Carl Hempel (from *Philosophy of Natural Science*)

SCIENTIFIC INQUIRY: INVENTION AND TEST

2.1 A case history as an example

As a simple illustration of some important aspects of scientific inquiry let us consider Semmelweis' work on childbed fever. Ignaz Semmelweis, a physician of Hungarian birth, did this work during the years from 1844 to 1848 at the Vienna General Hospital. As a member of the medical staff of the First Maternity Division in the hospital, Semmelweis was distressed to find that a large proportion of the women who were delivered of their babies in that division contracted a serious and often fatal illness known as puerperal fever or childbed fever. In 1844, as many as 260 out of 3,157 mothers in the First Division, or 8.2 per cent, died of the disease; for 1845, the death rate was 6.8 per cent, and for 1846, it was 11.4 per cent. These figures were all the more alarming because in the adjacent Second Maternity Division of the same hospital, which accommodated almost as many women as the First, the death toll from childbed fever was much lower: 2.3, 2.0, and 2.7 per cent for the same years. In a book that he wrote later on the causation and the prevention of childbed fever, Semmelweis describes his efforts to resolve the dreadful puzzle.[1]

He began by considering various explanations that were current at the time; some of these he rejected out of hand as incompatible with well-established facts; others he subjected to specific tests.

One widely accepted view attributed the ravages of puerperal fever to "epidemic influences", which were vaguely described as "atmospheric-cosmic-telluric changes" spreading over whole districts and causing childbed fever in women in confinement. But how, Semmelweis reasons, could such influences have plagued the First Division for years and yet spared the Second? And how could this view be reconciled with the fact that while the fever was raging in the hospital, hardly a case occurred in the city of Vienna or in its surroundings: a genuine epidemic, such as cholera, would not be so selective. Finally, Semmelweis notes that some of the women admitted to the First Division, living far from the hospital, had been overcome by labor on their way and had given birth in the street: yet despite these adverse conditions, the death rate from childbed fever among these cases of "street birth" was lower than the average for the First Division.

On another view, overcrowding was a cause of mortality in the First Division. But Semmelweis points out that in fact the crowding was heavier in the Second Division, partly as a result of the desperate efforts of patients to avoid assignment to the notorious First Division. He also rejects two similar conjectures that were current, by noting that there were no differences between the two Divisions in regard to diet or general care of the patients.

In 1846, a commission that had been appointed to investigate the matter attributed the prevalence of illness in the First Division to injuries resulting from rough examination by the medical students, all of whom received their obstetrical training in the First

Division. Semmelweis notes in refutation of this view that (a) the injuries resulting naturally from the process of birth are much more extensive than those that might be caused by rough examination; (b) the midwives who received their training in the Second Division examined their patients in much the same manner but without the same ill effects; (c) when, in response to the commission's report, the number of medical students was halved and their examinations of the women were reduced to a minimum, the mortality, after a brief decline, rose to higher levels than ever before.

 Various psychological explanations were attempted. One of them noted that the First Division was so arranged that a priest bearing the last sacrament to a dying woman had to pass through five wards before reaching the sickroom beyond: the appearance of the priest, preceded by an attendant ringing a bell, was held to have a terrifying and debilitating effect upon the patients in the wards and thus to make them more likely victims of childbed fever. In the Second Division, this adverse factor was absent, since the priest had direct access to the sickroom. Semmelweis decided to test this conjecture. He persuaded the priest to come by a roundabout route and without ringing of the bell, in order to reach the sick chamber silently and unobserved. But the mortality in the First Division did not decrease.

A new idea was suggested to Semmelweis by the observation that in the First Division the women were delivered lying on their backs; in the Second Division, on their sides. Though he thought it unlikely, he decided "like a drowning man clutching at a straw", to test whether this difference in procedure was significant. He introduced the use of the lateral position in the First Division, but again, the mortality remained unaffected.

