PHILOSOPHY OF SCIENCE: A Formal Approach

HENRY E. KYBURG, JR.
The University of Rochester

© 1968

The Macmillan Company
NEW YORK

Collier-Macmillan Limited
LONDON

Chapter 1

THE CONCEPT OF A FORMAL SYSTEM

I

A number of notions may be taken as key concepts in the philosophy of science: entropy, for example, or information, or explanation, or problem-solving. I have chosen to organize this book around the concept of a *formal system*. To some extent, this is a matter of taste—all of these concepts are related, of course, and each finds its proper role to play in the framework supplied by whichever concept is taken as fundamental. But I think it is also more than a matter of taste, although to show that it is more than that will require (at least) the contents of this book.

The study of the philosophy of science from the point of view of a formalist has one disadvantage: if one is not formalistically inclined, one may tire of the formalities and abstractions, one may get fed up with axioms and rules of inference and want to return to the exciting realities of moon rockets and psychedelic drugs. This disadvantage I shall try to compensate for in two ways. First, in this chapter, I shall try to exhibit the way in which the formal view of science can be used as an effective tool in the investigation of some of the serious philosophical questions often raised by or about the scientific enterprise. Looking at science as a formal system, or as a collection of formal systems, does yield results. Second, in this chapter and throughout the book, I shall attempt to keep the explicit treatment of the formalities of particular systems of knowledge to a minimum. This is certainly appropriate; the philosophy of science is not science; our aim is not to become familiar with the details of bodies of scientific knowledge. Neither is our aim the popularization of science. Our purpose is rather to become familiar with those details that are relevant to philosophical problems. For example, we might be concerned with the details of a *test* of a scientific hypothesis, in order to explore the philosophical problem of the way in which empirical support is given to theoretical assertions. We might be concerned with the analysis of some small group of terms from a scientific theory, in order to attempt to clarify how theoretical terms get their meanings, and how those meanings are related to the meanings of words of ordinary language.

What are these philosophical problems of science, and how does this systematic approach hope to shed light on them? One group of problems centers around the motivation for doing science: what can science achieve for us? Can we properly expect to succeed? What are the limits, if any, that circumscribe the answers that science can give to our questions? Another group of problems centers around the justification and discovery of scientific theories: what are the roles played by observation, experiment, pure cogitation, in the creation of scientific theory? What is the nature of justification in science—i.e., what standards are called upon when the justification of a theory is at issue? What are the relations between discovery, test, and validation? What, if anything, is presupposed by the standards of validation? A final group of problems concerns the general philosophical implications of the scientific enterprise. These are problems that might be termed metaphysical: What does the fact that science exists at all tell us about the world? Or about the human mind? What is the import of psychology, or of biology, in its present state, for such classical issues as vitalism (the doctrine that there is a vital force differentiating living from nonliving matter) and teleology (the doctrine that goals are casually efficacious —e.g., that people grow eyes in order to see)? What has quantum mechanics to say that bears on such questions as causality, determinism, indeterminism, free will?

Since my object in this book is not to answer such questions as these, but rather to provide the logical and technical framework that is indispensable to the discussion of them, I shall not devote much space to addressing them. Some remarks will appear in the final chapter. But, since these questions are the

ones that bring people to the philosophy of science, I had better show how the formal systems that I am going to talk about are relevant to them.

What do we do science for? Among the answers are *control* of our environment, *explanation* of individual happenings, *prediction* of future events, *understanding* of a general kind of phenomenon. Each of these answers raises new questions: to what extent can we expect to succeed in controlling our environment? What are the limits of our control? In physics, for example, one learns that a perpetual motion machine is impossible. Does this represent a limit to our control of our environment, for example, to our ability to transform heat into useful work? Or might it be regarded as a defect of our system of physics, to be removed by the next Newton, or another Einstein? To answer such a question as this, we must be able to examine the logical relationships that hold among various statements of physics. What laws, or sets of laws, would have to be rejected in order to construct a system of physics in which the existence of a perpetual motion machine was no longer impossible? Or would we find it *logically impossible* to construct such a system of physics? In addition to these laws, we would have to reject, perhaps, various consequences of these laws. What are they? In order to understand why it is that no reasonable man spends his time looking for a perpetual motion machine, we must be able to examine the role that the statement "A perpetual motion machine is impossible" plays in the edifice of physics; or rather, the role that its formal counterpart, the second law of thermodynamics (the entropy of a closed system tends to increase), plays. In order to do this, we will find it advantageous to look at physics in a formal and systematic way, as a coherent and highly organized network of well-supported statements.

According to one well-known and generally accepted point of view (Hempel's), explanation and prediction are essentially the same: we *explain* an event that has already occurred, we *predict* an event that has yet to occur. The difference lies in whether the event is in the future or in the past. (We can sometimes convert a prediction into an explanation by waiting.) In either case, what we do is this: we are given a set of laws; we are given a set of boundary conditions—i.e., the set of circumstances to which we are going to apply these laws; and from these two things we infer deductively (or statistically) a statement asserting the occurrence of the event which is to be explained or predicted. Considered within the framework of a formal system, this view of explanation and prediction becomes even simpler. We begin with a formal system consisting of a set of statements (constituting the scientific subject matter of our explanation or prediction); we conjoin to this a new set of statements describing the circumstances of the predicted or explained event; we infer another statement from this complex system of statements, which inferred statement is a description of the event to be explained or predicted.

As an example, suppose we want an explanation of the fact that the tide was so high it nearly covered the dock yesterday. Newtonian theory implies that when the earth, the sun, and the moon are in a straight line, the tides will be extreme. Astronomical observation informs us that the sun, moon, and earth were in line yesterday. Common observation has shown us that in the ordinary course of events the dock is a foot higher than the high water level; an extreme tide is a tide a foot higher than normal. We abbreviate this whole explanation by saying simply that the dock was awash because the sun, moon, and earth were in a straight line.

To put the matter in terms of formal relationships (and to avoid some obscurities that can all too easily arise in connection with "facts"), we can say: Let N be a complex statement embodying the relevant parts of Newtonian theory; let A be a complex statement representing the astronomical state of affairs (including statements about the relative masses of the earth, sun, and moon, as well as their relative positions); let C be a statement of the physical characteristics of the dock; and let W be the statement that the dock was awash at high tide today. The explanation of W lies, on this reconstruction, simply in the purely logical fact that the sentence W follows deductively from the conjunction of the sentences N, A, and C, together with the fact that we have reason to accept N and A and C. The case of prediction, it is often argued, is just the same: we could use the same example, but take the date to be the day before the monthly springs of the tide, and regard the calculation to be the grounds of a *prediction* for the next day.

There is, of course, much more to be said about both prediction and explanation than this. One thing in particular seems to be open to question, and that is the symmetry of prediction and explanation. To use a well-known example, due to Michael Scriven, we would not predict of an individual who had syphilis that he would develop paresis, because most syphilitics do not develop paresis; but as an explanation of the fact that a certain individual had paresis a physician would cite the fact that he had had syphilis. It might be that there is a way of reestablishing the symmetry, even in this case. But what counts here is simply the fact that again the formal and systematic framework seems to be an ideal one within which to examine these questions. What we are after, ultimately, is not an understanding of prediction and explanation as they apply to this particular application of scientific theory or that one, but a general understanding of scientific prediction and explanation.

There are those (Bar-Hillel, for example) who regard such expressions as "scientific prediction" (only prophets make predictions) and "scientific explanation" (it is superstitious to seek an *explanation* for an event) as inherently inconsistent. But even this extreme thesis is defended within a formal framework; indeed it boils down to the claim that the science *is* the formal system, the set of axioms and theorems the scientist works with, and not any use that the scientist makes of it.

Consider the question of our scientific understanding of the world. Does a body of scientific knowledge increase the depth of our understanding of the world? Does it make the world seem more familiar, or ourselves more at home in it? In what sense, if any, is this true? How does it come about? Understanding is perhaps analogous to explanation, but concerned with classes of phenomena rather than with single events. We explain the fact that a particular crow stole a shiny button on a particular occasion by saying that all crows like shiny things and will often take them back to their nests; but we understand the button-stealing behavior of crows in general when we see how that type of behavior fits in with the rest of our knowledge of the behavior of crows. It is then that it becomes a familiar, homey, fact. We can claim to *understand* it only when we have a general system of knowledge, into which that bit of knowledge fits coherently. At least, that is one meaning of the word "understand". Again, though there remains a great amount to learn about what *fitting in coherently* is, it is clear that the framework provided by a relatively formal point of view is a natural one for discussing these questions.

The problems of the justification and of the discovery of scientific theories are generally regarded as quite distinct; the question of justification arises when we are given a theory or hypothesis or law, and a body of empirical evidence that is or might be alleged to support the theory. Under what conditions, we ask in general, does empirical evidence support a scientific theory or law? Discovery is another and knottier problem. Discovery is at issue when we are given a body of empirical data, and asked to discover the law or theory (or to explain the historical discovery of the law or theory, if any exists) which will render that body of data comprehensible, understandable. How should we go about discovering the appropriate law?

The answer to the first question requires the development of an inductive logic, or a theory of evidential support. In this logic, as in deductive logic, we will be concerned with forms of argument, rather than with the particulars of individual cases. In order to expose the formal structure of these arguments, we shall, as in the case of deductive logic, want to work within formalized language systems, in which the syntax of the arguments can be spelled out explicitly and in detail. There also are those, like K. R. Popper, who take the logic of evidential support to be simply the logic of tests of hypotheses; but this is essentially just classical deductive logic—we *deduce* a consequence of the hypothesis which we can then put to experimental test. Deductive logic, as we know, is best clarified within a formal and systematic framework.

