it is perfectly symmetrical. Any reason one might give to suppose it will come up heads can be matched by an equally good reason to suppose it will land tails up. We say that the two sides are equally possible, and we conclude that the probability of heads is 1/2. If, however, we toss the coin a large number of times and find that it lands heads up in about 3/4 of all tosses and tails up in about 1/4 of all tosses, we do have good reason to prefer one outcome to the other, so we would not declare them equally possible. The basic idea behind the principle of indifference is this: when we have no reason to consider one outcome more probable than another, we should not arbitrarily choose one outcome to favor over another. This seems like a sound principle of probabilistic reasoning.

There is, however, a profound difficulty connected with the principle of indifference; its use can lead to outright inconsistency. The problem is that it can be applied in different ways to the same situation, yielding incompatible values for a particular probability. Again, consider an example, namely, the case of Joe, the sloppy bartender. When a customer orders a 3:1 martini (3 parts of gin to 1 part of dry vermouth), Joe may mix anything from a 2:1 to a 4:1 martini, and there is no further information to tell us where in that range the mix may lie. According to the principle of indifference, then, we may say that there is a fifty-fifty chance that the mix will be between 2:1 and 3:1, and an equal chance that it will be between 3:1 and 4:1. Fair enough. But there is another way to look at the same situation. A 2:1 martini contains 1/3 vermouth, and a 4:1 martini contains 1/5 vermouth. Since we have no further information about the proportion of vermouth we can apply the principle of indifference once more. Since 1/3 = 20/60 and 1/5 = 12/60, we can say that there is a fifty-fifty chance that the proportion of vermouth is between 20/60 and 16/60 and an equal chance that it is between 16/60 and 12/60. So far, so good?

Unfortunately, no. We have just contradicted ourselves. A 3:1 martini contains 25 percent vermouth, which is equal to 15/60, not 16/60. The principle of indifference has told us both that there is a fifty-fifty chance that the proportion of vermouth is between 20/60 and 16/60, and also that there is a fifty-fifty chance that it is between
20/60 and 15/60. The situation is shown graphically in Figure 2.2. As the graph shows, the same result occurs for those who prefer their martinis drier; the numbers are, however, not as easy to handle.

We must recall, at this point, our first axiom, which states, in part, that the probability of a given outcome under specified conditions is a \textit{unique} real number. As we have just seen, the classical interpretation of probability does not furnish unique results; we have just found two different probabilities for the same outcome. Thus, it turns out, the classical interpretation is \textit{not} an admissible interpretation of probability.

You might be tempted to think the case of the sloppy bartender is an isolated and inconsequential fictitious example. Nothing could be farther from the truth. This example illustrates a broad range of cases in which the principle of indifference leads to contradiction. The source of the difficulty lies in the fact that we have two quantities—the ratio of gin to vermouth and the proportion of vermouth—that are interdefinable; if you know one you can calculate the other. However, as Figure 2.2 clearly shows, the definitional relation is not linear; the graph is not a straight line. We can state generally: Whenever there is a nonlinear definitional relationship between two quantities, the principle of indifference can lead to a similar contradiction. To convince yourself of this point, work out the details of another example. Suppose there is a square piece of metal inside of a closed box. You cannot see it. But you are told that its area is somewhere between 1 square inch and 4 square inches, but nothing else is known about the area. First apply the principle of indifference to the area of the square, and then apply it to the length of the side which is, of course, directly

![Figure 2.2](image-url)
ascertainable from the area. (For another example, involving a car on a racetrack, see Salmon 1967, 66–67.)

Although the classical interpretation fails to provide a satisfactory basic definition of the probability concept, that does not mean that the idea of the ratio of favorable to equiprobable possible outcomes is useless. The trouble lies with the principle of indifference, and its aim of transforming ignorance of probabilities into values of probabilities. However, in situations in which we have positive knowledge that we are dealing with alternatives that have equal probabilities, the strategy of counting equiprobable favorable cases and forming the ratio of favorable to equiprobable possible cases is often handy for facilitating computations.

2. The frequency interpretation. The frequency interpretation has a venerable history, going all the way back to Aristotle (4th century B.C.), who said that the probable is that which happens often. It was first elaborated with precision and in detail by the English logician John Venn (1866, [1888] 1962). The basic idea is easily illustrated. Consider an ordinary coin that is being flipped in the standard way. As it is flipped repeatedly a sequence of outcomes is generated:

\[
\text{H T H T T H H T T H T T T H H H . . . .10}
\]

We can associate with this sequence of results a sequence of relative frequencies—that is, the proportion of tosses that have resulted in heads up to a given point in the sequence—as follows:

\[
1/1, \ 1/2, \ 2/3, \ 2/4, \ 2/5, \ 2/6, \ 3/7, \ 4/8, \ 4/9, \ 4/10, \ 5/11, \ 5/12, \ 5/13/ \ 5/14, \ 5/15, \ 6/16, \ 6/17, \ 7/18, \ 7/19, \ 7/20, \ 7/21, \ 8/22, \ 9/23, \ 10/24, \ 11/25, \ . . .
\]

The denominator in each fraction represents the number of tosses made up to that point; the numerator represents the number of heads up to that point. We could, of course, continue flipping the coin, recording the results, and tabulating the associated relative frequencies. We are reasonably convinced that this coin is fair and that it was flipped in an unbiased manner. Thus, we believe that the probability of heads is 1/2. If that belief is correct, then, as the number of tosses increases, the relative frequencies will become and remain close to 1/2. The situation is shown graphically in Figure 2.3. There is no particular number of tosses at which the fraction of heads is and remains precisely 1/2; indeed, in an odd number of tosses the ratio cannot possibly equal 1/2. Moreover, if, at some point in the sequence, the relative frequency does equal precisely 1/2, it will necessarily differ from that value on the next flip. Instead of saying that the relative frequency must equal 1/2 in any particular number of throws, we say that it approaches 1/2 in the long run.

Although we know that no coin can ever be flipped an infinite number of times, it is useful, as a mathematical idealization, to think in terms of a potentially infinite sequence of tosses. That is, we imagine that, no matter how many throws have been

---

10 These are the results of 25 flips made in an actual trial by the authors.
made, it is still possible to make more; that is, there is no particular finite number $N$ at which point the sequence of tosses is considered complete. Then we can say that the limit of the sequence of relative frequencies equals the probability; this is the meaning of the statement that the probability of a particular sort of occurrence is, by definition, its long run relative frequency.

What is the meaning of the phrase "limit of the relative frequency"? Let $f_1, f_2, f_3, \ldots$ be the successive terms of the sequence of relative frequencies. In the example above, $f_1 = 1$, $f_2 = 1/2$, $f_3 = 2/3$, and so on. Suppose that $p$ is the limit of the relative frequency. This means that the values of $f_n$ become and remain arbitrarily close to $p$ as $n$ becomes larger and larger. More precisely, let $\delta$ be any small number greater than 0. Then, there exists some finite integer $N$ such that, for any $n > N$, $f_n$ does not differ from $p$ by more than $\delta$.

Many objections have been lodged against the frequency interpretation of probability. One of the least significant is that mentioned above, namely, the finitude of all actual sequences of events, at least within the scope of human experience. The reason this does not carry much weight is the fact that science is full of similar sorts of idealizations. In applying geometry to the physical world we deal with ideal straight lines and perfect circles. In using the infinitesimal calculus we assume that certain quantities—such as electric charge—can vary continuously, when we know that they are actually discrete. Such practices carry no danger provided we are clearly aware of the idealizations we are using. Dealing with infinite sequences is technically easier than dealing with finite sequences having huge numbers of members.

A much more serious problem arises when we ask how we are supposed to ascertain the values of these limiting frequencies. It seems that we observe some limited portion of such a sequence and then extrapolate on the basis of what has been observed. We may not want to judge the probability of heads for a certain coin on the basis of 25 flips, but we might well be willing to do so on the basis of several hundred. Nevertheless, there are several logical problems with this procedure. First, no matter how many flips we have observed, it is always possible for a long run of heads to occur that would raise the relative frequency of heads well above 1/2. Similarly, a long run of future tails could reduce the relative frequency far below 1/2.

Another way to see the same point is this. Suppose that, for each $n$, $m/n$ is the fraction of heads to tosses as of the $n$th toss. Suppose also that $f_n$ does have the
limiting value \( p \). Let \( a \) and \( b \) be any two fixed positive integers where \( a \leq b \). If we add the constant \( a \) to every value of \( m \) and the constant \( b \) to every value of \( n \), the resulting sequence \( (m + a)/(n + b) \) will converge to the very same value \( p \). That means that you could attach any sequence of \( b \) tosses, \( a \) of which are heads, to the beginning of your sequence, without changing the limiting value of the relative frequency. Moreover, you can chop off any finite number \( b \) of members, \( a \) of which are heads, from the beginning of your sequence without changing the limiting frequency \( p \). As \( m \) and \( n \) get very large, the addition or subtraction of fixed numbers \( a \) and \( b \) has less and less effect on the value of the fraction. This seems to mean that the observed relative frequency in any finite sample is irrelevant to the limiting frequency. How, then, are we supposed to find out what these limiting frequencies—probabilities—are?

It would seem that things could not get much worse for the frequency interpretation of probability, but they do. For any sequence, such as our sequence of coin tosses, there is no guarantee that any limit of the relative frequency even exists. It is logically possible that long runs of heads followed by longer runs of tails followed by still longer runs of heads, and so on, might make the relative frequency of heads fluctuate between widely separated extremes throughout the infinite remainder of the sequence. If no limit exists there is no such thing as the probability of a head when this coin is tossed.