At last, early in 1847, an accident gave Semmelweis the decisive clue for his solution of the problem. A colleague of his, Kolletschka, received a puncture wound in the finger, from the scalpel of a student with whom he was performing an autopsy, and died after an agonizing illness during which he displayed the same symptoms that Semmelweis had observed in the victims of childbed fever. Although the role of micro-organisms in such infections had not yet been recognized at the time, Semmelweis realized that "cadaveric matter" which the student's scalpel had introduced into Kolletschka's blood stream had caused his colleague's fatal illness. And the similarities between the course of Kolletschka's disease and that of the women in his clinic led Semmelweis to the conclusion that his patients had died of the same kind of blood poisoning: he, his colleagues, and the medical students had been the carriers of the infectious material, for he and his associates used to come to the wards directly from performing dissections in the autopsy room, and examine the women in labor after only superficially washing their hands, which often retained a characteristic foul odor.

Again, Semmelweis put his idea to a test. He reasoned that if he were right, then childbed fever could be prevented by chemically destroying the infectious material adhering to the hands. He therefore issued an order requiring all medical students to wash their hands in a solution of chlorinated lime before making an examination. The mortality from childbed fever promptly began to decrease, and for the year 1848 it fell to

1.27 per cent in the First Division, compared to 1.33 in the Second.

In further support of his idea, or of his *hypothesis,* as we will also say, Semmelweis notes that it accounts for the fact that the mortality in the Second Division consistently was so much lower: the patients there were attended by midwives, whose training did not include anatomical instruction by dissection of cadavers.

The hypothesis also explained the lower mortality among "street births": women who arrived with babies in arms were rarely examined after admission and thus had a better chance of escaping infection. Similarly, the hypothesis accounted for the fact that the victims of childbed fever among the newborn babies were all among those whose mothers had contracted the disease during labor; for then the infection could be transmitted to the baby before birth, through the common bloodstream of mother and child, whereas this was impossible when the mother remained healthy.

Further clinical experiences soon led Semmelweis to broaden his hypothesis. On one occasion, for example, he and his associates, having carefully disinfected their hands, examined first a woman in labor who was suffering from a festering cervical cancer; then they proceeded to examine twelve other women in the same room, after only routine washing without renewed disinfection. Eleven of the twelve patients died of puerperal fever. Semmelweis concluded that childbed fever can be caused not only by cadaveric material, but also by "putrid matter derived from living organisms."

2.2 Basic steps in testing a hypothesis

We have seen how, in his search for the cause of childbed fever, Semmelweis examined various hypotheses that had been suggested as possible answers. How such hypotheses are arrived at in the first place is an intriguing question which we will consider later. First, however, let us examine how a hypothesis, once proposed, is tested.

Sometimes, the procedure is quite direct. Consider the conjectures that differences in crowding, or in diet, or in general care account for the difference in mortality between the two divisions. As Semmelweis points out, these conflict with readily observable facts. There are no such differences between the divisions; the hypotheses are therefore rejected as false.

But usually the test will be less simple and straightforward. Take the hypothesis attributing the high mortality in the First Division to the dread evoked by the appearance of the priest with his attendant. The intensity of that dread, and especially its effect upon childbed fever, are not as directly ascertainable as are differences in crowding or in diet, and Semmelweis uses an indirect method of testing. He asks himself: Are there any readily observable effects that should occur if the hypothesis were true? And he reasons: *If* the hypothesis were true, *then* an appropriate change in the priest's procedure should be followed by a decline in fatalities. He checks this implication by a simple experiment and finds it false, and he therefore rejects the hypothesis.

Similarly, to test his conjecture about the position of the women during delivery, he

reasons: *If* this conjecture should be true, *then* adoption of the lateral position in the First Division will reduce the mortality. Again, the implication is shown false by his experiment, and the conjecture is discarded.

In the last two cases, the test is based on an argument to the effect that *if* the contemplated hypothesis, say H, is true, *then* certain observable events (e.g., decline in mortality) should occur under specified circumstances (e.g., if the priest refrains from walking through the wards, or if the women are delivered in lateral position); or briefly, if H is true, then so is *I,* where I is a statement describing the observable occurrences to be expected. For convenience, let us say that *I* is inferred from, or implied by, H; and let us call I a test *implication of the hypothesis H.* (We will later give a more accurate description of the relation between *I* and H.)