The existence of a logic of discovery is more problematic. Most writers today discount the existence of such a logic, although there are exceptions, like N. R. Hanson. If there is such a thing—that is, if, in addition to a *psychology* of theory-creating or law-finding (we can, of course, look for the *empirical* facts of the scientist's mental activities), there is a *logic* that can provide clues as to the kinds of hypotheses that are worth testing—then it, too, will be the sort of thing that can be explored most effectively in a relatively general and abstract way. It is true that in exploring any of these problems we must keep in constant touch with concrete scientific realities; we must avoid cutting ourselves loose from the machines and test tubes and living beasts which are our subject matter; but we are doing philosophy, not science, and we are looking for broad logical generalities. Even if these generalities did not exist, even if every case of scientific discovery or scientific justification or scientific test had to be dealt with completely in its own terms, and in isolation from every other case, still we could perhaps only uncover that generalization by approaching these problems in an abstract and formal way.

The metaphysical presuppositions and implications of the body of scientific knowledge that we have can also be explored best when we have given that knowledge its clearest and most meticulously articulated form—that is, when we have given it the structure of a formal system. To answer the question, "What are the premises that current biological theory takes for granted, or are there none?" we need to be precisely aware of what the content of current biological theory is, and that content must be expressed in a formal enough manner so that questions about logical entailment, logical contradiction, and logical independence, can be given clear and definite answers. The same is true, on some views, of the probability relationships that hold between statements of the formal system. And it would be true also of probability relationships holding between those statements and others that might be regarded as in some sense metaphysical. Thus it might be maintained that although in modern scientific biology no concept of a *vital force* is required, and although no such concept is presupposed by the biological system as it stands, yet there are probability considerations that render the existence of such a force more probable than not. I shall not maintain this—I don't think it is true—but the point here is that either to establish or to refute such a view requires the biological system in question to be laid out in enough detail (and that the relevant sense of probability be closely enough defined) so that a fair degree of uniformity of opinion can be expected concerning whether or not, in point of fact, the biological facts cited do lend probabilistic force to the claim that a vital force exists.

It is sometimes claimed that philosophy makes no progress, and is still fretting, in the same way, about the very problems that bothered Greek philosophers. There may be a sense in which this is true, though I doubt that it is a very important sense; surely the arguments get better, even if the conclusions are often the same. But I think this is very definitely not the case in the philosophy of science; no one, I think, who has examined the development of the philosophy of science over the past hundred years, and in particular over the past fifty years, can fail to see the enormous strides that have been made in the analysis of evidential support, in the analysis of causality and determinism, and

in the examination of the nature of the scientific enterprise. It is also clear that this progress is due in large part to the increasing tendency of philosophers of science to forego juicy generalizations in favor of the meticulous, point-by-point examination of philosophical and scientific claims. In the past thirty-five years, this has involved more and more the formalization of existing scientific languages, or the creation of artificial languages of science, or (at the very least) the supposition that such languages are available. It has become clearer and clearer to many that the necessary condition of progress in the philosophy of science is the willingness to deal in abstractions, and in many cases to operate with carefully delineated formal systems. There are still those who think otherwise. But the plan of this book hinges on the belief that the most important concept for the philosophy of science is that of a formal system, and that the concept of a formal system, and the related concept of a body of rational beliefs, together provide the framework in which the solution of many of the traditional problems of the philosophy of science have often been sought successfully.

II

What is a formal system? Before we can define, or even describe, a formal system, we must make some distinctions. The most important of these is the distinction between object language and metalanguage. We look upon a formal system as a kind of language, with a vocabulary, a grammar, etc. At the same time, we want to talk about the formal system, and we require a language for this. For talking about a formal system (for talking about any of the several formal systems we will be talking about), we shall use English, supplemented with a few technical terms which will be carefully defined as they are introduced. This we shall call the *metalanguage*. The formal system, regarded as a language, will be called the *object language*. In the metalanguage, when we talk about expressions of the object language, we must use names for those expressions; in general we shall form names for expressions, just as we do in ordinary English, by enclosing the expression in quotation marks. Thus, just as Chicago is a large windy city, so "Chicago" is a certain word in English which is used as the name of a city. Another way of referring to a certain expression will be by displaying it on a separate line:

Chicago is a city.

This is one way of referring to a certain sentence that we might also refer to with the help of quotation marks, namely, "Chicago is a city."

We wish, in the metalanguage, not only to describe expressions of the object language, but also to make assertions concerning these expressions—e.g., that such and such an expression is a theorem, that this expression and that are contradictories, that one expression entails another. We also want to refer to meanings. Thus we want to be able to say in our metalanguage that the expression "red" in the object language means *red*. And sometimes, we are concerned with the complicated emotive or affective components in the use of the object language—e.g., with what is communicated when one person says to another of a third, "He is a red."

Within this broad range of metalinguistic activity, we can make a three-fold distinction between syntax, semantics, and pragmatics. *Syntax* is the part of the study of languages concerned with the bare arrangement of expressions of certain types—grammar (as opposed to diction) in English is primarily a matter of syntax. To know that "John between is" is not a sentence of English, we do now have to know what the words mean. Neither do we have to know who John and Mary are, nor even what loving is, to know that "John loves Mary," and "John doesn't love Mary," are contradictory. In the study of formal systems or formal languages, we generally give syntactical descriptions of the expressions that are regarded as forming the language; of the statements of the language that are to be regarded as axioms; and of the rules of inference of the language. (E.g., a rule of inference might be: From an expression of the form "If... then ___" and an expression of the form "...", the expression of the form "___" may be inferred, where the blanks of the two sorts are occupied by any sentential expressions of the object language.) Thus such metalinguistic terms as "... is a well-formed expression of such and such a language," "... is a theorem of such and such a language," etc. are syntactical. The part of a formal or linguistic system which is concerned simply with the manipulation of the expressions of that formal system according to the syntactical classes ("predicate", "variable", "parenthesis") that they fall into, is called a *calculus*. (A detailed definition will be given shortly.) The calculus of a formal system has, with some justice, been likened to a game played with the expressions of the object language. The justice of this metaphor lies in the fact that what we do in the calculus of a formal system depends on syntax alone, and the justification of any move we make, any inference we make, hinges solely on syntactical criteria. Meaning does not enter into our concerns at all, or need not, so long as we are merely playing the syntactical calculus game.

Of course, the game, as such, is not generally of much interest. We are interested in the manipulation of the expressions of the object language, according to the syntactical rules of manipulation, because these expressions do have meaning; if they had no meaning, we wouldn't be interested in the system. *Semantics* is the study of the meanings of the expressions of the object language. In our study of formal systems of various branches of science, we shall be concerned not only with the calculus of a formal system, but with the interpretation of that calculus. We are concerned not only with the expressions of the calculus of the formal system, but with the things that those expressions denote or designate or stand for. Semantics will be as much a matter for our study as syntax.

Although pragmatic considerations will guide us in our choice of formal systems for study, and in our formulation of the parts of those formal systems, *pragmatics* as such will not be of much concern to us. We are going to be concerned with the formal features of scientific languages and with the relationships between those languages and the world they purport to describe, but not specifically with the relations between the languages and the users of the languages, which is the relation with which pragmatics is primarily concerned.

One concept, however, which will be important in much of what follows is pragmatic. That is the concept of *relative interpersonal uniformity*. We shall say that one term has a higher degree of interpersonal uniformity than another (in a certain kind of situation) when users of the language are more uniform (potentially, as well as in historical fact) in their use of the first term than in their use of the second term. For example, in judging ambient temperature, a thermometer reading has a higher degree of interpersonal uniformity than intuitive judgments represented by terms like "hot", "cool", "cold", and "just right". This notion of interpersonal uniformity does involve essential reference to the users of the language, and thus is pragmatic. If we were to consider seriously the question of measuring interpersonal uniformity, we should have to do so by conducting empirical tests—it would be a scientific, empirical question. When we make use of this concept, it will be clear enough which terms have greater uniformity, without making empirical tests. But if tests were required, they would be part of the sociology of science or the linguistics of science; this is not to say that the philosophy of science is properly another science—not philosophy—but merely to point out that here as elsewhere the lines between analytical philosophy and empirical science are not clear and sharp.

One other matter should be mentioned now. I shall distinguish between a strict formal system, and a not-so-strict formal system. I shall regard bodies of scientific knowledge as they stand to be not-so-strict formal systems; that is, I shall take them to have roughly the same structure and components, implicitly, as a strict formal system, as I define it below. A not-so-strict formal system can be turned into a formal system by making explicit what is only implicit. This is a slight oversimplification, because there is more than one way, generally, of formalizing a bit of science—i.e., of turning it from a not-so-strict formal system into a strict formal system. Nevertheless, formalization is a useful procedure, as I shall attempt to bring out later.

In order to explain what a formal system is, in general, let me begin with the standard characterization of a strict formal system, and then proceed to show where the definition can be loosened along the seams while still keeping the garment the same.

A strict formal system consists of two parts:

I. A calculus
II. An interpretation.

Although there are some semantic systems (Carnap has presented some) in which the interpretation is provided formally, it is generally the case, even in a strict formal system, that the interpretation is treated more or less casually and informally. It is the interpretation, however, that provides the connection between the syntactical game of the calculus and the real world, and thus the interpretation is of the greatest importance to us in attempting to understand the connection between scientific theories and the world of toadstools and toolsheds.

The calculus of a formal system consists of the part of the formal system that can be dealt with by syntax alone. This part of a formal system is sometimes called a syntactical system. Though the calculus is quite literally meaningless without an accompanying interpretation, it is generally much easier to describe the syntax (the calculus) than the semantics (the interpretation), and (therefore?) the calculi of formal systems have received more attention than their interpretations.