In spite of these difficulties, the frequency concept of probability seems to be used widely in the sciences. In Chapter 1, for instance, we mentioned the spontaneous decay of \( ^{14}C \) atoms, commenting that the half-life is 5730 years. That is the rate at which atoms of this type have decayed in the past; we confidently predict that they will continue to do so. The relative frequency of disintegration of \( ^{14}C \) atoms within 5730 years is 1/2. This type of example is of considerable interest to archaeologists, physicists, and geophysicists. In the biological sciences it has been noted, for example, that there is a very stable excess of human male births over human female births, and that is expected to continue. Social scientists note, however, that human females, on average, live longer than human males. This frequency is also extrapolated.

It is easy to prove that the frequency interpretation satisfies the axioms of probability laid down in the preceding section. This interpretation is, therefore, admissible. Its main difficulty lies in the area of ascertainability. How are we to establish values of probabilities of this sort? This question again raises Hume's problem of justification of induction.

A further problem remains. Probabilities of the frequency variety are used in two ways. On the one hand, they appear in statistical laws, such as the law of radioactive decay of unstable species of nuclei. On the other hand, they are often applied in making predictions of single events, or finite classes of events. Pollsters, for example, predict outcomes of single elections on the basis of interviews with samples of voters. If, however, probability is defined as a limiting frequency in a potentially infinite sequence of events, it does not seem to make any sense to talk about probabilities of single occurrences. The problem of the single case raises a problem about the applicability of the frequency interpretation of probability.

Before we leave the frequency interpretation a word of caution is in order. The
frequency interpretation and the classical interpretations are completely different from one another, and they should not be confused. When the classical interpretation refers to possible outcomes and favorable outcomes it is referring to types or classes of events—for example, the class of all cases in which heads comes up is one possible outcome; the class of cases in which tails comes up is one other possible outcome. In this example there are only two possible outcomes. These classes—not their members—are what you count for purposes of the classical interpretation. In the frequency interpretation, it is the members of these classes that are counted. If the coin is tossed a large number of times there are many heads and many tails. In the frequency interpretation, the numbers of items of which ratios are formed keep changing as the number of individual events increases. In the classical interpretation, the probability does not depend in any way on how many heads or tails actually occur.

3. The propensity interpretation. The propensity interpretation is a relatively recent innovation in the theory of probability. Although suggested earlier, particularly by Charles Saunders Peirce, it was first clearly articulated by Popper (1957b, 1960). It was introduced specifically to deal with the problem of the single case.

The sort of situation Popper originally envisaged was a potentially infinite sequence of tosses of a loaded die that was biased in such a way that side 6 had a probability of 1/4. The limiting frequency of 6 in this sequence is, of course, 1/4. Suppose, however, that three of the tosses were not made with the biased die, but rather with a fair die. Whatever the outcomes of these three throws, they would have no effect on the limiting frequency. Nevertheless, Popper maintained, we surely want to say that the probability of 6 on those three tosses was 1/6—not 1/4. Popper argued that the appropriate way to deal with such cases is to associate the probability with the chance setup that produces the outcome, rather than to define it in terms of the sequence of outcomes themselves. Thus, he claims, each time the fair die is thrown, the mechanism—consisting of the die and the thrower—has a causal tendency or propensity of 1/6 to produce the outcome 6. Similarly, each time the loaded die is tossed, the mechanism has a propensity of 1/4 to produce the outcome 6.

Although this idea of propensity—probabilistic causal tendency—is important and valuable, it does not provide an admissible interpretation of the probability calculus. This can easily be seen in terms of the case of the frisbee factory introduced in the preceding section. That example, we recall, consisted of two machines, each of which had a certain propensity or tendency to produce defective frisbees. For the new machine the propensity was 0.01; for the old machine it was 0.02. Using the rule of total probability we calculated the propensity of the factory to produce faulty frisbees; it was 0.012. So far, so good.

The problem arises in connection with Bayes’s rule. Having picked a defective frisbee at random from the day’s production, we asked for the probability that it was produced by the new machine; the answer was 2/3. This is a perfectly legitimate probability, but it cannot be construed as a propensity. It makes no sense to say that this frisbee has a propensity of 2/3 to have been produced by the new machine. Either it was produced by the new machine or by the old. It does not have a tendency of 1/3 to have been produced by the old machine and a tendency of 2/3 to have been produced by the new one. The basic point is that causes pre-
cede their effects and causes produce their effects, even if the causal relationship has probabilistic aspects. We can speak meaningfully of the causal tendency of a machine to produce a faulty product. Effects do not produce their causes. It does not make sense to talk about the causal tendency of the effect to have been produced by one cause or another.

Bayes's rule enables us to compute what are sometimes called inverse probabilities. Whereas the rule of total probability enables us to calculate the forward probability of an effect, given suitable information about antecedent causal factors, Bayes's rule allows us to compute the inverse probability that a given effect was produced by a particular cause. These inverse probabilities are an integral part of the mathematical calculus of probability, but no propensities correspond to them. For this reason the propensity interpretation is not an admissible interpretation of the probability calculus.

4. The subjective interpretation. Both the frequency interpretation and the propensity interpretation are regarded by their proponents as types of physical probabilities. They are objective features of the real world. But probability seems to many philosophers and mathematicians to have a subjective side as well. This aspect has something to do with the degree of conviction with which an individual believes in one proposition or another. For instance, Mary Smith is sure that it will be cold in Montana next winter—that is, in some place in that state the temperature will fall below 50 degrees Fahrenheit between 21 December and 21 March. Her subjective probability for this event is extremely close to 1. Also, she disbelieves completely that Antarctica will be hot any time during its summer—that is, she is sure that the temperature will not rise above 100 degrees Fahrenheit between 21 December and 21 March. Her subjective probability for real heat in Antarctica in summer is very close to 0. She neither believes in rain in Pittsburgh tomorrow, nor disbelieves in rain in Pittsburgh tomorrow; her conviction for either one of these alternatives is just as strong as for the other. Her subjective probability for rain tomorrow in Pittsburgh is just about 1/2. As she runs through the various propositions in which she might believe or disbelieve she finds a range of degrees of conviction spanning the whole scale from 0 to 1. Other people will, of course, have different degrees of conviction in these same propositions.

It is easy to see immediately that subjective degrees of commitment do not provide an admissible interpretation of the probability calculus. Take a simple example. Many people believe that the probability of getting a 6 with a fair die is 1/6, and that the outcomes of successive tosses are independent of one another. They also believe that we have a fifty-fifty chance of getting 6 at least once in three throws. As we saw in the previous section, however, that probability is significantly below 1/2. Therefore, the preceding set of degrees of conviction violate the mathematical calculus of probability. Of course, not everyone makes that particular mistake, but extensive empirical research has shown that most of us do make various kinds of mistakes in dealing with probabilities. In general, a given individual's degrees of conviction fail to satisfy the mathematical calculus.

5. Personal probabilities. What if there were a person whose degrees of conviction did not violate the probability calculus? That person's subjective proba-
bilities would constitute an admissible interpretation. Whether there actually is any such person, we can think of such an organization of our degrees of conviction as an ideal.

Compare this situation with deductive logic. One of its main functions is to help us avoid certain types of logical errors. Anyone who believes, for example, that all humans are mortal and Socrates is human, but that Socrates is immortal, is guilty of self-contradiction. Whoever wants to believe only what is true must try to avoid contradictions, for contradictions cannot possibly be true. In this example, among the three statements, "All humans are mortal," "Socrates is human," and "Socrates is immortal," at least one must be false. Logic does not tell us which statement is false, but it does tell us to make some change in our set of beliefs if we do not want to believe falsehoods. A person who avoids logical contradictions—inhconsistencies—has a consistent set of beliefs.

A set of degrees of conviction that violate the calculus of probability is said to be incoherent. Anyone who holds a degree of conviction of 1/6 that a fair die, when tossed, will come up 6, and who also considers successive tosses independent (whose degree of conviction in 6 on the next toss is not affected by the outcome of previous tosses), and who is convinced to the degree 1/2 that 6 will come up at least once in three tosses, is being incoherent. So also is anyone who assigns two different values to the probability that a martini mixed by Joe, the sloppy bartender, is between 3:1 and 4:1.

A serious penalty results from being incoherent. A person who has an incoherent set of degrees of conviction is vulnerable to a Dutch book. A Dutch book is a set of bets such that, no matter what the outcome of the event on which the bets are made, the subject loses. Consider a very simple example. The negation rule of the probability calculus tells us that \( Pr(B|A) \) and \( Pr(\neg B|A) \) must add up to 1. Suppose someone has a degree of conviction of 2/3 that the next toss of a particular coin will result in heads, and also a degree of conviction of 2/3 that it will result in tails. This person should be willing to bet at odds of 2 to 1 that the coin will come up heads, and also at odds of 2 to 1 that it will come up tails. These bets constitute a Dutch book because, if the coin comes up heads the subject wins $1 but loses $2, and if it comes up tails the subject loses $2 and wins $1. Since these are the only possible outcomes, the subject loses $1 no matter what happens.

It has been proved in general that a person is subject to a Dutch book if and only if that person holds an incoherent set of degrees of conviction. Thus, we can look at the probability calculus as a kind of system of logic—the logic of degrees of conviction. Conforming to the rules of the probability calculus enables us to avoid certain kinds of blunders in probabilistic reasoning, namely, the type of error that makes one subject to a Dutch book. In light of these considerations, personal probabilities have been defined as coherent sets of degrees of conviction. It follows immediately that personal probabilities constitute an admissible interpretation of the probability calculus, for they have been defined in just that way.