In our last two examples, experiments show the test implication to be false, and the hypothesis is accordingly rejected. The reasoning that leads to the rejection may be schematized as follows:

2a)   If H is true, then so is I

But (as the evidence shows) *I is* not true.

\-

H is not true.

Any argument of this form, called *modus tollens* in logic,[2] is deductively valid; that is, if its premisses (the sentences above the horizontal line) are true, then its conclusion (the sentence below the horizontal line) is unfailingly true as well. Hence, if the premisses of (2a) are properly established, the hypothesis H that is being tested must indeed be rejected.

Next, let us consider the case where observation or experiment bears out the test implication *I.* From his hypothesis that childbed fever is blood poisoning produced by cadaveric matter, Semmelweis infers that suitable antiseptic measures will reduce fatalities from the disease. This time, experiment shows the test implication to be true. But this favorable outcome does not conclusively prove the hypothesis true, for the underlying argument would have the form

2b)   If *H is* true, then so is I.

(As the evidence shows) I is true.

\-

H is true.

And this mode of reasoning, which is referred to as the *fallacy of affirming the consequent,* is deductively invalid, that is, its conclusion may be false even if its premisses are true.[3] This is in fact illustrated by Semmelweis' own experience. The initial version of his account of childbed fever as a form of blood poisoning presented infection with cadaveric matter essentially as the one and only source of the disease; and he was right in reasoning that if this hypothesis should be true, then destruction of cadaveric particles by antiseptic washing should reduce the mortality. Furthermore, his experiment did show the test implication to be true. Hence, in this case, the premisses of (26) were both true. Yet, his hypothesis was false, for as he later discovered, putrid material from living organisms, too, could produce childbed fever.

Thus, the favorable outcome of a test, i.e., the fact that a test implication inferred from a hypothesis is found to be true, does not prove the hypothesis to be true. Even if many implications of a hypothesis have been borne out by careful tests, the hypothesis may still be false. The following argument still commits the fallacy of affirming the consequent:

2c)   If H is true, then so are $I_1, I_2, ..., I_n$

(As the evidence shows) $I_1, I_2, ..., I_n$ are all true.

H is true.

This, too, can be illustrated by reference to Semmelweis' final hypothesis in its first version. As we noted earlier, his hypothesis also yields the test implications that among cases of street births admitted to the First Division, mortality from puerperal fever should be below the average for the Division, and that infants of mothers who escape the illness do not contract childbed fever; and these implications, too, were borne out by the evidence — even though the first version of the final hypothesis was false.

But the observation that a favorable outcome of however many tests does not afford conclusive proof for a hypothesis should not lead us to think that if we have subjected a hypothesis to a number of tests and all of them have had a favorable outcome, we are no better off than if we had not tested the hypothesis at all. For each of our tests might conceivably have had an unfavorable outcome and might have led to the rejection of the hypothesis. A set of favorable results obtained by testing different test implications, Ii,^,...,!,,, of a hypothesis, shows that as far as these particular implications are concerned, the hypothesis has been borne out; and while this result does not afford a complete proof of the hypothesis, it provides at least some support, some partial corroboration or confirmation for it. The extent of this support will depend on various aspects of the hypothesis and of the test data. These will be examined in Chapter 4.

Let us now consider another example,[4] which will also bring to our attention some

further aspects of scientific inquiry.

As was known at Galileo's time, and probably much earlier, a simple suction pump, which draws water from a well by means of a piston that can be raised in the pump barrel, will lift water no higher than about 34 feet above the surface of the well. Galileo was intrigued by this limitation and suggested an explanation for it, which was, however, unsound. After Galileo's death, his pupil Torricelli advanced a new answer. He argued that the earth is surrounded by a sea of air, which, by reason of its weight exerts pressure upon the surface below, and that this pressure upon the surface of the well forces water up the pump barrel when the piston is raised. The maximum length of 34 feet for the water column in the barrel thus reflects simply the total pressure of the atmosphere upon the surface of the well.