To fix our ideas, let us consider the calculus of ordinary sentential or propositional logic. It is generally convenient to distinguish four parts of a calculus:

A. Vocabulary
B. Formation Rules
C. Axioms
D. Rules of Inference.

The *vocabulary* of a formal system consists of a list of all those signs that are used in writing the statements of the system, except those signs which are introduced explicitly by definition. The elements of the vocabulary are the *primitive signs* of the system. In our sentential calculus, they will consist of sentential variables *X*, *Y*, *Z*, *X'*, *Y'*, *Z'*, . . . , logical constants, and punctuation. These variables (under the usual interpretation of this calculus) take as their values, declarative statements. ("Socrates is wise," "The moon is between the sun and the earth," "The head of a horse is the head of an animal," and so on.) Note that in our vocabulary for sentential logic we allow for an infinity of sentential variables—this is the sense of the three dots ". . ."—but that we do so on the basis of precisely four signs: "*X*", "*Y*", "*Z*", and "'". The logical constants will be just these four: "∼", "∨", "&", and "⊃". The first, "∼", will be interpreted as logical negation; it corresponds to the judicious insertion of "not" in a declarative sentence: "Socrates is not wise," "The moon is not between the sun and the earth," etc. The wedge, "∨", will be interpreted as logical disjunction, corresponding to the inclusive sense of the English "or". Logical conjunction is represented by "&" in our formal system for sentential logic; under the usual interpretation it corresponds to the English "and", and sometimes to "but", "as well as", "also", etc. The sense of "&" is also

represented in English by simple juxtaposition: "John is foolish; Socrates is wise." The horseshoe, "⊃", is sometimes represented in English by "If . . . then ___"; but unfortunately matters are complicated for our interpretation by the fact that the English locution "If . . . then ___" is often used, not to express a *conditional*, but to express an *implication*—i.e., a certain logical relation between ". . ." and "___". The conditional, in logic, is simply a truth functional connective like "and" "or" and "not"—that is, it is an operation which you perform on two simple sentences to get a single more complicated sentence; it is *truth functional*, because the truth value of the more complicated sentence is taken to be determined once the truth values of the simpler sentences are given. The truth-functional character of these logical signs allows us to take the following table as giving their interpretation. (But more will be said about their interpretation below.)

X	Y	$\sim X$	$(X \vee Y)$	$(X \& Y)$	$(X \supset Y)$
T	*T*	*F*	*T*	*T*	*T*
T	*F*	*F*	*T*	*F*	*F*
F	*T*	*T*	*T*	*F*	*T*
F	*F*	*T*	*F*	*F*	*T*

Punctuation consists of the left and right parentheses: "("and")".

The *formation rules* tell us how to construct the meaningful or well-formed expressions of the object language whose syntax we are describing. First we specify the sentential variables: a sentential variable is "*X*" or "*Y*" or "*Z*", or else it consists of a sentential variable followed by an accent. A sentential variable alone is a well-formed formula; and if ". . ." and "___" are well-formed formulas, so are "(. . . ∨ ___)", "(. . . & ___)", "∼(. . .)" and "(. . . ⊃ ___)". Furthermore, we shall stipulate that only the expressions that are classified as well-formed by these rules are to be well-formed formulas of our calculus.

Notice that these rules are expressed in such a way that, given any expression whatsoever made up from pieces of the primitive vocabulary, we can follow a definite procedure which will determine, after a finite number of steps, whether or not that expression is well formed. A well-formed expression can have only one of five forms: it must either be a sentential variable, a disjunction, a conjunction, a negation, or a conditional. If it is a sentential variable, it must be "*X*" or "*Y*" or "*Z*", or else it must consist of a sentential variable followed by an accent. If it is a negation, it must consist of "∼", followed by a well-formed expression enclosed in parentheses. If it is a disjunction, conjunction, or conditional, it must consist of two well-formed expressions, separated by the ampersand, the wedge, or the horseshoe, and enclosed in parentheses. The new well-formed expressions referred to in this analysis—the components of the conjunction, disjunction, or whatever—will be subject to the same analysis as the first one; they will be sentential variables, or disjunctions, or conjunctions, or . . ., but they will be at least one symbol shorter than the original expression. The process of analyzing a given expression to see if it is well formed will thus come to an end.

The *axioms* of any system consist of a set of statements in the object language of that system. Syntactically, they are an arbitrary set of statements; semantically, they are generally a set of statements *assumed* to be true, from which (pragmatically) interesting statements of the system can be deduced. Sometimes this set of statements is finite, and then the axioms can simply be exhibited; sometimes the set of axioms is infinite, and then they must be described by means of axiom *schemata* (e.g., "Every well-formed expression of the form "((. . . $\vee$ . . .) $\supset$ . . .)" is an axiom"), or by some other technique of description in the metalanguage (e.g., "Every statement consisting of a well-formed formula followed by a horseshoe, followed by the same well-formed formula, is an axiom.")

To express the *rules of inference*, we again need to use general descriptions in the metalanguage. I have used dots and dashes (. . ., ---) so far, for simplicity; but now it will help to introduce the concept of a metalinguistic variable. I shall use "A", "B", etc, as metalinguistic variables. They stand for expressions of the object language; they are variables taking as their values expressions of the object language. To refer to a particular statement or expression of the object language, we follow the perfectly standard (English language) technique of enclosing that statement or expression in quotation marks. To refer to an expression of a certain *form*—e.g., to a well-formed expression which is a conditional whose antecedent is a conjunction—we use metalinguistic variables: we speak of a statement of the form $((A \,\&\, B) \supset C)$.

Let us now take the syntactical system I have been describing discursively, and express it as a calculus.

Calculus I

A. Vocabulary:
   1. Nonlogical signs:
      primitive sentential variables: X, Y, Z;
      operator for generating new sentential variables: $'$.
   2. Logical signs: $\sim$, $\vee$, $\supset$, &.
   3. Punctuation: (,).

B. Formation Rules:
   1. "X", "Y", and "Z" are sentential variables.
   2. If A is a sentential variable, A followed by "$'$" is a sentential variable.

   3. The only sentential variables are those expressions satisfying 1 or 2.
   4. A sentential variable is a well-formed formula.
   5. If A and B are well-formed formulas, so are $\sim A$, $(A \vee B)$, $(A \,\&\, B)$, and $(A \supset B)$.
   6. An expression is a well-formed formula only if it satisfies 4 or 5.

C. Axioms:
   1. $((X \vee X) \supset X)$
   2. $((X \vee Y) \supset (Y \vee X))$
   3. $((X \supset Y) \supset ((Z \vee X) \supset (Z \vee Y)))$
   4. $(X \supset (X \vee Y))$
   5. $((\sim X \vee Y) \supset (X \supset Y))$
   6. $((X \supset Y) \supset (\sim X \vee Y))$
   7. $((X \vee Y) \supset \sim(\sim X \,\&\, \sim Y))$
   8. $(\sim(\sim X \,\&\, \sim Y) \supset (X \vee Y))$

D. Rules of Inference:
   1. From well-formed formulas of the form A and $(A \supset B)$, B may be inferred.
   2. If A is like B, except for containing some well-formed formula C wherever B contains a certain sentential variable, then A may be inferred from B.

This calculus for sentential logic is now completely specified. We have not said what a proof is, or a theorem; but these concepts are the same for any formal system. We define both notions perfectly generally:

Proof:

A proof of a well-formed formula A in a given formal system, from a sequence of well-formed formulas $B_1, B_2, \ldots, B_n$ as premises, consists of a sequence of well-formed formulas, each of which is one of the formulas B_i, or an axiom of the system, or may be inferred by the rules of inference of the system from earlier formulas in the sequence, the last formula of which is A. (Provided no substitution is made—except in axioms—for variables of the premises.)

THEOREM:

A is a theorem of a given formal system if and only if there is a proof of A in that system from a set of premises which has no members.

Just to illustrate these two definitions, I shall give the formal proof of a theorem here, before continuing on to consider the interpretation of this calculus in more detail.

THEOREM: $(\sim X \vee X)$.

Proof:

1.	$((X \supset Y) \supset ((Z \vee X) \supset (Z \vee Y)))$	axiom 3
2.	$(((X \vee X) \supset X) \supset ((\sim X \vee (X \vee X)) \supset (\sim X \vee X)))$	1, rule 2
3.	$((X \vee X) \supset X)$	axiom 1
4.	$((\sim X \vee (X \vee X)) \supset (\sim X \vee X))$	2, 3, Rule 1
5.	$(X \supset (X \vee Y))$	axiom 4
6.	$(X \supset (X \vee X))$	5, rule 2
7.	$((X \supset Y) \supset (\sim X \vee Y))$	axiom 6
8.	$((X \supset (X \vee X)) \supset (\sim X \vee (X \vee X)))$	7, rule 2
9.	$(\sim X \vee (X \vee X))$	6, 8, rule 1
10.	$(\sim X \vee X)$	4, 9, rule 1

Q.E.D.

Here is a proof of a formula from premises: the proof of the formula Z, from the premises Y, $(Z \vee X)$, and $(X \supset \sim Y)$.

1.	$((X \supset Y) \supset ((Z \vee X) \supset (Z \vee Y)))$	axiom 3
2.	$((X \supset \sim Y) \supset ((Z \vee X) \supset (Z \vee \sim Y)))$	1, rule 2
3.	$(X \supset \sim Y)$	premise
4.	$((Z \vee X) \supset (Z \vee \sim Y))$	2, 3, rule 1
5.	$(Z \vee X)$	premise
6.	$(Z \vee \sim Y)$	4, 5, rule 1
7.	$((X \vee Y) \supset (Y \vee X))$	axiom 2
8.	$((Z \vee \sim Y) \supset (\sim Y \vee Z))$	7, rule 2
9.	$(\sim Y \vee Z))$	6, 8, rule 1
10.	$((\sim X \vee Y) \supset (X \supset Y))$	axiom 5
11.	$((\sim Y \vee Z) \supset (Y \supset Z))$	10, rule 2
12.	$(Y \supset Z)$	9, 11, rule 1
13.	Y	premise
14.	Z	12, 13, rule 1

Q.E.D.

It should be observed that the axioms are also theorems, according to the definition of theoremhood on page 13. The proof for Axiom 1, in virtue of which it is a theorem, for example, is just Axiom 1 itself. The sequence of formulas in question happens to be very short: one formula.