One of the major motivations of those who accept the personalist interpretation of probability lies in the use of Bayes's rule; indeed, those who adhere to personal probabilities are often called "Bayesians." To see why, let us take another look at Bayes's rule (Rule 9):
\[
Pr(H \mid K, E) = \frac{Pr(H \mid K) \times Pr(E \mid K, H)}{Pr(H \mid K) \times Pr(E \mid K, H) + Pr(\neg H \mid K) \times Pr(E \mid K, \neg H)}
\]

provided that \(Pr(E \mid K) \neq 0\).

Consider the following simple example. Suppose that someone in the next room is flipping a penny, and that we receive a reliable report of the outcome after each toss. We cannot inspect the penny, but for some reason we suspect that it is a two-headed coin. To keep the arithmetic simple, let us assume that the coin is either two-headed or fair. Let \(K\) stand for our background knowledge and opinion, \(H\) for the hypothesis that the coin is two-headed, and \(E\) for the results of the flips. For any given individual \(Pr(H \mid K)\)—the prior probability—represents that person’s antecedent degree of conviction that the coin is two-headed before any of the outcomes have been reported. Probability \(Pr(E \mid K, H)\)—one of the likelihoods—is the probability of the outcome reported to us, given that the coin being flipped is two-headed. If an outcome of tails is reported, that probability obviously equals zero, and the hypothesis \(H\) is refuted.

If one or more heads are reported, that probability clearly equals 1. The probability \(Pr(\neg H \mid K)\) is the prior probability that the coin is not two-headed—that is, that it is fair. On pain of incoherence, this probability must equal \(1 - Pr(H \mid K)\). The probability \(Pr(E \mid \neg H, K)\) is also a likelihood; it is the probability of the outcomes reported to us given that the coin is not two-headed. The probability \(Pr(H \mid K, E)\)—the posterior probability—is the probability that the coin is two-headed given both our background knowledge and knowledge of the results of the tosses. That probability represents an assessment of the hypothesis in the light of the observational evidence (reported reliably to us).

Suppose that John’s prior personal probability that the coin is two-headed is \(1/100\). The result of the first toss is reported, and it is a head. Using Bayes’s rule, he computes the posterior probability as follows:

\[
\frac{1/100 \times 1}{1/100 \times 1 + 99/100 \times 1/2} = \frac{2}{101} \approx 0.02.
\]

After two heads the result would be

\[
\frac{1/100 \times 1}{1/100 \times 1 + 99/100 \times 1/4} = \frac{4}{103} \approx 0.04.
\]

After ten heads the result would be

\[
\frac{1/100 \times 1}{1/100 \times 1 + 99/100 \times 1/1024} = \frac{1024}{1123} \approx 0.91.
\]

Suppose Wes’s personal prior probability, before any outcomes are known, is much higher than John’s; Wes has a prior conviction of \(1/2\) that the coin is two-headed. After the first head, he makes the following calculation:

\[
\frac{1/2 \times 1}{1/2 \times 1 + 1/2 \times 1/2} = \frac{2}{3} = 0.67.
\]

After the second head, he has

The Confirmation of Scientific Hypotheses
\[
\frac{1/2 \times 1}{1/2 \times 1 + 1/2 \times 1/4} = \frac{4/5}{} = 0.80.
\]

After ten heads, he has

\[
\frac{1/2 \times 1}{1/2 \times 1 + 1/2 \times 1/1024} = \frac{1024/1025}{1025} > 0.99.
\]

These calculations show two things. First, they show how Bayes's rule can be used to ascertain the probability of a hypothesis if we have values for the prior probabilities. If we employ personal probabilities the prior probabilities become available. They are simply a person's degrees of conviction in the hypothesis prior to receipt of the observational evidence. In this kind of example the likelihoods can be calculated from assumptions we share concerning the behavior of fair and two-headed coins.

Second, these calculations illustrate a phenomenon known as washing out of the priors or swamping of the priors. Notice that we did two sets of calculations—one for John and one for Wes. We started with widely divergent degrees of conviction in the hypothesis; Wes's was 1/2 and John's was 1/100. As the evidence accumulated our degrees of conviction became closer and closer. After ten heads, Wes's degree of conviction is approximately 0.99 and John's is approximately 0.91. As more heads occur our agreement becomes even stronger. This illustrates a general feature of Bayes's rule. Suppose there are two people with differing prior probabilities—as far apart as you like provided neither has an extreme value of 0 or 1. Then, if they agree on the likelihoods and if they share the same observational evidence, their posterior probabilities will get closer and closer together as the evidence accumulates. The influence of the prior probabilities on the posterior probabilities decreases as more evidence becomes available. This phenomenon of washing out of the priors should help to ease the worry we might have about appealing to admittedly subjective degrees of conviction in our evaluations of scientific hypotheses.

Still, profound problems are associated with the personalistic interpretation of probability. The only restriction imposed by this interpretation on the values of probabilities is that they be coherent—that they satisfy the rules of mathematical probability. This is a very weak constraint. If we look at the rules of probability we note that, with a couple of trivial exceptions, the mathematical calculus of probability does not by itself furnish us with any values of probabilities. The exceptions are that a logically necessary proposition must have probability 1 and a contradiction must have probability 0. In all other cases, the rules of probability enable us to calculate some probability values from others. You plug in some probability values, turn the crank, and others come out. This means that there need be little contact between our personal probabilities and what goes on in the external world. For example, it is possible for a person to have a degree of conviction of 9/10 that the next toss of a coin will result in heads even though the coin has been tossed hundreds of times and has come up tails on the vast majority of these tosses. By suitably adjusting one's other probabilities one can have such personal probabilities as these without becoming incoherent. If our probabilities are to represent reasonable degrees of conviction some stronger restrictions surely appear to be needed.
6. The logical interpretation. One of the most ambitious twentieth-century attempts to deal with the problems of probability and confirmation was the construction of a theory of logical probability by Rudolf Carnap. Carnap was not the first to make efforts in that direction, but his was the most systematic and precise. In fact, Carnap maintained that there are two important and legitimate concepts of probability—relative frequencies and logical probabilities—but his main work was directed toward the latter. He referred to logical probability as degree of confirmation. Many philosophers refer to logical probability as inductive probability. The three terms are essentially interchangeable.

Carnap’s program was straightforward in intent. He believed that it is possible to develop a formal inductive logic along much the same lines as formal deductive logic. In fact, he constructed a basic logical language in which both deductive and inductive relations would reside. In deductive logic, if a statement \( E \) entails another statement \( H \), \( E \) supports \( H \) completely—if \( E \) is true \( H \) must also be true. In inductive logic, if \( E \) is evidence for a hypothesis \( H \), \( E \) provides some sort of partial support for \( H \); indeed, this type of partial support is often referred to as partial entailment.

The easiest way to understand what Carnap did is to work out the details of a simple and highly artificial example. Let us construct a language which deals with a universe containing only three entities, and each of these entities has or lacks one property. We let \( a \), \( b \), and \( c \) denote the three individuals, and we use \( F \) to designate the property. To make the example concrete, we can think of the individuals as three balls and the property as red. The notation \( Fa \) says that the first ball is red; \( \sim Fa \) says that it is not red. We need a few other basic logical symbols. We use \( x \) and \( y \) as variables for individuals, and \((x)\), which is known as the universal quantifier, means “for every \( x \).” The notation \((\exists x)\), which is known as the existential quantifier, means “there exists at least one \( x \) such that.” A dot “.” is used for the conjunction and; a wedge “\( \lor \)’’ for the disjunction or. That is about all of the logical equipment we will need.

The model universe we are discussing is a very simple place, and we can describe it completely; indeed, we can describe every logically possible state of this universe. Any such complete description of a possible state is a state description; there are eight:

1. \( Fa.Fb.Fc \)
2. \( Fa.Fb.\sim Fc \)
3. \( Fa.\sim Fb.Fc \)
4. \( \sim Fa.Fb.Fc \)
5. \( Fa.\sim Fb.\sim Fc \)
6. \( \sim Fa.Fb.\sim Fc \)
7. \( \sim Fa.\sim Fb.Fc \)
8. \( \sim Fa.\sim Fb.\sim Fc \)

Any consistent statement that we can form in this miniature language can be expressed by means of these state descriptions. For example, \((x)Fx\), which says that every ball is red, is equivalent to state description 1. The statement \((\exists x)Fx\), which says that at least one ball is red, is equivalent to the disjunction of state descriptions 1–7; that is, it says that either state description 1 or 2 or 3 or 4 or 5 or 6 or 7 is true. \( Fa \) is equivalent to the disjunction of state descriptions 1, 2, 3, and 5. \( Fa.Fb \) is equivalent to the disjunction of state descriptions 1 and 2. If we agree to admit—just for the sake of convenience—that there can be disjunctions with only one term, we
can say that every consistent statement is equivalent to some disjunction of state
descriptions. The state descriptions in any such disjunction constitute the range of
that statement. A contradictory statement is equivalent to the denial of all eight of the
state descriptions. Its range is empty.

In the following discussion, $H$ is any statement that is being taken as a hy-
pothesis and $E$ any statement that is being taken as evidence. In this discussion any
consistent statement that can be formulated in our language can serve as a statement
of evidence, and any statement—consistent or inconsistent—can serve as a hy-
pothesis. Now, consider the hypothesis $\exists x Fx$ and evidence $Fc$. Clearly this ev-
eidence deductively entails this hypothesis; if the third ball is red at least one must
be red. If we look at the ranges of this evidence and this hypothesis, we see that
the range of $Fc$ (state descriptions 1, 3, 4, 7) is entirely included in the range of
$\exists x Fx$ (state descriptions 1–7). This situation always holds. If one statement en-
tails another, the range of the first is included within the range of the second. This
means that every possible state of the universe in which the first is true is a possible
state of the universe in which the second is true. If two statements have identical
ranges, they are logically equivalent, and each one entails the other. If two state-
ments are logically incompatible with one another, their ranges do not overlap at
all—that is, there is no possible state of the universe in which they can both be
ture. We see, then, that deductive relationships can be represented as relationships
among the ranges of the statements involved.