It is evidently impossible to determine by direct inspection or observation whether this account is correct, and Torricelli tested it indirectly. He reasoned that *if* his conjecture were true, *then* the pressure of the atmosphere should also be capable of supporting a proportionately shorter column of mercury; indeed, since the specific gravity of mercury is about 14 times that of water, the length of the mercury column should be about 34/14 feet, or slightly less than 2 1/2 feet. He checked this test implication by means of an ingeniously simple device, which was, in effect, the mercury barometer. The well of water is replaced by an open vessel containing mercury; the barrel of the suction pump is replaced by a glass tube sealed off at one end. The tube is completely filled with mercury and closed by placing the thumb tightly over the open end. It is then inverted, the open end is submerged in the mercury well, and the thumb is withdrawn; whereupon the mercury column in the tube drops until its length is about 30 inches—just as predicted by Torricelli's hypothesis.

A further test implication of that hypothesis was noted by Pascal, who reasoned that if the mercury in Torricelli's barometer is counterbalanced by the pressure of the air above the open mercury well, then its length should decrease with increasing altitude, since the weight of the air overhead becomes smaller. At Pascal's request, this implication was checked by his brother-in-law, Perier, who measured the length of the mercury column in the Torricelli barometer at the foot of the Puy-de-Dome, a mountain some 4,800 feet high, and then carefully carried the apparatus to the top and repeated the measurement there while a control barometer was left at the bottom under the supervision of an assistant. Perier found the mercury column at the top of the mountain more than three inches shorter than at the bottom, whereas the length of the column in the control barometer had remained unchanged throughout the day.

2.3 The role of induction in scientific inquiry

We have considered some scientific investigations in which a problem was tackled by proposing tentative answers in the form of hypotheses that were then tested by deriving from them suitable test implications and checking these by observation or experiment.

But how are suitable hypotheses arrived at in the first place? It is sometimes held that they are inferred from antecedently collected data by means of a procedure called *inductive inference, as* contradistinguished from deductive inference, from which it differs in important respects.

In a deductively valid argument, the conclusion is related to the premisses in such a way that if the premisses are true then the conclusion cannot fail to be true as well. This requirement is satisfied, for example, by any argument of the following general form:

> If *p,* then *q.*
>
> It is not the case that *q.*
>
> -----------------------------
>
> It is not the case that *p.*

Brief reflection shows that no matter what particular statements may stand at the places marked by the letters 'p' and *'q',* the conclusion will certainly be true if the premisses are. In fact, our schema represents the argument form called *modus tollens,* to which we referred earlier.

Another type of deductively valid inference is illustrated by this example:

Any sodium salt, when put into the flame of a Bunsen burner, turns the flame yellow.

This piece of rock salt is a sodium salt.

This piece of rock salt, when put into the flame of a Bunsen burner, will turn the flame yellow.

Arguments of the latter kind are often said to lead from the general (here, the premiss about all sodium salts) to the particular (a conclusion about the particular piece of rock salt). Inductive inferences, by contrast, are sometimes described as leading from premises about particular cases to a conclusion that has the character of a general law or principle. For example, from premises to the effect that each of the particular samples of various sodium salts that have so far been subjected to the Bunsen flame test did turn the flame yellow, inductive inference supposedly leads to the general conclusion that all sodium salts, when put into the flame of a Bunsen burner, turn the flame yellow. But in this case, the truth of the premisses obviously does *not* guarantee the truth of the conclusion; for even if it is the case that all samples of sodium salts examined so far did turn the Bunsen flame yellow, it remains quite possible that new kinds of sodium salt might yet be found that do not conform to this generalization. Indeed, even some kinds of sodium salt that have already been tested with positive result might conceivably fail to satisfy the generalization under special physical con- ditions (such as very strong magnetic fields or the like) in which they have not yet been examined. For this reason, the premisses of an inductive inference are often said to imply the conclusion only with more or less high probability, whereas the premisses of a

deductive inference imply the conclusion with certainty.