The usual interpretation of the calculus that I have just presented, the interpretation that justifies calling it a *sentential* calculus, is the interpretation mentioned earlier: The sentential variables, "X", "Y", etc., are taken to range over (to take as their values) declarative sentences of English. The logical signs are taken to represent the truth-functional aspects of their English counterparts.

Under this interpretation, the calculus represents the totality of all those expressions in English that are true by virtue of their truth-functional form alone (i.e., all the expressions like "It is raining or it is not raining"). If we let the variable "X" take the value "It is raining," then the conclusion of the first theorem proved is "It is not raining or it is raining." The calculus also embodies all of those valid arguments in English whose validity hinges only on their truth-functional structure. An argument in English is truth-functionally valid—valid, that is, because of its truth functional occurrences of "and", "or", "not", "if... then...", etc.—if and only if there is a corresponding derivation in the sentential calculus. Of course, there are many logical truths of English that are not true by virtue of their truth functional structure: "Everything is either red or not red," for example. And there are arguments that are valid, but not valid merely in virtue of their truth functional structure: "All men are mortal, and Socrates is a man; therefore Socrates is mortal," for example. Nevertheless, important parts of any kind of argument can be represented in this calculus.

An example of an argument that can be completely represented in our sentential calculus is the following:

> If the diamonds were really stolen, it was not an amateur job. Either the client is trying to defraud the insurance company, or the diamonds were really stolen. It was an amateur job.
>
> Therefore the client is trying to defraud the insurance company.

Suppose we let the variable "X" have the value "The diamonds were really stolen," and the variable "Y" have the value "It was not an amateur job," and the variable "Z" have the value "The client is trying to defraud the insurance company." Then the premises of this argument in English have the same form as the premises of the sample proof on page 14; and the conclusion of the argument in English is precisely the conclusion of the proof on page 14. The argument in English is truth-functionally valid, and there exists a corresponding proof in the sentential calculus.

That this is always the case—i.e., that if there is a proof in the calculus corresponding to an argument in English, then the English argument is truth-functionally valid—hinges on the fact that the rules of inference are truth preserving (we can never get from a true sentence to a false one using the two rules of inference of the calculus) and the fact that the axioms of the calculus are always true, regardless of the truth values of their components. (Regardless of the truth values of the sentences that the variables are taken to stand for, the axioms remain true, under the truth table interpretation of the connectives given on page 11). The detailed proof of the fact that there is a proof in the calculus if and only if the corresponding English argument is valid is beyond the scope of our discussion here, though it may be made plausible by reflection on the two facts mentioned.

The purpose of the formal system whose calculus has been presented above is to establish a standard mechanism to facilitate the assessment of validity in sentential argument. What happens is this: we find some arguments in our native tongue to be universally regarded as satisfactory. Some of the satisfactory arguments, arguments that we and all ordinary people regard as valid, involve nothing more than the English connectives "but", "and", "or", "not", etc., in their truth functional roles. That is to say, the validity of the argument rests solely on the relation between the truth-functional connections among sentences of the premises, and the truth-functional connections among sentences of the conclusion. What comes first is the recognition of a particular class (whose boundaries may be quite vague) of satisfactory arguments, of arguments that anyone *in his right mind* would regard as valid. This bare recognition is not science; it is not logic; it becomes logic when we begin to systematize and organize these types of arguments; and it becomes logic in the modern sense when we begin to construct a formal system, with a calculus such as Calculus I as its syntactical part, in which we can represent all possible arguments of truth-functional type.

A formal system of sentential logic is more than a calculus. It is a calculus together with an interpretation which provides some connection between the calculus and the things that are really of concern to us. As logicians, as mathematicians, or simply as playful people, we may be concerned with the calculus *qua* calculus. As scientists, or philosophers of science, or (applied) logicians, we are concerned with interpreted systems and not with pure calculi. Calculus I is intended to be a sentential calculus, and its interpretation is therefore relatively straightforward. But there are still alternatives even among standard interpretations. One alternative is the one given above. Another is this: we could take the variables "X", "Y", "Z", etc., to stand for any propositions (i.e., we could interpret the variables as taking particular propositions as their values), and we could take the logical connectives "$\sim$", "$\vee$", etc., to stand for the negation, disjunction, etc., of propositions. Propositions themselves, we must take to be the meanings of sentences. Now the negation, disjunction, etc., of propositions we may simply take to be truth-functional, so that the truth tables, rather than constituting an interpretation of the corresponding symbols, elucidate their given character. The calculus, thus interpreted, is called a propositional logic. The interpretation is perfectly coherent and sound, though it is a different one from the interpretation I offered first. It should be noticed, however, that this interpretation leaves vague the relation between sentences and propositions (and it is only sentences that can be exhibited), between propositional negation, and particular negating expressions of a given language, between propositional disjunction and the disjunctive locutions of a given language. In short, it leaves vague the relation between propositional logic and common discourse.

Although there is nothing very objectionable in this interpretation, I think it is

more profitable, and throws far more light on the function of formal systems as embodiments of scientific knowledge, to locate the vagueness and imprecision in the interpretation itself, rather than in the relation between the entities referred to in the interpretation and the entities (sentences, in this case) that we encounter in the world. The interpretation of Calculus I given first brings this out. The variables "X", "Y", "Z", etc., stand for sentences of (say) English; the logical constants correspond to the logical particles of English, insofar as the use of those particles in English is truth-functional. That is to say that "$\sim$" corresponds to ordinary English negation, insofar as ordinary English negation is simply truth functional, which is pretty far. Similarly, although conjunction in English sometimes contains elements of meaning that involve time only the aspects of English conjunctions such as "and", "but", ";", etc., that are truth functional are represented in the formal system of sentential logic by "&". Another way of expressing precisely the same thing, although it sounds different, is to say that the logical constants are to be *defined* by truth tables. Thus we can give an interpretation of the logical constants by means of truth tables, together with the concept of truth. The concept of truth focuses attention on the truth-functional aspects of the particles: to say that conjunction is to be such that the conjunction of A and B is true when A is true and when B is true, but not otherwise, is merely a roundabout but precise way of saying that logical conjunction is to function like conjunction in English insofar as conjunction in English is truth-functional.

In the interpretation we are now considering, we must also decide, more or less explicitly, what sentences of English the variables "X", "Y", etc., are to take as their values. Generally we say: simple declarative sentences. This characterization is all right; we understand what is meant. But here, too, there is an element of vagueness; it is not hard to imagine sentences whose form is declarative, but whose function is quite otherwise ("You haven't finished your assignment yet!"), and sentences whose form seems not to be declarative, but whose function is: ("Ouch!"). And there are any number of kinds of sentences whose truth-value is dependent on the circumstances of their utterance: "Today is Tuesday," "I am hot," "You are a liar," "That car is moving too fast." It is obviously necessary to take account of these systematic ambiguities in any application of the formal system to ordinary discourse.

It was stated earlier that one function of the formal system of sentential logic was to provide a simple mechanism by means of which to assess the validity of relatively complicated arguments. The concept of validity, applied to complicated arguments in ordinary language, was asserted to be equivalent to the concept of provability within the calculus. Whether or not a certain sequence of object language statements in the calculus is (say) a proof of A from premises $B_1 \ldots B_n$, is a question that admits of a simple, straightforward answer, that can be determined by inspection of the sequence of statements. It is characteristic of all formal systems that there is always a mechanical procedure for deciding whether or not an alleged proof of a particular statement is indeed a proof: we need simply inspect the sequence of statements that is alleged to be the proof, and see that each statement of the sequence is (*a*) an axiom, (*b*) a premise, if we are concerned with a proof from certain premises, or (*c*) derivable by one of the official rules of inference from earlier statements in the sequence.

This characteristic of formal systems has given rise to the (somewhat justified) assertion that they are designed to eliminate the necessity of intelligence. It takes a much higher order of intelligence to assess the validity of an argument directly than it does to check the steps in the proof (in a formal system) of the theorem that corresponds to that argument. In general this is the case: formal systems in the sciences are calculational devices as well as codifications of the facts.

III

Calculus I is not the only way of representing the syntax of sentential logic. Calculus II, borrowed from Hilbert and Ackermann, is easier to present than Calculus I, and easier to prove things *about*. It has the same interpretation as Calculus I, ordinarily; but, like the calculus of any formal system, it admits of other interpretations than that for which it was designed.

Calculus II

A. Vocabulary:
  1. Nonlogical signs:
     sentential variables: X, Y, Z, X', Y', . . .
  2. Logical constants: $\sim$, $\vee$.
  3. Punctuation: (,).

B. Formation Rules:
  1. A sentential variable is a well-formed formula.
  2. If A and B are well-formed formulas, so are $\sim A$ and $(A \vee B)$.
  3. An expression is a well-formed formula only if it satisfies 1 or 2.

C. Axioms
  1. $(\sim(X \vee X) \vee X)$.
  2. $(\sim X \vee (X \vee Y))$.
  3. $(\sim(X \vee Y) \vee (Y \vee X))$.
  4. $(\sim(\sim X \vee Y) \vee (\sim(Z \vee X) \vee (Z \vee Y)))$.

D. Rules of Inference:
  1. From well-formed formulas of the form A and $(\sim A \vee B)$, B may be inferred.
  2. If A is like B, except for containing some well-formed formula C, wherever B contains a particular sentential variable, then A may be inferred from B.

There are a number of differences between Calculus I and Calculus II. The most striking of these differences is that the logical constants "&" and "$\supset$" do not appear in Calculus II. Nevertheless, if we introduce "$\sim$" and "$\vee$" by means of a truth table interpretation, as we did before, then we can *define* "$\supset$" and "&" contextually:

D1. "$A \supset B$" is defined to be "$(\sim A \vee B)$."
D2. "$A \,\&\, B$" is defined to be "$\sim(\sim A \vee \sim B)$."