Let us now turn to inductive relationships. Consider the hypothesis $(x)Fx$ and
the evidence $Fa$. This evidence obviously does not entail the hypothesis, but it seems
reasonable to suppose that it provides some degree of inductive support or confirma-
tion. The range of the evidence $(1, 2, 3, 5)$ is not completely included in the range
of the hypothesis $(1)$, but it does overlap that range—the two ranges have state
description 1 in common. What we need is a way of expressing the idea of confirma-
tion in terms of the overlapping of ranges. When we take any statement $E$ as
evidence, we are accepting it as true; in so doing we are ruling out all possible states
of the universe that are incompatible with the evidence $E$. Having ruled out all of
those, we want to know to what degree the possible states in which the evidence holds
true are possible states in which the hypothesis also holds true. This can be expressed
in the form of a ratio, $\operatorname{range} (E,H)/\operatorname{range} (E)$, and this is the basic idea behind the
concept of degree of confirmation.

Consider the range of $(x)Fx$; this hypothesis holds in one state description out
of eight. If, however, we learn that $Fa$ is true, we rule out four of the state descrip-
tions, leaving only four as possibilities. Now the hypothesis holds in one out of four.
If we now discover that $Fb$ is also true, our combined evidence $Fa.Fb$ holds in only
two state descriptions, and our hypothesis holds in one of the two. It looks reasonable
to say that our hypothesis had a probability of 1/8 on the basis of no evidence, a
probability of 1/4 on the basis of the first bit of evidence, and a probability of 1/2 on
the two pieces of evidence. (This suggestion was offered by Wittgenstein 1922). But
appearances are deceiving in this case.

If we were to adopt this suggestion as it stands, Carnap realized, we would
rule out altogether the possibility of learning from experience; we would have no
basis at all for predicting future occurrences. Consider, instead of \((x)Fx\), the hypothesis \(Fc\). By itself, this hypothesis holds in four \((1, 3, 4, 7)\) out of eight state descriptions. Suppose we find as evidence that \(Fa\). The range of this evidence is four state descriptions \((1, 2, 3, 5)\), and the hypothesis holds in two of them. But \(4/8 = 2/4\), so the evidence has done nothing to support the hypothesis. Moreover, if we learn that \(Fb\) is true our new evidence is \(Fa.Fb\), which holds in two state descriptions \((1, 2)\), and our hypothesis holds in only one of them, giving us a ratio of \(1/2\). Hence, according to this way of defining confirmation, what we observe in the past and present has no bearing on what will occur in the future. This is an unacceptable consequence. When we examined the hypothesis \((x)Fx\) in the preceding paragraph we appeared to be achieving genuine confirmation, but that was not happening at all. The hypothesis \((x)Fx\) simply states that \(a, b,\) and \(c\) all have property \(F\). When we find out by observing the first ball that it is red, we have simply reduced the predictive content of \(h\). At first it predicted the color of three balls; after we examine the first ball it predicts the color of only two balls. After we observe the second ball, the hypothesis predicts the color of only one ball. If we were to examine the third ball and find it to be red, our hypothesis would have no predictive content at all. Instead of confirming our hypothesis we were actually simply reducing its predictive import.

In order to get around the foregoing difficulty, Carnap proposed a different way of evaluating the ranges of statements. The method adopted by Wittgenstein amounts to assigning equal weights to all of the state descriptions. Carnap suggested assigning unequal weights on the following basis. Let us take another look at our list of state descriptions in Table 2.2:

<table>
<thead>
<tr>
<th>State Description</th>
<th>Weight</th>
<th>Structure Description</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. (Fa.Fb.Fc)</td>
<td>(1/4)</td>
<td>(All F)</td>
<td>(1/4)</td>
</tr>
<tr>
<td>2. (Fa.Fb.\neg Fc)</td>
<td>(1/12)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. (Fa.\neg Fb.Fc)</td>
<td>(1/12)</td>
<td>(2 F, 1 \neg F)</td>
<td>(1/4)</td>
</tr>
<tr>
<td>4. (\neg Fa.Fb.Fc)</td>
<td>(1/12)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. (Fa.\neg Fb.\neg Fc)</td>
<td>(1/12)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. (\neg Fa.Fb.\neg Fc)</td>
<td>(1/12)</td>
<td>(1 F, 2 \neg F)</td>
<td>(1/4)</td>
</tr>
<tr>
<td>7. (\neg Fa.\neg Fb.Fc)</td>
<td>(1/12)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. (\neg Fa.\neg Fb.\neg Fc)</td>
<td>(1/4)</td>
<td>(No F)</td>
<td>(1/4)</td>
</tr>
</tbody>
</table>

Carnap noticed that state descriptions 2, 3, and 4 make similar statements about our miniature universe; they say that two entities have property \(F\) and one lacks it. Taken together, they describe a certain structure. They differ from one another in identifying the ball that is not red, but Carnap suggests that that is a secondary consideration. Similarly, state descriptions 5, 6, and 7, taken together describe a certain structure, namely, a universe in which one individual has property \(F\) and two lack it. Again, they differ in identifying the object that has this property. In contrast,
state description 1, all by itself, describes a particular structure, namely, all threeentities have property F. Similarly, state description 8 describes the structure in which no object has that property.

Having identified the structure descriptions, Carnap proceeds to assign equal weights to them (each gets 1/4); he then assigns equal weights to the state descriptions within each structure description. The resulting system of weights is shown above. These weights are then used as a measure of the ranges of statements; this system of measures is called \( m^* \). A confirmation function \( c^* \) is defined as follows:\(^{12}\)

\[
c^*(H|E) = \frac{m^*(H.E)}{m^*(E)}.
\]

To see how it works, let us reconsider the hypothesis \( Fc \) in the light of different bits of evidence. First, the range of \( Fc \) consists of state description 1, which has weight 1/4, and 3, 4, and 7, each of which has weight 1/12. The sum of all of them is 1/2; that is, the probability of our hypothesis before we have any evidence. Now, we find that \( Fa \); its measure is 1/2. The range of \( Fa.Fc \) is state descriptions 1 and 3, whose weights are, respectively, 1/4 and 1/12, for a total of 1/3. We can now calculate the degree of confirmation of our hypothesis on this evidence:

\[
c^*(H|E) = \frac{m^*(E.H)}{m^*(E)} = \frac{1}{3} \div \frac{1}{2} = \frac{2}{3}.
\]

Carrying out the same sort of calculation for evidence \( Fa.Fb \) we find that our hypothesis has degree of confirmation 3/4. If, however, our first bit of evidence had been \( \sim Fa \), the degree of confirmation of our hypothesis would have been 1/3. If our second bit of evidence had been \( \sim Fb \), that would have reduced its degree of confirmation to 1/4. The confirmation function \( c^* \) seems to do the right sorts of things. When the evidence is what we normally consider to be positive, the degree of confirmation goes up. When the evidence is what we usually take to be negative, the degree of confirmation goes down. Clearly, \( c^* \) allows for learning from experience.

A serious philosophical problem arises, however. Once we start playing the game of assigning weights to state descriptions, we face a huge plethora of possibilities. In setting up the machinery of state descriptions and weights, Carnap demands only that the weights for all of the state descriptions add up to 1, and that each state description have a weight greater than 0. These conditions are sufficient to guarantee an admissible interpretation of the probability calculus. Carnap recognized the obvious fact that infinitely many confirmation functions satisfying this basic requirement are possible. The question is how to make an appropriate choice. It can easily be shown that choosing a confirmation function is precisely the same as assigning prior probabilities to all of the hypotheses that can be stated in the given language.

Consider the following possibility for a measure function:

\(^{11}\) The measure of the range of any statement \( H \) can be identified with the prior probability of that statement in the absence of any background knowledge \( K \). It is an \textit{a priori} prior probability.

\(^{12}\) Wittgenstein’s measure function assigns the weight ⅛ to each state description; the confirmation function based upon it is designated \( c? \).
TABLE 2.3

<table>
<thead>
<tr>
<th>State Description</th>
<th>Weight</th>
<th>Structure Description</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Fa.Fb.Fc</td>
<td>1/20</td>
<td>All F</td>
<td>1/20</td>
</tr>
</tbody>
</table>

(The idea of a confirmation function of this type was given in Burks 1953; the philosophical issues are further discussed in Burks 1977, Chapter 3.) This method of weighting, which may be designated $m^\circ$, yields a confirmation function $C^\circ$, which is a sort of counterinductive method. Whereas $m^*$ places higher weights on the first and last state descriptions, which are state descriptions for universes with a great deal of uniformity (either every object has the property, or none has it), $m^\circ$ places lower weights on descriptions of uniform universes. Like $c^*$, $c^\circ$ allows for “learning from experience,” but it is a funny kind of anti-inductive “learning.” Before we reject $m^\circ$ out of hand, however, we should ask ourselves if we have any a priori guarantee that our universe is uniform. Can we select a suitable confirmation function without being totally arbitrary about it? This is the basic problem with the logical interpretation of probability.