The idea that in scientific inquiry, inductive inference from antecedently collected data leads to appropriate general principles is clearly embodied in the following account of how a scientist would ideally proceed: ,   :

> If we try to imagine how a mind of superhuman power and reach, but normal so far as the logical processes of its thought are concerned, . . . would use the scientific method, the process would be as follows:
>
> First, all facts would be observed and recorded, *without selection* or a *priori* guess as to their relative importance. Secondly, the observed and recorded facts would be analyzed, compared, and classified, *without hypothesis or postulates* other than those necessarily involved in the logic of thought. Third, from this analysis of the facts generalizations would be inductively drawn as to the relations, classificatory or causal, between them. Fourth, further research would be deductive as well as inductive, employing inferences from previously established generalizations.[5]

This passage distinguishes four stages in an ideal scientific inquiry:

(1) observation and recording of all facts, (2) analysis and classification of these facts, (3) inductive derivation of generalizations from them, and (4) further testing of the generalizations. The first two of these stages are specifically assumed not to make use of any guesses or hypotheses as to how the observed facts might be interconnected; this restriction seems to have been imposed in the belief that such preconceived ideas would introduce a bias and would jeopardize the scientific objectivity of the investigation.

But the view expressed in the quoted passage—I will call it *the narrow inductivist conception of scientific inquiry*—is untenable, for several reasons. A brief survey of these can serve to amplify and to supplement our earlier remarks on scientific procedure.

First, a scientific investigation as here envisaged could never get off the ground. Even its first phase could never be carried out, for a collection of *all* the facts would have to await the end of the world, so to speak; and even all the facts *up to now* cannot be collected, since there are an infinite number and variety of them. Are we to examine, for example, all the grains of sand in all the deserts and on all the beaches, and are we to record their shapes, their weights, their chemical composition, their distances from each other, their constantly changing temperature, and their equally changing distance from the center of the moon? Are we to record the floating thoughts that cross our minds in the tedious process? The shapes of the clouds overhead, the changing color of the sky? The construction and the trade name of our writing equipment? Our own life histories and those of our fellow investigators? All these, and untold other things, are, after all, among "all the facts up to now".

Perhaps, then, all that should be required in the first phase is that all the *relevant* facts be collected. But relevant to what? Though the author does not mention this, let us suppose that the inquiry is concerned with a specified *problem.* Should we not then begin by collecting all the facts—or better, all available data—relevant to that problem? This notion still makes no clear sense. Semmelweis sought to solve one specific problem, yet he collected quite different kinds of data at different stages of his inquiry. And rightly so; for what particular sorts of data it is reasonable to collect is not determined by the problem under study, but by a tentative answer to it that the investigator entertains in the form of a conjecture or hypothesis. Given the conjecture that mortality from childbed fever was increased by the terrifying appearance of the priest and his attendant with the death bell, it was relevant to collect data on the consequences of having the priest change his routine; but it would have been totally irrelevant to check what would happen if doctors and students disinfected their hands before examining their patients. With respect to Semmelweis' eventual contamination hypothesis, data of the latter kind were clearly relevant, and those of the former kind totally irrelevant.

Empirical "facts" or findings, therefore, can be qualified as logically relevant or irrelevant only in reference to a given hypothesis, but not **in** reference to a given problem.

Suppose now that a hypothesis H has been advanced as a tentative answer to a research problem: what kinds of data would be relevant to *H?* Our earlier examples suggest an answer: A finding is relevant to H if either its occurrence or its nonoccurrence can be inferred from H. Take Torricelli's hypothesis, for example. As we saw, Pascal inferred from it that the mercury column in a barometer should grow shorter if the barometer were carried up a mountain. Therefore, any finding to the effect that this did indeed happen in a particular case is relevant to the hypotheses; but so would be the finding that the length of the mercury column had remained unchanged or that it had decreased and then increased during the ascent, for such findings would refute Pascal's test implication and would thus disconfirm Torricelli's hypothesis. Data of the former kind may be called positively, or favorably, relevant to the hypothesis; those of the latter kind negatively, or unfavorably, relevant.

In sum, the maxim that data should be gathered without guidance by antecedent hypotheses about the connections among the facts under study is self-defeating, and it is certainly not followed in scientific inquiry. On the contrary, tentative hypotheses are needed to give direction to a scientific investigation. Such hypotheses determine, among other things, what data should be collected at a given point in a scientific investigation.