A check of the truth tables of "$(\sim A \vee B)$" and "$\sim(\sim A \vee \sim B)$" will show that they are precisely the same, in every possible case, as the truth tables we took in Calculus I to define "$\supset$" and "&". If we introduce these two definitions it turns out that every theorem of Calculus I is also a theorem of Calculus II, and every theorem of Calculus II is a theorem of Calculus I. The two calculi are thus equivalent. They contain the same body of theorems.

Another striking difference is that there are only half as many axioms, and half as many formation rules; and the formation rules of Calculus II are simpler. The difference in the number of axioms and the difference in simplicity of the formation rules are direct consequences of the fact that there are fewer logical constants in Calculus II than in Calculus I. The reduction in the number of formation rules (through the deletion of the formation rules concerning the construction of sentential variables) is due to the fact that the vocabulary is intended to contain an infinite number of sentential variables; that is the intent of the three dots that follow "Y'". Since the *primitive* vocabulary contains an infinite number of primitive variables, we do not need formation rules to enable us to construct an infinite number of sentential variables from a finite number of primitive variables and a variable operator.

A third difference would quickly become apparent if we were to begin to construct proofs of theorems in Calculus II; we would find it far more tedious and awkward than it is in Calculus I.

Calculus I and Calculus II, together with their normal interpretations, are simply two different ways of formalizing the same subject matter, truth functional logic. For some purposes one is advantageous, for some purposes the other. Given the usual interpretation of the sentential variables and of the logical constants, every theorem of one calculus will also be a theorem of the other.

Given a certain set of statements, it is possible to construct more than one calculus that will yield that set of statements as theorems. (Both Calculus I and Calculus II yield all the truth-functional truths as theorems.) It is also possible, given a calculus, to provide it with more than one interpretation. Calculus II, for example, can be taken as an arithmetic of the two numbers 0 and 1, if we take "$\vee$" to represent ordinary multiplication, and "$\sim$" to be interpreted thus: "~ 1" is to be interpreted as 0; "~ 0" is to be interpreted as 1. The variables, "X", "Y", "Z", etc., take as values, on this interpretation, the numbers 0, and 1; the well-formed formulas also represent numbers (0 and 1); the axioms, and the theorems, represent the number 0 however the numbers 0 and 1 are assigned to the sentential variables.

Yet another interpretation of this calculus (one that is useful to engineers!) takes the variables "X", "Y", "Z", etc., to represent the *states* of switches (open—no current can flow; closed—current can flow). The operator "$\sim$" changes the open state into the closed state, and the closed state into the open state. "$\sim X$" represents the state of a switch that is open when the switch corresponding to X is closed and closed when the switch corresponding to X is open. If "X" and "Y" stand for the states of two switches, "$(X \vee Y)$" stands for the state of the compound switch, consisting of those two switches in parallel (Fig. 1): If either of them is closed, current can flow. "$(X \& Y)$" (defined by D2)

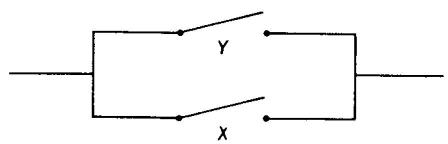

Fig. 1

stands for the state of the compound switch consisting of the two switches in series (Fig. 2): if either of them is open, no current can flow. The axioms of the

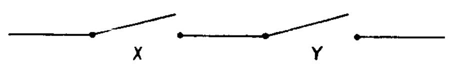

Fig. 2

calculus represent switching systems through which current can flow, regardless of the positions of the switches. The theorems of the calculus represent the set of all possible switching circuits with this characteristic.

IV

In most of the formal systems that will be discussed here, the subject matter is not logic, but physics or chemistry or biology. The formal systems of logic play a special role in each of these other formal systems; instead of restricting the rules of inference to two or three (sometimes to one!), as we do in setting up formal systems of pure logic, we will generally allow any valid principle of inference to serve as a rule. The logical formal system lurks in the background, providing both a standard of validity and a means of assessing validity; but we are interested neither in the economy nor in the elegance of the logical formal system, so much as we are interested in the economy or elegance of the overall formal system. We will therefore choose to embed in most of our formal systems a system of

logic which is messy in the sense that it incorporates a great many rules, but which is simple in that it is easy to construct proofs in that system of logic, and also in the sense that the allowable inferences are just (about) those inferences that one would intuitively expect to be allowable.

The vocabulary and the formation rules cannot be specified in general, because they depend on the terminology required for the science whose formalization is being considered. Thus "Patient A is schizophrenic," and "Patient B's I.Q. is over 110," represent the sort of sentences we might find in a body of psychological knowledge, whereas "$\sum M_i = 0$", "$E = mc^2$", and the like are the kinds of sentences we might expect to find in a body of physics. Axioms are required in order to be able to prove logical truths as theorems; but since the point of these applied formal systems is not to codify logical truths, but psychological, or physical, or chemical, or sociological truths, these axioms are of less importance than the rules of inference. I begin therefore by listing a number of rules of inference that will be taken for granted (unless otherwise stipulated) in each of the formal systems that follow. The first group concerns inferences that are truth-functionally valid; they can therefore each be justified by reference to Calculus I or to Calculus II: it is possible to show that any inference that is validated by the following truth functional rules of inference could also be validated by the axioms and rules of inference of either Calculus I or Calculus II.

Rules of Inference

Modus Ponens: From premises A and $(A \supset B)$, B may be inferred.
Modus Tollens: From premises $\sim B$ and $(A \supset B)$, $\sim A$ may be inferred.

> If the temperature falls below 20°F, the pond freezes.
> The pond did not freeze last night. Therefore the temperature did not fall below 20°F last night.

Hypothetical Syllogism: From premises $(A \supset B)$ and $(B \supset C)$, $(A \supset C)$ may be inferred.

> If the temperature falls below 20°F the horsetrough freezes.
> If the horsetrough freezes, the horse can't drink. Therefore, if the temperature falls below 20°F, the horse can't drink.

Disjunctive Syllogism: From premises $(A \vee B)$ and $\sim A$, B may be inferred.

> Tom will come to the party or Mary will come to the party.
> Tom won't be able to come to the party. Therefore Mary will come to the party.

Simplification: From premise $(A \& B)$, A may be inferred.

> There are two black cards and three red cards in my hand.
> Therefore there are two black cards in my hand.

Conjunction: From premises A and B, $(A \& B)$ may be inferred.

> There are two black cards in my hand. There are three red cards in my hand. Therefore there are two black cards in my hand and three red cards in my hand.

Addition: From the premise A, $(A \vee B)$ may be inferred.

> The die landed with a five-spot up. Therefore the die landed with either a five or a three up.

Replacement: If A and B are logically equivalent, and C is like D, except for containing one or more occurrences of A where D contains occurrences of B, C may be inferred from D.

In order to apply this rule, we must know what logical equivalence is. In general we say that if $(A \supset B)$ and $(B \supset A)$ are theorems, in a pure logical calculus, such as Calculus I or Calculus II, then A and B are truth-functionally, and thus logically equivalent. A useful list of truth-functional equivalences (which may be considered part of the rule of replacement) follows.

DeMorgan's Laws

$\sim(A \vee B)$	is equivalent to	$(\sim A \& \sim B)$
$\sim(A \& B)$	is equivalent to	$(\sim A \vee \sim B)$

Distributive Laws

$(A \& (B \vee C))$	is equivalent to	$((A \& B) \vee (A \& C))$
$(A \vee (B \& C))$	is equivalent to	$((A \vee B) \& (A \vee C))$

Associative Laws

$(A \vee (B \vee C))$	is equivalent to	$((A \vee B) \vee C)$
$(A \& (B \& C))$	is equivalent to	$((A \& B) \& C)$

Commutative Laws

$(A \vee B)$	is equivalent to	$(B \vee A)$
$(A \& B)$	is equivalent to	$(B \& A)$

Double Negation

$\sim\sim A$	is equivalent to	A

Transposition

$(A \supset B)$	is equivalent to	$(\sim B \supset \sim A)$

Exportation

$((A \& B) \supset C)$	is equivalent to	$(A \supset (B \supset C))$

Conditional Law

$(A \supset B)$	is equivalent to	$(\sim A \vee B)$

Conditionalization: B may be inferred from premises $A_1, A_2, \ldots, A_n$ if and only if $(A_n \supset B)$ may be inferred from premises $A_1, A_2, \ldots, A_{n-1}$.

This is an exceedingly valuable rule of inference; in proving a conclusion of the form $(B \supset D)$, it allows us to take as a *conditional premise* the antecedent B, of the conditional; we then construct a proof of D, from premises that include B; we then argue by conditionalization that $(B \supset D)$ follows from the remainder of the premises. We can argue, for example, that a conclusion of the form $(A \supset (A \& B))$ follows from a premise of the form $(A \supset B)$. The schematic outline of the proof would be

1. $(A \supset B)$	premise
2. A	conditional premise
3. B	1,2, Modus Ponens
4. $(A \& B)$	2,3, conjunction
5. $(A \supset (A \& B))$	2–4, conditionalization

Another use of conditional proof is in the reconstruction of *reductio ad absurdum* arguments. Assume as a conditional premise, the denial of the conclusion: $\sim C$. Derive a contradiction, $(P \& \sim P)$. Conditionalize: $(\sim C \supset (P \& \sim P))$ is yielded by the remainder of the premises. But $\sim(P \& \sim P)$ is a theorem (is obtainable from purely logical axioms), and so by Modus Tollens, we have $\sim\sim C$, or C.

The definitions of "proof from premises" and of "theorem" are unchanged from page 13. Here are two examples of arguments in ordinary English, and the corresponding proofs.

> If we urge Smith to come, he will come to visit us now. If Smith doesn't come to visit us now, then he'll visit us some other time. But either we urge Smith to come, or he'll not visit us another time. Therefore Smith will come to visit us now.