**Part IV: Confirmation and Probability**

2.9 THE BAYESIAN ANALYSIS OF CONFIRMATION

We now turn to the task of illustrating how the probabilistic apparatus developed above can be used to illuminate various issues concerning the confirmation of scientific statements. Bayes’s theorem (Rule 9) will appear again and again in these illustrations, justifying the appellation of Bayesian confirmation theory.

Various ways are available to connect the probabilistic concept of confirmation back to the qualitative concept, but perhaps the most widely followed route utilizes an incremental notion of confirmation: $E$ confirms $H$ relative to the background knowledge $K$ just in case the addition of $E$ to $K$ raises the probability of $H$, that is, $Pr(H|E.K) > Pr(H|K)$.[13] Hempel’s study of instance confirmation in terms of a

---

[13] Sometimes, when we say that a hypothesis has been confirmed, we mean that it has been rendered highly probable by the evidence. This is a high probability or absolute concept of confirmation, and it should be carefully distinguished from the incremental concept now under discussion (see Carnap 1962, Salmon 1973, and Salmon 1975). Salmon (1973) is the most elementary discussion.
two-place relation can be taken to be directed at the special case where $K$ contains no information. Alternatively, we can suppose that $K$ has been absorbed into the probability function in the sense that $Pr(K) = 1$, in which case the condition for incremental confirmation reduces to $Pr(H|E) > Pr(H)$. (The unconditional probability $Pr(H)$ can be understood as the conditional probability $Pr(H|T)$, where $T$ is a vacuous statement, for example, a tautology. The axioms of Section 2.7 apply only to conditional probabilities.)

It is easy to see that on the incremental version of confirmation, Hempel’s consistency condition is violated as is

**Conjunction condition:** If $E$ confirms $H$ and also $H'$ then $E$ confirms $H \cdot H'$.

It takes a bit more work to construct a counterexample to the special consequence condition. (This example is taken from Carnap 1950 and Salmon 1975, the latter of which contains a detailed discussion of Hempel’s adequacy conditions in the light of the incremental notion of confirmation.) Towards this end take the background knowledge to contain the following information. Ten players participate in a chess tournament in Pittsburgh; some are locals, some are from out of town; some are juniors, some are seniors; and some are men ($M$), some are women ($W$). Their distribution is given by

<table>
<thead>
<tr>
<th></th>
<th>Locals</th>
<th>Out-of-towners</th>
</tr>
</thead>
<tbody>
<tr>
<td>Juniors</td>
<td>$M, W, W$</td>
<td>$M, M$</td>
</tr>
<tr>
<td>Seniors</td>
<td>$M, M$</td>
<td>$W, W, W$</td>
</tr>
</tbody>
</table>

And finally, each player initially has an equal chance of winning. Now consider the hypotheses $H$: an out-of-towner wins, and $H'$: a senior wins, and the evidence $E$: a woman wins. We find that

$$Pr(H|E) = 3/5 > Pr(H) = 1/2$$

so $E$ confirms $H$. But

$$Pr(H \lor H'|E) = 3/5 < (Pr(H \lor H') = 7/10.$$ 

So $E$ does not confirm $H \lor H'$; in fact $E$ confirms $\sim(H \lor H')$ and so disconfirms $H \lor H'$ even though $H \lor H'$ is a consequence of $H$.

The upshot is that on the incremental conception of confirmation, Hempel’s adequacy conditions and, hence, his definition of qualitative confirmation, are inadequate. However, his adequacy conditions fare better on the high probability conception of confirmation according to which $E$ confirms $H$ relative to $K$ just in case $Pr(H|E.K) > r$, where $r$ is some number greater than 0.5. But this notion of

\[\text{As would be the case if learning from experience is modeled as change of probability function through conditionalization; that is, when } K \text{ is learned, } Pr_{\text{old}}(\cdot) = Pr_{\text{new}}(\cdot | K) \text{. From this point of view, Bayes’s theorem (Rule 9) describes how probability changes when a new fact is learned.}\]
confirmation cannot be what Hempel has in mind; for he wants to say that the observation of a single black raven (E) confirms the hypothesis that all ravens are black (H), although for typical K, Pr(H\E\K) will surely not be as great as 0.5. Thus, in what follows we continue to work with the incremental concept.

The probabilistic approach to confirmation coupled with a simple application of Bayes's theorem also serves to reveal a kernel of truth in the H-D method. Suppose that the following conditions hold:

(i) H, K ⊨ E; (ii) 1 > Pr(H\K) > 0; and (iii) 1 > Pr(E\K) > 0.

Condition (i) is the basic H-D condition. Conditions (ii) and (iii) say that neither H nor E is known on the basis of the background information K to be almost surely false or almost surely true. Then on the incremental conception it follows, as the H-D methodology would have it, that E confirms H on the basis of K. By Bayes's theorem

\[ Pr(H\mid E\cdot K) = \frac{Pr(H\mid K)}{Pr(E\mid K)} \]

since by (i),

\[ Pr(E\mid H\cdot K) = 1. \]

It then follows from (ii) and (iii) that

\[ Pr(H\mid E\cdot K) > Pr(H\mid K). \]

Notice also that the smaller Pr(E\mid K) is, the greater the incremental confirmation afforded by E. This helps to ground the intuition that "surprising" evidence gives better confirmational value. However, this observation is really double-edged as will be seen in Section 2.10.

The Bayesian analysis also affords a means of handling a disquieting feature of the H-D method, sometimes called the problem of irrelevant conjunction. If the H-D condition (i) holds for H, then it also holds for H.X where X is anything you like, including conjuncts to which E is intuitively irrelevant. In one sense the problem is mirrored in the Bayesian approach, for assuming that 1 > Pr(H.X\K) > 0, it follows that E incrementally confirms H.X. But since the special consequence condition does not hold in the Bayesian approach, we cannot infer that E confirms the consequence X of H.X. Moreover, under the H-D condition (i), the incremental confirmation of a hypothesis is directly proportional to its prior probability. Since Pr(H\mid K) ≥ Pr(H.X\mid K), with strict inequality holding in typical cases, the incremental confirmation for H will be greater than for H.X.

Bayesian methods are flexible enough to overcome various of the shortcomings of Hempel's account. Nothing, for example, prevents the explication of confirmation in terms of a Pr-function which allows observational evidence to boost the probability of theoretical hypotheses. In addition the Bayesian approach illuminates the paradoxes of the ravens and Goodman's paradox.

In the case of the ravens paradox we may grant that the evidence that the individual a is a piece of white chalk can confirm the hypothesis that "All ravens are black" since, to put it crudely, this evidence exhausts part of the content of the

The Confirmation of Scientific Hypotheses 91
hypothesis. Nevertheless, as Suppes (1966) has noted, if we are interested in subjecting the hypothesis to a sharp test, it may be preferable to do outdoor ornithology and sample from the class of ravens rather than sampling from the class of nonblack things. Let $a$ denote a randomly chosen object and let

$$
Pr(Ra.Ba) = p_1, \quad Pr(Ra.\neg Ba) = p_2
$$
$$
Pr(\neg Ra.Ba) = p_3, \quad Pr(\neg Ra.\neg Ba) = p_4.
$$

Then

$$
Pr(\neg Ba|Ra) = p_2 \neq (p_1 + p_2)
$$
$$
Pr(Ra|\neg Ba) = p_2 \neq (p_2 + p_4)
$$

Thus, $Pr(\neg Ba|Ra) > Pr(Ra|\neg Ba)$ just in case $p_4 > p_1$. In our world it certainly seems true that $p_4 > p_1$. Thus, Suppes concludes that sampling ravens is more likely to produce a counterinstance to the ravens hypothesis than is sampling the class of nonblack things.

There are two problems here. The first is that it is not clear how the last statement follows since $a$ was supposed to be an object drawn at random from the universe at large. With that understanding, how does it follow that $Pr(\neg Ba|Ra)$ is the probability that an object drawn at random from the class of ravens is nonblack? Second, it is the anti-inductivists such as Popper (see item 4 in Section 2.8 above and 2.10 below) who are concerned with attempts to falsify hypotheses. It would seem that the Bayesian should concentrate on strategies that enhance absolute and incremental probabilities. An approach due to Gaifman (1979) and Horwich (1982) combines both of these points.

Let us make it part of the background information $K$ that $a$ is an object drawn at random from the class of ravens while $b$ is an object drawn at random from the class of nonblack things. Then an application of Bayes’s theorem shows that

$$
Pr(H|Ra.Ba.K) > Pr(H|\neg Rb.\neg Bb.K)
$$

just in case

$$
1 > Pr(\neg Rb|K) > Pr(Ba|K).
$$

To explore the meaning of the latter inequality, use the principle of total probability to find that

$$
Pr(Ba|K) = Pr(Ba|H.K) \cdot Pr(H|K) + Pr(Ba|\neg H.K) \cdot Pr(\neg H|K)
$$

and that

$$
Pr(\neg Rb|K) = Pr(H|K) + Pr(\neg Rb|\neg H.K) \cdot Pr(\neg H|K).
$$

So the inequality in question holds just in case

$$
1 > Pr(\neg Rb|\neg H.K) > Pr(Ba|\neg H.K),
$$

or

$$
Pr(\neg Ba|\neg H.K) > Pr(Rb|\sim H.K) > 0,
$$
which is presumably true in our universe. For supposing that some ravens are nonblack, a random sample from the class of ravens is more apt to produce such a bird than is a random sample from the class of nonblack things since the class of nonblack things is much larger than the class of ravens. Thus, under the assumption of the stated sampling procedures, the evidence Ra.Ba does raise the probability of the ravens hypothesis more than the evidence ~Rb.~Bb does. The reason for this is precisely the differential propensities of the two sampling procedures to produce counterexamples, as Suppes originally suggested.