It is of interest to note that social scientists trying to check a hypothesis by reference to the vast store of facts recorded by the U.S. Bureau of the Census, or by other data-gathering organizations, sometimes find to their disappointment that the

values of some variable that plays a central role in the hypothesis have nowhere been systematically recorded. This remark is not, of course, intended as a criticism of data gathering: those engaged in the process no doubt try to select facts that might prove relevant to future hypotheses; the observation is simply meant to illustrate the impossibility of collecting "all the relevant data" without knowledge of the hypotheses to which the data are to have relevance.

The second stage envisaged in our quoted passage is open to similar criticism. A set of empirical "facts" can be analyzed and classified in many different ways, most of which will be un-illuminating for the purposes of a given inquiry. Semmelweis could have classified the women in the maternity wards according to criteria such as age, place of residence, marital status, dietary habits, and so forth; but information on these would have provided no clue to a patient's prospects of becoming a victim of childbed fever. What Semmelweis sought were criteria that would be significantly connected with those prospects; and for this purpose, as he eventually found, it was illuminating to single out those women who were attended by medical personnel with contaminated hands; for it was with this characteristic, or with the corresponding class of patients, that high mortality from childbed fever was associated.

Thus, if a particular way of analyzing and classifying empirical findings is to lead to an explanation of the phenomena concerned, then it must be based on hypotheses about how those phenomena are connected; without such hypotheses, analysis and classification are blind.

Our critical reflections on the first two stages of inquiry as envisaged in the quoted passage also undercut the notion that hypotheses are introduced only in the third stage, by inductive inference from antecedently collected data. But some further remarks on the subject should be added here. Induction is sometimes conceived as a method that leads, by means of mechanically applicable rules, from observed facts to corresponding general principles. In this case, the rules of inductive inference would provide effective canons of scientific discovery; induction would be a mechanical procedure analogous to the familiar routine for the multiplication of integers, which leads, in a finite number of predetermined and mechanically performable steps, to the corresponding product.

Actually, however, no such general and mechanical induction procedure is available at present; otherwise, the much studied problem of the causation of cancer, for example, would hardly have remained unsolved to this day. Nor can the discovery of such a procedure ever be expected. For—to mention one reason—scientific hypotheses and theories are usually couched in terms that do not occur at all in the description of the empirical findings on which they rest, and which they serve to explain. For example, theories about the atomic and subatomic structure of matter contain terms such as 'atom', 'electron', 'proton', 'neutron', 'psi-function', etc.; yet they are based on laboratory findings about the spectra of various gases, tracks in cloud and bubble chambers, quantitative aspects of chemical reactions, and so forth—all of which can be described

without the use of those "theoretical terms". Induction rules of the kind here envisaged would therefore have to provide a mechanical routine for constructing, on the basis of the given data, a hypothesis or theory stated in terms of some quite novel concepts, which are nowhere used in the description of the data themselves. Surely, no general mechanical rule of procedure can be expected to achieve this. Could there be a general rule, for example, which, when applied to the data available to Galileo concerning the limited effectiveness of suction pumps, would, by a mechanical routine, produce a hypothesis based on the concept of a sea of air?

To be sure, mechanical procedures for inductively "inferring" a hypothesis on the basis of given data may be specifiable for situations of special, and relatively simple, kinds. For example, if the length of a copper rod has been measured at several different temperatures, the resulting pairs of associated values for temperature and length may be represented by points in a plane coordinate system, and a curve may be drawn through them in accordance with some particular rule of curve fitting. The curve then graphically represents a general quantitative hypothesis that expresses the length of the rod as a specific function of its temperature. But note that this hypothesis contains no novel terms; it is expressible in terms of the concepts of temperature and length, which are used also in describing the data. Moreover, the choice of "associated" values of temperature and length as data already presupposes a guiding hypothesis; namely, that with each value of the temperature, exactly one value of the length of the copper rod is associated, so that its length is indeed a function of its temperature alone. The mechanical curve-fitting routine then serves only to select a particular function as the appropriate one. This point is important; for suppose that instead of a copper rod, we examine a body of nitrogen gas enclosed in a cylindrical container with a movable piston as a lid, and that we measure its volume at several different temperatures. If we were to use this procedure in an effort to obtain from our data a *general* hypothesis representing the volume of the gas as a function of its temperature, we would fail, because the volume of a gas is a function both of its temperature and of the pressure exerted upon it, so that at the same temperature, the given gas may assume different volumes.