Let us write "S" as an abbreviation of "Smith does come to visit us now," "U" as an abbreviation of "We urge Smith to come," and "O" as an abbreviation of "Smith will visit us some other time." The premises of the argument become

1. $(U \supset S)$
2. $(\sim S \supset O)$
3. $(U \vee \sim O)$

And the proof of the conclusion becomes

1. $(U \supset S)$	premise
2. $(\sim S \supset O)$	premise
3. $(\sim O \supset \sim\sim S)$	2, replacement
4. $(U \vee \sim O)$	premise
5. $(\sim\sim U \vee O)$	4, replacement
6. $(\sim U \supset O)$	5, replacement

7. $(\sim U \supset \sim\sim S)$	3, 6, hypothetical syllogism
8. $(\sim U \supset S)$	7, replacement
9. $((U \supset S) \& (\sim U \supset S))$	1, 8, Conjunction
10. $((\sim U \vee S) \& (\sim U \vee S))$	9, replacement
11. $((\sim U \& \sim\sim U) \vee S)$	10, distributive law
12. $((\sim U \& U) \vee S)$	11, replacement.
13. $\sim(\sim U \& U)$	An instance of the axioms of this system
14. S	12, 13, disjunctive syllogism

When it rains, the crows are excitable. When it doesn't rain plants don't thrive. Either farming is a poor proposition, or plants thrive. The crows aren't excitable. Therefore farming is a poor proposition.

Let "R" abbreviate "It rains," "C" abbreviate "The crows are excitable," "P" abbreviate "The plants thrive," and "F" abbreviate "Farming is a poor proposition." The proof of the validity of the argument becomes:

1. $(R \supset C)$	premise
2. $(\sim R \supset \sim P)$	premise
3. $(F \vee P)$	premise
4. $\sim C$	premise
5. $\sim R$	1,4, modus tollens
6. $\sim P$	2,5, modus ponens
7. $(P \vee F)$	3, replacement
8. F	6,7, disjunctive syllogism

The logical framework of most of the systems we shall deal with must be taken to contain more than sentential logic. We must suppose that it contains a standard quantification theory as well. Quantification theory is the part of logic that corresponds to the implicit logic of the English terms "all", "some", "every", "there is...", etc. The key idea behind the treatment of the logic of these terms is the analysis of sentences into subject and predicate form. ("Subject" and "predicate" are not to be understood in their grammatical sense.) The statement "Socrates is mortal", for example, would be analyzed into a subject, "Socrates", and a predicate, "is mortal". If we symbolize the predicate "is mortal" by "M", and the subject "Socrates" by "s", we can symbolize the whole statement by "$M(s)$". If we abbreviate "John" by the symbol "j", we can represent the statement, "John is mortal" by "$M(j)$". If we let "H" stand for the predicate "is human", we can symbolize "John is human" and "Socrates is human" by "$H(j)$" and "$H(s)$", respectively. We can, of course, combine these sentences by means of the truth-functional connectives "$\sim$", "$\vee$", "$\supset$", etc. Thus "Socrates is human and John is human" becomes "$(H(s) \& H(j))$"; "Socrates is human or mortal", becomes "$(H(s) \vee M(s))$"; "If John is human, then John is mortal" becomes "$(H(j) \supset M(j))$"; and "If Socrates is human, then he is mortal", becomes "$(H(s) \supset M(s))$".

These last two sentences are true (it happens), and indeed they remain true no matter whose name you put in place of Socrates' name, because all humans are mortal. One way of paraphrasing this last clause is this: "Whatever you may name (or refer to), if it is human, then it is mortal." Since naming is not relevant to the mortality of all humans (and does not enter into our first statement), we can also paraphrase: "Whatever it is, if it is human, it is mortal." Replacing "it" by "x", we have "Whatever x is, $(H(x) \supset M(x))$". One standard way of symbolizing the phrase "Whatever x is" is simply to enclose the "x" in parentheses: we can thus represent "All men are mortal", as "$(x)(H(x) \supset M(x))$". "Everything is either human or not human", becomes "Whatever x may be, x is human or x is not human", or "$(x)(H(x) \vee \sim H(x))$". The sentence, "Everything is human," becomes "$(x)(H(x))$".

The expression "(x)" is called a quantifier; it is called a *universal* quantifier, for its function is to embody the universality expressed in English by such words as "all", "every", as well as "no", and "none". Thus "None of the players is a star," becomes "Whatever x may be, if it is one of the players, it is not a star." Symbolizing "is one of the players" by "P", and "is a star" by "S", we have, "$(x)(P(x) \supset \sim S(x))$". "No humans are mortal", becomes "$(x)(H(x) \supset \sim M(x))$".

Corresponding to the English locutions, "some", "there is a", "there are", 'is the existential quantifier,' "$(\exists x)$"; "Some humans are stars," can be paraphrased, "Something is such that it is a human, and it is a star," which in turn can be phrased, "For some x, $H(x)$ and $S(x)$," or, finally, "$(\exists x)(H(x) \& S(x))$".

The formulas we construct with quantifiers and sentential connectives may themselves be combined by means of sentential connectives. Thus "Either some human is immortal, or all men are mortal," becomes "$((\exists x)(H(x) \& \sim M(x)) \vee (x)(H(x) \supset M(x)))$".

There is nothing special about the letter "x"; we shall use a whole alphabet of variables, "x", "y", "z", "x'", "y'" in the same way. (Thus the last sentence symbolized could alternatively be symbolized: "$((\exists y)(H(y) \& \sim M(y)) \vee (z)(H(z) \supset M(z)))$".) These variables are called individual variables; they take individuals as their values. Before starting a discussion on the four basic modes of inference involving these quantifiers, we must define the *scope* of a quantifier. The scope of a quantifier consists of all the material between the first left parenthesis (not part of another quantifier) following that quantifier, and the right parenthesis that is the mate of the left parenthesis. For example:

$$(z)\ (\ (\ x\ )\ (\exists y)\ (\ H(\ x\ ) \supset P\ (\ y\ )\ \ ) \supset M(\ z\ )\ \ )$$

①②③④⑤⑥⑦⑧ ⑨⑩ ⑪⑫⑬ ⑭ ⑮⑯

The scope of the first quantifier "(z)" is the material between parenthesis ③, the first left parenthesis not part of another quantifier, and parenthesis ⑯, its mate. The scope of the quantifier "(x)" is all the material between parenthesis ⑧ and parenthesis ⑬. Parenthesis ⑥ is the first left parenthesis after the quantifier "(x)", but it is part of the quantifier "$(\exists y)$", and thus doesn't count. The scope of the quantifier "$(\exists y)$" is just the same: the material between parenthesis ⑧and parenthesis ⑬.

An occurrence of an individual variable is *bound* if it occurs within the scope of a quantifier of that variable, and is said to be bound to that quantifier. The occurrence of the variable within the quantifier is also bound. Thus in the preceding formula every occurrence of every variable is bound to the corresponding quantifier. An occurrence of an individual variable that is not a bound occurrence of that variable, is a *free* occurrence of that variable. For example:

$$(x)((x)(H(x)) \supset ((M(x) \vee (\exists y)(H(y))) \vee M(y)))$$
① ② ③ ④ ⑤ ⑥ ⑦

The occurrence of "x" in "$H(x)$" is bound to quantifier ②; the scope of that quantifier is "$H(x)$". The occurrence of "x" in "$M(x)$" is bound to quantifier ①, whose scope extends from the second left parenthesis to the last right parenthesis. Occurrence ⑤ of "y", like occurrences ① and ② of "x", is bound because it occurs in a quantifier. The occurrence of "y" in "$H(y)$" is bound to quantifier ⑤; since ⑤ is the only quantifier of the variable "y", and its scope is only "$H(y)$", the "y" in "$M(y)$" does not fall within the scope of any quantifier of "y", and is thus not bound, but is a free occurrence of "y".

I shall use lower case "v_1", "v_2", ... as metalinguistic variables taking as their values individual variables (e.g., "x", "y") of the object language. I shall use Greek letters "ϕ", "ψ", "χ", etc., as metalinguistic variables taking as their values well formed expressions of the object language that may (or may not) contain free individual variables. I shall use "n_1", "n_2", etc., as metalinguistic variables taking as their values proper names ("John", "Socrates", "5", "$\sqrt{2}$") of the object language, or definite descriptions ("the father of John", "the largest prime number less than 100", "the melting point of lead", etc.). Definite descriptions may also contain free variables: "the father of x", "the smallest prime number greater than y", etc. I shall say that an occurrence of (a name or definite description) n_1 in an expression $\phi(n_1)$ is a *free occurrence of* n_1 provided n_1 is a proper name, or a definite description that does not contain any free variable that becomes bound in $\phi(n_1)$ by a quantifier occurring in $\phi(n_1)$. "$\phi(v_1)$" is a variable-expression of the metalanguage, taking as its values well-informed expressions of the object language which contain at least one free occurrence of the individual variable v_1. "$\phi(v_2)$" and "$\phi(n_2)$" then denote the result of replacing all free occurrences of v_1 in $\phi(v_1)$ by v_2 or n_2, respectively, subject to the condition that v_2 (or n_2) must be free in $\phi(v_2)$ (or $\phi(n_2)$) at all those places at which v_1 occurs free in $\phi(v_1)$.

The quantification rules that follow are adapted from Copi.

Universal Instantiation: From $(v_1)\phi(v_1)$, $\phi(v_2)$ and $\phi(n_2)$ may be inferred.

Existential Generalization: From $\phi(v_2)$ or $\phi(n_2)$, $(\exists v_1)\phi(v_1)$ may be inferred.

Existential Instantiation: From $(\exists v_1)\phi(v_1)$, $\phi(v_2)$ may be inferred, provided v_2 occurs free in no earlier step.