The Bayesian analysis also casts light on the problems of induction, old and new, Humean and Goodmanian. Russell (1948) formulated two categories of induction by enumeration:

Induction by simple enumeration is the following principle: "Given a number n of α’s which have been found to be β’s, and no α which has been found to be not a β, then the two statements: (a) ‘the next α will be a β,’ (b) ‘all α’s are β’s,’ both have a probability which increases as n increases, and approaches certainty as a limit as n approaches infinity.’"

I shall call (a) "particular induction" and (b) "general induction." (1948, 401)

Between Russell’s "particular induction" and his "general induction" we can interpolate another type, as the following definitions show (note that Russell’s "α" and "β" refer to properties, not to individual things):

\[ \lim_{n \to \infty} Pr(P_{a_{n+1}}|P_{a_1}, \ldots, P_{a_n}, K) = 1. \]

Def. Relative to K, the predicate "P" is weakly projectible over the sequence of individuals \( a_1, a_2, \ldots \) just in case

\[ \lim_{n \to \infty} Pr(P_{a_{n+1}} \ldots P_{a_{n+m}}|P_{a_1}, \ldots, P_{a_n}, K) = 1. \]

Def. Relative to K, "P" is strongly projectible over \( a_1, a_2, \ldots \) just in case

\[ \lim_{n, m \to \infty} Pr(P_{a_{n+1}} \ldots P_{a_{n+m}}|P_{a_1}, \ldots, P_{a_n}, K) = 1. \]

(The notation \( \lim_{m, n \to \infty} \) indicates the limit as \( m \) and \( n \) both tend to infinity in any manner you like.) A sufficient condition for both weak and strong probability is that the general hypothesis \( H \): \( (i)P_a \) receives a nonzero prior probability. To see that it is sufficient for weak projectibility, we follow Jeffreys’s (1957) proof. By Bayes’s theorem

\[ Pr(H|P_{a_1}, \ldots, P_{a_{n+1}}, K) = \frac{Pr(P_{a_1}, \ldots, P_{a_{n+1}}|H, K) \cdot Pr(H|K)}{Pr(P_{a_1}, \ldots, P_{a_{n+1}}|K)} \]

\[ = \frac{Pr(H|K)}{Pr(P_{a_1}|K) \cdot Pr(P_{a_2}|P_{a_1}, K) \cdots \cdot Pr(P_{a_{n+1}}|P_{a_1}, \ldots, P_{a_n}, K)} \]

15 Equation \( \lim_{n \to \infty} x_n = L \) means that, for any real number \( \epsilon > 0 \), there is an integer \( N > 0 \) such that, for all \( n > N \), \( |x_n - L| < \epsilon \).

The Confirmation of Scientific Hypotheses
Unless $Pr(P_{a_{n+1}} | P_{a_1}, \ldots, P_{a_n}, K)$ goes to 1 as $n \rightarrow \infty$, the denominator on the right-hand side of the second equality will eventually become less than $Pr(H|K)$, contradicting the truth of probability that the left-hand side is no greater than 1.

The posit that

$$\text{(P) } Pr([\exists i)P_{a_i}|K] > 0$$

is not necessary for weak projectibility. Carnap’s systems of inductive logic (see item 6 in Section 2.8 above) are relevant examples since in these systems (P) fails in a universe with an infinite number of individuals although weak projectibility can hold in these systems.\(^\text{16}\) But if we impose the requirement of countable additivity

$$\text{(CA) } \lim_{n \rightarrow \infty} Pr(P_{a_1}, \ldots, P_{a_n}|K) = Pr([\exists i)P_{a_i}|K]$$

then (P) is necessary as well as sufficient for strong projectibility.

Also assuming (CA), (P) is sufficient to generate a version of Russell’s “general induction,” namely

$$\text{(G) } \lim_{n \rightarrow \infty} Pr([\exists i)P_{a_i}|P_{a_1}, \ldots, P_{a_n}, K] = 1.$$  

(Russell 1948 lays down a number of empirical postulates he thought were necessary for induction to work. From the present point of view these postulates can be interpreted as being directed to the question of which universal hypotheses should be given nonzero priors.)

Humean skeptics who regiment their beliefs according to the axioms of probability cannot remain skeptical about the next instance or the universal generalization in the face of ever-increasing positive instances (and no negative instances) unless they assign a zero prior to the universal generalization. But

$$Pr([\exists i)P_{a_i}|K] = 0$$

implies that

$$Pr([\exists i) \sim P_{a_i}|K] = 1,$$

which says that there is certainty that a counterinstance exists, which does not seem like a very skeptical attitude.

\(^\text{16}\) A nonzero prior for the general hypothesis is a necessary condition for strong projectibility but not for weak projectibility. The point can be illustrated by using de Finetti’s representation theorem, which says that if $P$ is exchangeable over $a_1, a_2, \ldots$ (which means roughly that the probability does not depend on the order) then:

$$Pr(P_{a_1}, P_{a_2}, \ldots, P_{a_n} | K) = f_0^{1/0*} d\mu(\theta)$$

where $\mu(\theta)$ is a uniquely determined measure on the unit interval $0 \leq \theta \leq 1$. For the uniform measure $d\mu(\theta) = d(\theta)$ we have

$$Pr(P_{a_{n+1}} | P_{a_1}, \ldots, P_{a_n}, K) = n + 1/n + 2$$

and

$$Pr(P_{a_{n+1}} \cdot \ldots \cdot P_{a_{n+m}} | P_{a_1}, \ldots, P_{a_n}, K) = m + 1/n + m + 1.$$
Note also that the above results on instance induction hold whether "P" is a normal or a Goodmanized predicate—for example, they hold just as well for \( P^*a_i \) which is defined as

\[
[(i \leq 2000).Pa_i] \lor [(i > 2000).\neg Pa_i],
\]

where \( Pa_i \) means that \( a_i \) is purple. But this fact just goes to show how weak the results are; in particular, they hold only in the limit as \( n \to \infty \) and they give no information about how rapidly the limit is approached.

Another way to bring out the weakness is to note that (P) does not guarantee even a weak form of Hume projectibility.

*Def.* Relative to \( K \), "P" is weakly Hume projectible over the doubly infinite sequence \( \ldots a_{-2}, a_{-1}, a_0, a_1, a_2, \ldots \) just in case for any \( n \),

\[
\lim_{k \to \infty} Pr(Pa_0 Pa_{n-1}, \ldots Pa_{n-k} | K) = 1.
\]

(To illustrate the difference between the Humean and non-Humean versions of projectibility, let \( Pa_n \) mean that the sun rises on day \( n \). The non-Humean form of projectibility requires that if you see the sun rise on day 1, on day 2, and so on, then for any \( \varepsilon > 0 \) there will come a day \( N \) when your probability that the sun will rise on day \( N + 1 \) will be at least \( 1 - \varepsilon \). By contrast, Hume projectibility requires that if you saw the sun rise yesterday, the day before yesterday, and so on into the past, then eventually your confidence that the sun will rise tomorrow approaches certainty.)

If (P) were sufficient for Hume projectibility we could assign nonzero priors to both (i)\( Pa_i \) and (i)\( P^*a_i \), with the result that as the past instances accumulate, the probabilities for \( Pa_{2001} \) and for \( P^*a_{2001} \) both approach 1, which is a contradiction.

A sufficient condition for Hume projectibility is exchangeability.

*Def.* Relative to \( K \), "P" is exchangeable for \( Pr \) over the \( a_i \)'s just in case for any \( n \) and \( m \)

\[
Pr(\pm Pa_n, \ldots \pm Pa_{n+m} | K) = Pr(\pm Pa_n, \ldots \pm Pa_{n+m} | K)
\]

where \( \pm \) indicates that either \( P \) or its negation may be chosen and \( \{a_i\} \) is any permutation of the \( a_i \)'s in which all but a finite number are left fixed. Should we then use a \( Pr \)-function for which the predicate "purple" is exchangeable rather than the Goodmanized version of "purple"? Bayesianism per se does not give the answer anymore than it gives the answer to who will win the presidential election in the year 2000. But it does permit us to identify the assumptions needed to guarantee the validity of one form or another of induction.

Having touted the virtues of the Bayesian approach to confirmation, it is now only fair to acknowledge that it is subject to some serious challenges. If it can rise to these challenges, it becomes all the more attractive.

### 2.10 CHALLENGES TO BAYESIANISM

1. *Nonzero priors.* Popper (1959) claims that "in an infinite universe . . . the probability of any (non-tautological) universal law will be zero." If Popper were right
and universal generalizations could not be probabilified, then Bayesianism would be
worthless as applied to theories of the advanced sciences, and we would presumably
have to resort to Popper’s method of corroboration (see item 4 in Section 2.8 above).

To establish Popper’s main negative claim it would suffice to show that the prior
probability of a universal generalization must be zero. Consider again $H$: $\langle i \rangle Pa_i$. Since for any $n$

$$H \vdash Pa_1, Pa_2, \ldots, Pa_n,$$

$$Pr(H|K) \leq \lim_{n \to \infty} Pr(Pa_1, \ldots, Pa_n|K).$$

Now suppose that

(1) For all $n$, $Pr(Pa_1, \ldots, Pa_n|K) = Pr(Pa_1|K) \cdot \ldots \cdot Pr(Pa_n|K)$

and that

(E) For any $m$ and $n$, $Pr(Pa_m|K) = Pr(Pa_n|K)$.