Thus, even in these simple cases, the mechanical procedures for the construction of a hypothesis do only part of the job, for they presuppose an antecedent, less specific hypothesis (i.e., that a certain physical variable is a function of one single other variable), which is not obtainable by the same procedure.

There are, then, no generally applicable "rules of induction", by which hypotheses or theories can be mechanically derived or inferred from empirical data. The transition from data to theory requires creative imagination. Scientific hypotheses and theories are not *derived* from observed facts, but *invented* in order to account for them. They constitute guesses at the connections that might obtain between the phenomena under study, at uniformities and patterns that might underlie their occurrence. "Happy guesses" [6] of this kind require great ingenuity, especially if they involve a radical departure from

current modes of scientific thinking, as did, for example, the theory of relativity and quantum theory. The inventive effort required in scientific research will benefit from a thorough familiarity with current knowledge in the field. A complete novice will hardly make an important scientific discovery, for the ideas that may occur to him are likely to duplicate what has been tried before or to run afoul of well-established facts or theories of which he is not aware.

Nevertheless, the ways in which fruitful scientific guesses are arrived at are very different from any process of systematic inference. The chemist Kekule, for example, tells us that he had long been trying unsuccessfully to devise a structural formula for the benzene molecule when, one evening in 1865, he found a solution to his problem while he was dozing in front of his fireplace. Gazing into the flames, he seemed to see atoms dancing in snakelike arrays. Suddenly, one of the snakes formed a ring by seizing hold of its own tail and then whirled mockingly before him. Kekule awoke in a flash: he had hit upon the now famous and familiar idea of representing the molecular structure of benzene by a hexagonal ring. He spent the rest of the night working out the conse-quences of this hypothesis.[7]

This last remark contains an important reminder concerning the objectivity of science. In his endeavor to find a solution to his problem, the scientist may give free rein to his imagination, and the course of his creative thinking may be influenced even by scientifically questionable notions. Kepler's study of planetary motion, for example, was inspired by his interest in a mystical doctrine about numbers and a passion to demonstrate the music of the spheres. Yet, scientific objectivity is safeguarded by the principle that while hypotheses and theories may be freely invented and *proposed* in science, they can be *accepted* into the body of scientific knowledge only if they pass critical scrutiny, which includes in particular the checking of suitable test implications by careful observation or experiment.

Interestingly, imagination and free invention play a similarly important role in those disciplines whose results are validated exclusively by deductive reasoning; for example, in mathematics. For the rules of deductive inference do not afford mechanical rules of discovery, either. As illustrated by our statement of *modus tollens* above, those rules are usually expressed in the form of general schemata, any instance of which is a deductively valid argument. If premises of the specified kind are given, such a schema does indeed specify a way of proceeding to a logical consequence. But for any set of premises that may be given, the rules of deductive inference specify an infinity of validly deducible conclusions. Take, for example, one simple rule represented by the following schema:

$\underline{p}$

$p \lor q$

It tells us, in effect, that from the proposition that *p* is the case, it follows that *p or q is* the case, where *p* and *q* may be any propositions whatever. The word 'or' is here