Universal Generalization: From $\phi(v_1)$, $(v_2)\phi(v_2)$ may be inferred, provided v_1 occurs free in $\phi(v_1)$ at all and only those places in which v_2 occurs free in $\phi(v_2)$, and provided no variable introduced by existential instantiation is free in any unconditionalized assumption of the argument.

Principle of Identity:

From $\phi(v_1)$ and $v_2 = v_1$, we may infer $\phi(v_2)$.
From $\phi(v_1)$ and $n_2 = v_1$, we may infer $\phi(n_2)$.

Four useful equivalences involving quantifiers, that may be considered to be added to the replacement rule, are listed below:

$(\exists v)\phi(v)$	is equivalent to	$\sim(v) \sim \phi(v)$
$(\exists v) \sim \phi(v)$	is equivalent to	$\sim(v)\phi(v)$
$\sim(\exists v)\phi(v)$	is equivalent to	$(v) \sim \phi(v)$
$\sim(\exists v) \sim \phi(v)$	is equivalent to	$(v)\phi(v)$

These principles of inference suffice for a logic involving sentential connectives, quantifiers, and identity. It is the logic we shall suppose to be built into each of the formal systems to follow.

Let us look at a couple of proofs by way of illustration. It should be borne in mind that we are here trying to elucidate the structure of formal systems, and not trying to learn logic for its own sake; nevertheless, we shall in the future call on some of the logical principles laid down in this chapter.

Consider the argument:

> All vegetables either taste good, or are nutritious. No legumes that are vegetables taste good. Therefore if all legumes are vegetables, then all legumes are nutritious.

Let us write "$L(x)$" for "x is a legume," "$V(x)$" for "x is a vegetable," "$N(x)$" for "x is nutritious," and "$G(x)$" for "x tastes good." The premises become

(1) $(x)(V(x) \supset (G(x) \vee N(x)))$
(2) $(x)((L(x) \,\&\, V(x)) \supset \sim G(x))$

and the conclusion becomes

(3) $((x)(L(x) \supset V(x)) \supset (y)(L(y) \supset N(y)))$.

The proof goes like this:

1.	$(x)(L(x) \supset V(x))$	conditional premise
2.	$L(x) \supset V(x)$	1, universal instantiation
3.	$L(x)$	conditional premise
4.	$V(x)$	2,3, modus ponens
5.	$(x)(V(x) \supset (G(x) \vee N(x)))$	premise (1)
6.	$V(x) \supset (G(x) \vee N(x)$	5, universal instantiation
7.	$(G(x) \vee N(x))$	6,4, modus ponens
8.	$(L(x) \& V(x))$	3,4, conjunction
9.	$(x)(L(x) \& V(x)) \supset \sim G(x))$	premise (2)
10.	$((L(x) \& V(x)) \supset \sim G(x))$	9, universal instantiation
11.	$\sim G(x)$	8,10, modus ponens
12.	$N(x)$	7,11, disjunctive syllogism
13.	$(L(x) \supset N(x))$	conditionalization, 3–12
14.	$(y)(L(y) \supset N(y))$	3, universal generalization

(Observe that in lines 3–12, "x" is a variable that is free in a premise of each of these lines. In line 13 this no longer is the case: line 3 is *not* a premise of line 13. We are therefore free to generalize on the variable "x" in line 13.)

15.	$((x)(L(x) \supset V(x)) \supset (y)(L(y) \supset N(y)))$	conditionalization, lines, 1–14

The following argument involves relational predicates, as well as simple predicates.

> There is a movie star who is liked by everyone who likes at least one movie star. Everybody likes some movie star. Therefore there is a movie star who is liked by everybody.

Let us write "$M(x)$" for "x is a movie star" and "$L(x,y)$" for "x likes y". The premises are

(1) $(\exists x)(M(x) \& (y)((\exists z)(M(z) \& L(y,z)) \supset L(y,x)))$,
(2) $(y)(\exists z)(M(z) \& (y,z))$,

and the conclusion is

(3) $(\exists x)(M(x) \& (y)(L(y,x)))$.

The formalization of the argument is as follows:

1.	$(\exists x)(M(x) \& (y)((\exists z)(M(z) \& L(y,z)) \supset L(y,x)))$	premise
2.	$(y)(\exists z)(M(z) \& L(y,z))$	premise
3.	$\sim(\exists x)(M(x) \& (y)(L(y,x)))$	conditional premise
4.	$(x)(\sim (M(x) \& (y)(L(y,x))))$	3, replacement
5.	$(M(x) \& (y)((\exists z)(M(z) \& L(y,z)) \supset L(y,x)))$	1, EI
6.	$\sim(M(x) \& (y)(L(y,x)))$	4, UI

7.	$\sim M(x) \vee \sim(y)(L(y,x))$	6, De Morgan, replacement
8.	$M(x)$	5, simplification
9.	$\sim\sim M(x)$	8, double negation
10.	$\sim(y)(L(y,x))$	7,9, disjunctive syllogism
11.	$(\exists y)(\sim L(y,x))$	10, replacement
12.	$\sim L(y,x)$	11, EI

(Observe that "y" has occurred free in no earlier line.)

13.	$(y)((\exists z)(M(z) \& L(y,z)) \supset L(y,x)))$	5, replacement and simplification
14.	$((\exists z)(M(z) \& L(y,z)) \supset L(y,x)))$	13, UI
15.	$\sim(\exists z)(M(z) \& L(y,z))$	12, modus tollens
16.	$(\exists z)(M(z) \& L(y,z))$	2, UI
17.	$(P \& \sim P)$	15,16, conjunction
18.	$(\sim(\exists x)(M(x) \& (y)(L(y,x))) \supset (P \& \sim P))$	conditionalization

(Observe that the particular form of the contradiction we arrive at is irrelevant; it is thus abbreviated by $(P \& \sim P)$.)

19.	$\sim(P \& \sim P)$	derivable from logical axioms
20.	$(\exists x)(M(x) \& (y)(L(y,x)))$	18,19, modus tollens, replacement

(Observe that at no point did we have to quantify over variables.)

We are now in a position to give a general description of the kind of formal system we shall be concerned with from here on. It is a formal system which includes, as a framework, the quantificational logic just outlined. In addition to the rules of inference adumbrated above, it will include a primitive vocabulary, formation rules, axioms, and an interpretation.

V

In the general case, the vocabulary will include, beside the logical signs, "$\sim$", "$\vee$", etc., a vocabulary of variables "x", "y", "z", etc., and punctuation "(" and ")", and a number of nonlogical signs. These may be divided into two groups: individual constants, generally single lower-case letters, standing for individuals, like Socrates, John, a certain eclipse of the sun, a particular crow, etc. In addition to individual constants, there will be predicate constants, generally represented by capital letters, standing for properties ("$R(x)$" abbreviates "x is red", "$P(x)$" abbreviates "x is prime", "$D(x)$" represents "x is a dream"), standing for two-place relations ("$G(x,y)$" abbreviates "x is greater than y," "$M(x,y)$" represents "x is the mother of y")

standing for three place relations ("$B(x,y,z)$" abbreviates "y lies between x and z", "$G(x,y,z)$" abbreviates "x gives y to z"), and so on.

The formation rules will help to sort out the predicates; thus if we are going to interpret "$G(x,y)$" as meaning that x is greater than y, our formation rules will generally rule out such expressions as "$G(x)$", "$G(x,y,z)$", as not being well-formed formulas at all. In general, the formation rules will have this form:

A. A one-place predicate constant, followed by an individual variable enclosed in parentheses is a well-formed formula.
B. A two-place predicate constant, followed by two individual variables separated by a comma and enclosed in parentheses, is a well-formed formula.
C. A three-place predicate constant, followed by . . ., and so on.
D. If S is a well-formed formula according to (a), (b), (c), and S^* is like S, except for containing individual constants at one or more places where S contains individual variables, then S^* is a well-formed formula.
E. If S and T are well-formed formulas, then $\sim S$ and $(S \vee T)$ are well-formed formulas.
F. If S is a well-formed formula, and v an individual variable, then $(v)S$ and $(\exists v)S$ are well-formed formulas.
G. These are all the well-formed formulas there are.

The axioms will in general include appropriate axioms for deriving logical truths (such as $\sim(P \;\&\; \sim P)$, used above). They will also generally include certain laws concerning the subject matter being formalized. In formalizing a bit of physics, for example, we might take as axioms, Newton's laws; or in formalizing a bit of chemistry, we might take as axioms certain basic chemical principles like the law of constant proportions.

The interpretation of the logical constants will follow the lines already indicated; the interpretation of the individual variables has been indicated roughly, that is, "$(x)(\ldots)$" is interpreted as meaning that whatever *it*, x, may be, . . . ; and "$(\exists x)(\ldots)$" is interpreted as, meaning that there exists, somewhere, something, x, such that it . . . . The problematic part of the interpretation of a formal system generally concerns the interpretation of the nonlogical constants; and we shall find that this part of the interpretation really is problematic more often than not. There are certain special cases in which we do not encounter much difficulty: We can interpret a predicate like ". . . is red" as applying to just those things that are red; to say "x is red" is simply to say that x is red, and it will be true to say "x is red" if and only if x is red; and we can find out whether or not x is in fact red, most of the time, the moment an individual constant replaces the individual variable "x". But most scientific predicates are not so straightforward as "is red".

Even in a simple language with simple color words, like "red", "blue", "green", etc., we must recognize and deal with the problem that color

words are essentially vague: their range of application is not completely determined. There is a class of things that I would unhesitatingly call "blue"; there is a class of things that I would unhesitatingly call "not blue"; but there is also a class of things that would cause me to hesitate. Should I say that this is blue or shouldn't I? I could come to a decision, in each case, simply by tossing a coin: but I might very well be inconsistent in the sense that I classified something (thus arbitrarily) as blue on one occasion, which I might classify as not-blue on another. More important in a social institution like science than the fact that I find some things that are hard to classify as blue or not blue, is the fact that you, too, will find an area of vagueness in your application of the word "blue". Furthermore, the area of vagueness that you find will not correspond precisely to the area of vagueness that I have found.