Then except for the uninteresting case that $Pr(Pa_n|K) = 1$ for each $n$, it follows that

$$\lim_{n \to \infty} Pr(Pa_1, \ldots, Pa_n|K) = 0$$

and thus that $Pr(H|K) = 0$.

Popper’s argument can be attacked in various places. Condition (E) is a form of
exchangeability, and we have seen above that it cannot be expected to hold for all
predicates. But Popper can respond that if (E) does fail then so will various forms of
inductivism (e.g., Hume projectibility). The main place the inductivist will attack is
at the assumption (1) of the independence of instances. Popper’s response is that the
rejection of (1) amounts to the postulation of something like a causal connection
between instances. But this a red herring since the inductivist can postulate a prob-
abilistic dependence among instances without presupposing that the instances are
cemented together by some sort of causal glue.

In another attempt to show that probabilistic methods are ensnared in inconsist-
encies, Popper cites Jeffreys’s proof sketched above that a non-zero prior for $\langle i \rangle Pa_i$
guarantees that

$$\lim_{n \to \infty} Pr(Pa_{n+1}|Pa_1, \ldots, Pa_n, K) = 1.$$ 

But, Popper urges, what is sauce for the goose is sauce for the gander. For we can do
the same for a Goodmanized $P^*$, and from the limit statements we can conclude that
for some $r > 0.5$ there is a sufficiently large $N$ such that for any $N' > N$, the
probabilities for $P_{aq_N}$ and for $P_{aq_{N'}}$ are both greater than $r$, which is a contradiction
for appropriately chosen $P^*$. But the reasoning here is fallacious and there is in
fact no contradiction lurking in Jeffreys’s limit theorem since the convergence is
not supposed to be uniform over different predicates—indeed, Popper’s reasoning
shows that it cannot be.

96 The Confirmation of Scientific Hypotheses
Of course, none of this helps with the difficult questions of which hypotheses should be assigned nonzero priors and how large the priors should be. The example from item 5 in Section 2.8 above suggests that the latter question can be ignored to some extent since the accumulation of evidence tends to swamp differences in priors and force merger of posterior opinion. Some powerful results from advanced probability theory show that such merger takes place in a very general setting (on this matter see Gaifman and Snir 1982).

2. Probabilification vs. inductive support. Popper and Miller (1983) have argued that even if it is conceded that universal hypotheses may have nonzero priors and thus can be probabilified further and further by the accumulation of positive evidence, the increase in probability cannot be equated with genuine inductive support. This contention is based on the application of two lemmas from the probability calculus:

Lemma 1. \( \Pr(\neg H|E,K) \times \Pr(\neg E|K) = \Pr(H \lor \neg E|K) - \Pr(H \lor \neg E|E,K) \). 

Lemma 1 leads easily to

Lemma 2. If \( \Pr(H|E,K) < 1 \) and \( \Pr(E|K) < 1 \) then 

\[ \Pr(H \lor \neg E|E,K) < \Pr(H \lor \neg E|K). \]

Let us apply Lemma 2 to the case discussed above where Bayesianism was used to show that under certain conditions the H-D method does lead to incremental confirmation. Recall that we assumed that 

\[ H, K \vdash E; 1 > \Pr(E|K) > 0; \text{ and } 1 > \Pr(H|K) > 0 \]

and then showed that

\[ \Pr(H|E,K) > \Pr(H|K), \]

which the inductivists want to interpret as saying that \( E \) inductively supports \( H \) on the basis of \( K \). Against this interpretation, Popper and Miller note that \( H \) is logically equivalent to \( (H \lor E). (H \lor \neg E) \). The first conjunct is deductively implied by \( E \), leading Popper and Miller to identify the second conjunct as the part of \( H \) that goes beyond the evidence. But by Lemma 2 this part is countersupported by \( E \), except in the uninteresting case that \( E,K \) makes \( H \) probabilistically certain.

Jeffrey (1984) has objected to the identification of \( H \lor \neg E \) as the part of \( H \) that goes beyond the evidence. To see the basis of his objection, take the case where

\[ H: (i)Pa_i \text{ and } E: Pa_1, \ldots, Pa_n. \]

Intuitively, the part of \( H \) that goes beyond this evidence is \( (i) \{ (i > n) \ldots Pa_i \} \) and not the Popper-Miller \( (i)Pa_i \lor \neg(Pa_1, \ldots, Pa_n) \).

Gillies (1986) restated the Popper-Miller argument using a measure of inductive support based on the incremental model of confirmation: (leaving aside \( K \) ) the support given by \( E \) to \( H \) is \( S(H, E) = \Pr(H|E) - \Pr(H) \). We can then show that

Lemma 3. \( S(H, E) = S(H \lor E, E) + S(H \lor \neg E, E) \).
Gillies suggested that $S(H \vee EE, )$ be identified as the deductive support given $H$ by $E$ and $S(H \vee \neg E, E)$ as the inductive support. And as we have already seen, in the interesting cases the latter is negative. Dunn and Hellman (1986) responded by dualizing. Hypothesis $H$ is logically equivalent to $(H.E) \vee (H.\neg E)$ and $S(H, E) = S(H.E, E) + S(H.\neg E, E)$. Identify the second component as the deductive counterpart. Since this is negative, any positive support must be contributed by the first component which is a measure of the nondeductive support.

3. The problem of old evidence. In the Bayesian identification of the valid kernel of the H-D method we assumed that $Pr(E|K) < 1$, that is, there was some surprise to the evidence $E$. But this is often not the case in important historical examples. When Einstein proposed his general theory of relativity ($H$) at the close of 1915 the anomalous advance of the perihelion of Mercury ($E$) was old news, that is, $Pr(E|K) = 1$. Thus, $Pr(H|E.K) = Pr(H|K)$, and so on the incremental conception of confirmation, Mercury’s perihelion does not confirm Einstein’s theory, a result that flies in the face of the fact that the resolution of the perihelion problem was widely regarded as one of the major triumphs of general relativity. Of course, one could seek to explain the triumph in nonconfirmational terms, but that would be a desperate move.

Garber (1983) and Jeffrey (1983) have suggested that Bayesianism be given a more human face. Actual Bayesian agents are not logically omniscient, and Einstein for all his genius was no exception. When he proposed his general theory he did not initially know that it did in fact resolve the perihelion anomaly, and he had to go through an elaborate derivation to show that it did indeed entail the missing 43° of arc per century. Actual flesh and blood scientists learn not only empirical facts but logicomathematical facts as well, and if we take the new evidence to consist in such facts we can hope to preserve the incremental model of confirmation. To illustrate, let us make the following assumptions about Einstein’s degrees of belief in 1915:

(a) $Pr(H|K) > 0$ (Einstein assigned a nonzero prior to his general theory.)
(b) $Pr(E|K) = 1$ (The perihelion advance was old evidence.)
(c) $Pr(H \vdash E|K) < 1$ (Einstein was not logically omniscient and did not invent his theory so as to guarantee that it entailed the 43°.)
(d) $Pr[(H \vdash E) \vee (H \vdash \neg E)|K] = 1$ (Einstein knew that his theory entailed a definite result for the perihelion motion.)
(e) $Pr[H.(H \vdash \neg E)|K] = Pr[H.(H \vdash \neg E).\neg E|K]$ (Constraint on interpreting $\vdash$ as logical implication.)

From (a)–(e) it can be shown that $Pr[H|(H \vdash E).K]. > Pr(H|K)$. So learning that his theory entailed the happy result served to increase Einstein’s confidence in the theory.

Although the Garber-Jeffrey approach does have the virtue of making Bayesian agents more human and, therefore, more realistic, it avoids the question of whether the perihelion phenomena did in fact confirm the general theory of relativity in favor of focusing on Einstein’s personal psychology. Nor is it adequate to dismiss this
concern with the remark that the personalist form of Bayesianism is concerned precisely with psychology of particular agents, for even if we are concerned principally with Einstein himself, the above calculations seem to miss the mark. We now believe that for Einstein in 1915 the perihelion phenomena provided a strong confirmation of his general theory. And contrary to what the Garber-Jeffrey approach would suggest, we would not change our minds if historians of science discovered a manuscript showing that as Einstein was writing down his field equations he saw in a flash of mathematical insight that \( H \rightarrow E \) or alternatively that he consciously constructed his field equations so as to guarantee that they entailed \( E \). "Did \( E \) confirm \( H \) for Einstein?" and "Did learning that \( H \rightarrow E \) increase Einstein’s confidence in \( H \)" are two distinct questions with possibly different answers. (In addition, the fact that agents are allowed to assign \( Pr(H \rightarrow E|K) < 1 \) means that the Dutch book justification for the probability axioms has to be abandoned. This is anathema for orthodox Bayesian personalists who identify with the betting quotient definition of probability.)

A different approach to the problem of old evidence is to apply the incremental model of confirmation to the counterfactual degrees of belief that would have obtained had \( E \) not been known. Readers are invited to explore the prospects and problems of this approach for themselves. (For further discussion of the problem of old evidence, see Howson 1985, Eells 1985, and van Fraassen 1988.)

2.11 CONCLUSION

The topic of this chapter has been the logic of science. We have been trying to characterize and understand the patterns of inference that are considered legitimate in establishing scientific results—in particular, in providing support for the hypotheses that become part of the corpus of one science or another. We began by examining some extremely simple and basic modes of reasoning—the hypothetico-deductive method, instance confirmation, and induction by enumeration. Certainly (pace Popper) all of them are frequently employed in actual scientific work.