understood in the "nonexclusive" sense, so that *'p* or *q' is* tantamount to 'either *p* or *q* or both *p* and *q'.* Clearly, if the premiss of an argument of this type is true, then so must be the conclusion; hence, any argument of the specified form is valid. But this one-rule alone entitles us to infer infinitely many different consequences from any one premiss. Thus, from 'the Moon has no atmosphere', it authorizes us to infer any statement of the form 'The Moon has no atmosphere, or *q',* where for *'q'* we may write any statement whatsoever, no matter whether it is true or false; for example, 'the Moon's atmosphere is very thin', 'the Moon is uninhabited', 'gold is denser than silver', 'silver is denser than gold', and so forth. (It is interesting and not difficult to prove that infinitely many different statements can be formed in English; each of these may be put in the place of the variable *'(q'.*) Other rules of deductive inference add, of course, to the variety of state-ments derivable from one premiss or set of premisses. Hence, if we are given a set of statements as premisses, the rules of deduction give no direction to our inferential procedures. They do not single out one statement as "the" conclusion to be derived from our premisses, nor do they tell us how to obtain interesting or systematically important conclusions; they provide no mechanical routine, for example, for deriving significant mathematical theorems from given postulates. The discovery of important, fruitful mathematical theorems, like the discovery of important, fruitful theories in empirical science, requires inventive ingenuity; it calls for imaginative, insightful guessing. But again, the interests of scientific objectivity are safeguarded by the demand for an *objective validation* of such conjectures. In mathematics, this means *proof* by deductive derivation from axioms. And when a mathematical proposition has been proposed as a conjecture, its proof or disproof still requires inventiveness and ingenuity, often of a very high caliber; for the rules of deductive inference do not even provide a general mechanical procedure for constructing proofs or disproofs. Their systematic role is rather the modest one of serving as *criteria of soundness for arguments* offered as proofs: an argument will constitute a valid mathematical proof if it proceeds from the axioms to the proposed theorem by a chain of inferential steps each of which is valid according to one of the rules of deductive inference. And to check whether a given argument is a valid proof in this sense is indeed a purely mechanical task.

Scientific knowledge, as we have seen, is not arrived at by applying some inductive inference procedure to antecedently collected data, but rather by what is often called "the method of hypothesis", i.e. by inventing hypotheses as tentative answers to a problem under study, and then subjecting these to empirical test. It will be part of such test to see whether the hypothesis is borne out by whatever relevant findings may have been gathered before its formulation; an acceptable hypothesis will have to fit the available relevant data. Another part of the test will consist in deriving new test implications from the hypothesis and checking these by suitable observations or experiments. As we noted earlier, even extensive testing with entirely favorable results does not establish a hypothesis conclusively, but provides only more or less strong support for it. Hence, while scientific inquiry is certainly not inductive in the narrow

sense we have examined in some detail, it may be said to be *inductive in a wider sense,* inasmuch as it involves the acceptance of hypotheses on the basis of data that afford no deductively conclusive evidence for it, but lend it only more or less strong "inductive support", or confirmation. And any "rules of induction" will have to be conceived, in analogy to the rules of deduction, as canons of validation rather than of discovery. Far from generating a hypothesis that accounts for given empirical findings, such rules will presuppose that both the empirical data forming the "premisses" of the "inductive argument" and a tentative hypothesis forming its "conclusion" are *given.* The rules of induction would then state criteria for the soundness of the argument. According to some theories of induction, the rules would determine the strength of the support that the data lend to the hypothesis, and they might express such support in terms of probabilities. In chapters 3 and 4 we will consider various factors that affect the inductive support and the acceptability of scientific hypotheses.

THE TEST OF A HYPOTHESIS: ITS LOGIC AND ITS FORCE

3.1 Experimental vs. nonexperimental tests.

Now we turn to *a.* closer scrutiny of the reasoning on which scientific tests are based and of the conclusions that may be drawn from their outcomes. As before, we will use the word 'hypothesis' to refer to whatever statement is under test, no matter whether it purports to describe some particular fact or event or to express a general law or some other, more complex, proposition.

Let us begin with a simple remark, to which we will frequently have to refer in the subsequent discussion: the test implications of a hypothesis are normally of a conditional character; they tell us that *under specified test conditions,* an outcome of a certain kind will occur. Statements to this effect can be put into the following explicitly conditional form:

3d] If conditions of kind C are realized, then an event of kind E will occur.

For example, one of the hypotheses considered by Semmelweis yielded the test implication

If the patients in the First Division are delivered in lateral position, then their mortality from childbed fever will decrease.

And one of the test implications of his final hypothesis was

If the persons attending the women in the First Division wash their hands in a solution of chlorinated lime, then mortality from childbed fever will decrease.

Similarly, the test implications of Torricelli's hypothesis included conditional statements such as

If a Torrcelli barometer is carried to increasing altitudes, then its mercury column will correspondingly decrease in length.

Such test implications are thus implications in a twofold sense: they are implications of