All of this could be made quite precise, if it seemed worthwhile; we could construct a three-dimensional continuum (hue, brightness, saturation) of color samples; we could find a distribution function for a given person that would characterize his use of the word "blue" (it would be expressed as the probability that he would classify the objects in a certain region of the three-dimensional color space as blue); and we could then measure the ways in which two people differed in their use of color terms. This would be interesting for certain parts of linguistics or psychology, but not for the philosophy of science; for one of the ways in which science progresses and changes is by the refinement of the vocabularies in its formal systems; and one of the ways in which vocabularies are refined is by replacing terms that cannot be applied very uniformly by all people who are in a position to use them (color terms, for example), by terms which can be applied with a higher degree of uniformity by the people who know what they mean (expressions for wavelengths, for example). We seek terms with a high degree of interpersonal uniformity. We shall have occasion to look more closely at this phenomenon later on, and in fact, to trace some of the changes that have occurred historically along these lines.

A complication in the interpretation of nonlogical formal systems that is of another order of magnitude, is that involved in the use of so-called theoretical terms. While there is a fairly obvious sense in which a person can point to something that is blue, and say, "There, see that? That's blue!" there is no corresponding sense in which he can point and say, "There, see that? That's a gravitational field with such and such characteristics," or "That's an ego," or "That's a gene." We shall see later that this distinction can by no means be drawn as sharply as this argument suggests, and that the difference between *theoretical terms* and *observational terms* had best be regarded as a difference of degree. But the difference in degree between terms like "cat" and "dog", on the one hand, and terms like "force" and "field" on the other, is a difference that should be kept in mind when it comes to discussing the interpretation of scientific calculi.

Yet another complication introduced in the theoretical formal systems of science is the fact that theoretical terms—the primitive, nonlogical theoretical terms, like "mass" and "charge"—are introduced in such a way that they are only partly (if at all) interpreted independently of the formal system in which they function. Thus such terms as "mass" and "force" do not have independent interpretations outside of the axioms of physics that relate them to each other. But on the other hand, the axioms of physics are not the only things that determine their meaning; there are also stipulations about their meanings embodied in the system in the form of meaning postulates which relate them ultimately and indirectly to things we can touch and see and feel.

VI

I have been writing as if scientific theories were presented as formal systems; this is not the case, of course. But it is generally possible to take a scientific theory, or a body of scientific knowledge, and to transform it, to formalize it, in such a way that it then has the characteristics of a strict formal system. There are generally a number of alternative formalizations possible, for a given body of knowledge, and these alternatives may not differ merely in the choice of axioms or in the choice of primitive terms. In a given science, for example, a given statement may be taken as partially determining the meaning of one of its component terms, or it may be taken as making an assertion about the things denoted by those terms. Its role will depend, in a given formalization, on our arbitrary choice. But I shall take as a principle that nothing that depends on the particular form of formalization can be terribly significant philosophically.

The formal systems with which we have to deal here are not all strict formal systems in the sense in which I have defined them. There is a large element of idealization in what I have proposed that we consider. Nor do I suggest that the proper form for a scientific system is that of a strict formal system. It may well be that to the extent that formalization and systematization are feasible, they are all to the good; but it is surely not feasible to start to develop a specific formal system to correspond to the theory of psychoanalysis. Nevertheless, it is feasible, and it is helpful to our understanding of the character of psychoanalytic theory, to look at psychoanalytic theory as a system having somewhat the same structure as a strict formal system. It is worthwhile to look at its vocabulary, and at the possibility of defining some terms with the help of others; it is worthwhile to attempt to isolate those fundamental principles or laws from which psychoanalytic conclusions are intended to follow; it is worthwhile considering what statements are intended to follow; it is worthwhile considering what statements might be construed as specifying (or helping to specify) the meanings of technical theoretical terms, rather than as embodying factual assertions about mental or behavioral events.

Thus the concept of a strict formal system is not being presented here as one

to which all sciences do conform, nor even as one to which all sciences ought to conform; but rather it is being offered only as an ideal to which all sciences approximate, more or less, according to the degree of their development, and which serves to make clear the role that is played by some of the procedures, concepts, terms, and so on, in those sciences as they are. The concept of a formal system, then, offers two immediate advantages as a fundamental concept for the philosophy of science: first, it plays a fundamental role in every branch of science, and even in the explanation and systematization of other concepts that might be taken as fundamental (e.g., entropy, or information, or problem-solving, or decision-making); and second, it is a flexible concept. If we understand it broadly, much scientific talk and theorizing and assessment of experimental evidence, just as it stands, can be understood as falling within the framework of a formal system that is not presented strictly, but could be so presented. And if we want to talk about the formal system as such, we can understand the term more strictly, so that what we are talking about is a well-known thing that has been investigated, in the abstract, in great detail.

BIBLIOGRAPHICAL NOTES TO CHAPTER 1

Two authors who take the concepts of information and of entropy as their key concepts, are David Hawkins, *The Language of Nature*, (San Francisco: Freeman and Co., 1964), and Leon Brillouin *Science and Information Theory* (New York: Academic Press, 1962). John Dewey takes problem-solving to be the key concept of science in *Logic: The Theory of Inquiry* (New York: Holt, Rinehart and Winston, 1938). Ernest Nagel, *The Structure of Science: Problems in the Logic of Scientific Explanation* (New York: Harcourt, Brace and World, 1961), Carl Hempel, *Aspects of Scientific Explanation* (New York: The Free Press, 1965), and R. B. Braithwaite, *Scientific Explanation* (Cambridge, England: Cambridge University Press, 1953), all take explanation as a focal concept. Among other authors who take the concept of a formal system as central, Rudolf Carnap is the most distinguished, *Introduction to Symbolic Logic and its Applications* (New York: Dover Publications, 1958).

Most textbooks in the philosophy of science are organized around problems of a philosophical nature that are encountered in one form or another in most bodies of scientific knowledge: the problem of confirmation of hypotheses, the problem of the role of theoretical terms and the status of theoretical entities, the problem of reducibility of one science to another (e.g. the reduction of psychology to physics), etc. Four fine texts organized along these lines are: Ernest Nagel, *The Structure of Science* (New York: Harcourt, Brace and World, 1961), Arthur Pap, *An Introduction to the Philosophy of Science* (New York: The Free Press, 1962), and Israel Scheffler, *The Anatomy of Inquiry* (New York: Alfred Knopf, 1963). These three books are standard textbooks; C. G. Hempel, *Philosophy of*

Natural Science (Englewood Cliffs, N.J.: Prentice-Hall, 1966) is a small, eminently readable paperback.

The extreme view attributed to Yehoshua Bar-Hillel may be found in two articles (and in items of the bibliography appended thereto): "Comments on 'Degree of Confirmation'" by Professor Karl Popper, *British Journal for the Philosophy of Science*, **6**, 1955–56; "Further Comments on Probability and Confirmation", *Ibid.* **7**, 1956–57.

Michael Scriven's most detailed treatment of the problems of explanation and prediction is: "Explanations, Predictions, and Laws," in *Minnesota Studies in the Philosophy of Science* (Feigl and Maxwell, eds.), Vol. III (Minneapolis: University of Minnesota Press, 1962).

N. R. Hanson's defense of his claim that there is a logic of discovery, is to be found in its most complete form in: N. R. Hanson, *Patterns of Discovery* (Cambridge, England: Cambridge University Press, 1958).

For an introduction to the general nature and content of formal systems and of the relationship between formal systems and scientific knowledge, the following three monographs of the *Foundations of the Unity of Science* (Chicago: University of Chicago Press), are good: Vol. I., No. 1 (1938): *Encyclopedia of Unified Science*; Vol. I., No. 2 (1938): C. W. Morris, *Foundations of the Theory of Signs*; Vol. I., No. 3 (1939): R. Carnap, *Foundations of Logic and Mathematics*.

There are a large number of perfectly good textbooks in logic that will provide a more detailed introduction to the logical concepts very briefly introduced in this chapter. Among them are

Quine, W. V. O. *Methods of Logic* (New York: Holt, Rinehart, and Winston, 1959).

Rescher, N. *Introduction to Logic* (New York: St. Martin's Press, 1964).

Kalish and Montague, *Logic* (New York: Harcourt, Brace and World, 1964).

Suppes, P. *Introduction to Logic* (Princeton, N.J.: Princeton Univ. Press, 1957).

Copi, I. *Symbolic Logic*, Third Edition (New York: Macmillan, 1967).

Finally, there are a number of collections of essays concerning the philosophy of science; a few of them are

Philosophy of Science (Danto and Morgenbesser, eds.) (New York: Meridian Books, 1960).

Current Issues in the Philosophy of Science (Feigl and Maxwell, Eds.) (New York: Holt, Rinehart, and Winston, 1961).

Minnesota Studies in the Philosophy of Science, Vol. I. (Feigl and Scriven, eds.), Minneapolis, 1956.

———, Vol. II (Feigl, Scriven, Maxwell, eds.), Minneapolis, 1958.

———, Vol. III (Feigl and Maxwell, eds.), Minneapolis, 1962.

Chapter 2

A SIMPLE FORMAL SYSTEM

I

The formal system I propose to discuss in this chapter is a small one, but it embodies in a form easy to grasp all the elements of more complex systems. Let us suppose that there is a certain community in which social exclusiveness is the supreme value. Since the best index of social exclusiveness is membership in a club, everyone in the community belongs to at least one club. Since club membership isn't much of an index of exclusiveness if everybody belongs to all the clubs, there is a rule to the effect that two individuals can both belong to no more than one club. On the other hand, since everyone in this society is very friendly, every pair of people belong to at least one club together. Furthermore, unless there were some clubs whose memberships did not overlap, we could hardly consider them exclusive; so it is also established that if a person does not belong to a given club, there is exactly one club to which he does belong which