We find—both in contemporary science and in the history of science—that scientists do advance hypotheses from which (with the aid of initial conditions and auxiliary hypotheses) they deduce observational predictions. The test of Einstein’s theory of relativity in terms of the bending of starlight passing close to the sun during a total solar eclipse is an oft-cited example. Others were given in this chapter. Whether the example is as complex as general relativity or as simple as Boyle’s law, the logical problems are the same. Although the H-D method contains a valid kernel—as shown by Bayes’s rule—it must be considered a serious oversimplification of what actually is involved in scientific confirmation. Indeed, Bayes’s rule itself seems to offer a schema far more adequate than the H-D method. But—as we have seen—it, too, is open to serious objections (such as the problem of old evidence).

When we looked at Hempel’s theory of instance confirmation, we discussed an example that has been widely cited in the philosophical literature—namely, the generalization "All ravens are black." If this is a scientific generalization, it is certainly at a low level, but it is not scientifically irrelevant. More complex examples raise the same logical problems. At present, practicing scientists are concerned with—and
excited by—such generalizations as, "All substances having the chemical structure given by the formula YBa$_2$Cu$_3$O$_7$ are superconductors at 70 kelvins." As if indoor ornithology weren't bad enough, we see, by Hempel's analysis, that we can confirm this latter-day generalization by observing black crows. It seems that observations by birdwatchers can confirm hypotheses of solid state physics. (We realize that bird-lovers would disapprove of the kind of test that would need to be performed to establish that a raven is not a superconductor at 70°K.) We have also noted, however, the extreme limitations of the kind of evidence that can be gathered in any such fashion.

Although induction by enumeration is used to establish universal generalizations, its most conspicuous use in contemporary science is connected with statistical generalizations. An early example is found in Rutherford's counting of the frequencies with which alpha particles bombarding a gold foil were scattered backward (more or less in the direction from which they came). The counting of instances led to a statistical hypothesis attributing stable frequencies to such events. A more recent example—employing highly sophisticated experiments—involves the detection of neutrinos emitted by the sun. Physicists are puzzled by the fact that they are detecting a much smaller frequency than current theory predicts. (Obviously probabilities of the type characterized as frequencies are involved in examples of the sort mentioned here.) In each of these cases an inductive extrapolation is drawn from observed frequencies. In our examination of induction by enumeration, however, we have found that it is plagued by Hume's old riddle and Goodman's new one.

One development of overwhelming importance in twentieth-century philosophy of science has been the widespread questioning of whether there is any such thing as a logic of science. Thomas Kuhn's influential work, The Structure of Scientific Revolutions (1962, 1970), asserted that the choice of scientific theories (or hypotheses) involves factors that go beyond observation and logic—including judgement, persuasion, and various psychological and sociological influences. There is, however, a strong possibility that, when he wrote about going beyond the bounds of observation and logic, the kind of logic he had in mind was the highly inadequate H-D schema, (see Salmon 1989 for an extended discussion of this question, and for an analysis of Kuhn's views in the light of Bayes's rule). The issues raised by the Kuhnian approach to philosophy of science are discussed at length in Chapter 4 of this book.

Among the problems we have discussed there are—obviously—many to which we do not have adequate solutions. Profound philosophical difficulties remain. But the deep and extensive work done by twentieth-century philosophers of science in these areas has cast a good deal of light on the nature of the problems. It is an area in which important research is currently going on and in which significant new results are to be expected.

**DISCUSSION QUESTIONS**

1. Select a science with which you are familiar and find a case in which a hypothesis or theory is taken to be confirmed by some item of evidence. Try to characterize the relationship between the
evidence and hypothesis or theory confirmed in terms of the schemas discussed here. If none of them is applicable, can you find a new schema that is?

2. If the prior probability of every universal hypothesis is zero how would you have to rate the probability of the statement that unicorns (at least one) exist? Explain your answer.

3. Show that accepting the combination of the entailment condition, the special consequence condition, and the converse consequence condition (see Section 2.4) entails that any E confirms any H.

4. Consider a population that consists of all of the adult population of some particular district. We want to test the hypothesis that all voters are literate,

\[(x) (Vx \supset Lx),\]

which is, of course, equivalent to

\[(x) (\sim Lx \supset \sim Vx).\]

Suppose that approximately 75 percent of the population are literate voters, approximately 15 percent are literate nonvoters, approximately 5 percent are illiterate nonvoters, and approximately 5 percent are illiterate voters—but this does not preclude the possibility that no voters are illiterate. Would it be best to sample the class of voters or the class of illiterate people? Explain your answer. (This example is given in Suppes 1966, 201.)

5. Goodman's examples challenge the idea that hypotheses are confirmed by their instances. Goodman holds that the distinction between those hypotheses that are and those that are not projectable on the basis of their instances is to be drawn in terms of 'entrenchment.' Predicates become entrenched as antecedents or consequents by playing those roles in universal conditionals that are actually projected. Call a hypothesis admissible just in case it has some positive instances, no negative instances, and is not exhausted. Say that H overriding H' just in case H and H' conflict, H is admissible and is better entrenched than H' (i.e., has a better entrenched antecedent and equally well entrenched consequent or vice versa), and H is not in conflict with some still better entrenched admissible hypothesis. Critically discuss the idea that H is projectable on the basis of its positive instances just in case it is admissible but not overridden.

6. Show that

\[H: (x) (\exists y) Rxy.(x) \sim Rxx.(x) (y) (z) [(Rxy.Ryz) \supset Rxz]\]

cannot be Hempel-confirmed by any consistent E.

7. It is often assumed in philosophy of science that if one is going to represent numerically the degree to which evidence E supports hypothesis H with respect to background B, then the numbers so produced — \(P(H|E,B)\) — must obey the probability calculus. What are the prospects of alternative calculi? (Hint: Consider each of the axioms in turn and ask under what circumstances each axiom could be violated in the context of a confirmation theory. What alternative axiom might you choose?)

8. If Bayes's rule is taken as a schema for confirmation of scientific hypotheses, it is necessary to decide on an interpretation of probability that is suitable for that context. It is especially crucial to think about how the prior probabilities are to be understood. Discuss this problem in the light of the admissible interpretations offered in this chapter.

9. William Tell gave his young cousin Wesley a two-week intensive archery course. At its completion, William tested Wes's skill by asking him to shoot arrows at a round target, ten feet in radius with a centered bull's-eye, five feet in radius.

"You have learned no control at all," scolded William after the test. "Of those arrows that hit the target, five are within five feet of dead center and five more between five and ten feet from dead center." "Not so," replied Wes, who had been distracted from archery practice by...
his newfound love of geometry. "That five out of ten arrows on the target hit the bull's-eye shows I do have control. The bullseye is only one quarter the total area of the target."

Adjudicate this dispute in the light of the issues raised in the chapter. Note that an alternative form of Bayes's rule which applies when one considers the relative confirmation accrued by two hypotheses $H_1$ and $H_2$ by evidence $E$ with respect to background $B$ is:

$$\frac{Pr(H_1|E,B)}{Pr(H_2|E,B)} = \frac{Pr(E|H_1,B)}{Pr(E|H_2,B)} \cdot \frac{Pr(H_1|B)}{Pr(H_2|B)}$$

10. Let $\{H_1, H_2, \ldots, H_n\}$ be a set of competing hypotheses. Say that $E$ selectively Hempel-confirms some $H_j$ just in case it Hempel-confirms $H_j$ but fails to confirm the alternative $H_k$. Use this notion of selective confirmation to discuss the relative confirmatory powers of black ravens versus nonblack nonravens for alternative hypotheses about the color of ravens.

11. Prove Lemmas 1, 2, and 3 of Section 2.10.

12. Discuss the prospects of resolving the problem of old evidence by using counterfactual degrees of belief, that is, the degrees of belief that would have obtained had the evidence $E$ not been known.

13. Work out the details of the following example, which was mentioned in Section 2.8. There is a square piece of metal in a closed box. You cannot see it. But you are told that its area is somewhere between 1 square inch and 4 square inches. Show how the use of the principle of indifference can lead to conflicting probability values.

14. Suppose there is a chest with two drawers. In each drawer are two coins; one drawer contains two gold coins, the other contains one gold coin and one silver coin. A coin will be drawn from one of these drawers. Suppose, further, that you know (without appealing to the principle of indifference) that each drawer has an equal chance of being chosen for the draw, and that, within each drawer, each coin has an equal chance of being chosen. When the coin is drawn it turns out to be gold. What is the probability that the other coin in the same drawer is gold? Explain how you arrived at your answer.

15. Discuss the problem of ascertaining limits of relative frequencies on the basis of observed frequencies in initial sections of sequences of events. This topic is especially suitable for those who have studied David Hume's problem regarding the justification of inductive inference in Part II of this chapter.

16. When scientists are considering new hypotheses they often appeal to plausibility arguments. As a possible justification for this procedure, it has been suggested that plausibility arguments are attempts at establishing prior probabilities. Discuss this suggestion, using concrete illustrations from the history of science or contemporary science.

17. Analyze the bootstrap confirmation of the perfect gas law in such a way that no "macho" bootstrapping is used, that is, the gas law itself is not used as an auxiliary to deduce instances of itself.

SUGGESTED READINGS


HEMPEL, CARL G. (1945), "Studies in the Logic of Confirmation," *Mind* 54: 1–26, 97–121. Reprinted in Hempel (1965, see Bibliography), with a 1964 Postscript added. This classic essay contains Hempel’s analysis of the Nicod criterion of confirmation, and it presents Hempel’s famous paradox of the ravens, along with his analysis of it.


HUME, DAVID (1748), *An Enquiry Concerning Human Understanding*. Many editions available. This is the philosophical classic on the problem of induction, and it is highly readable. Sections 4–7 deal with induction, causality, probability, necessary connection, and the uniformity of nature.

