FORMAL METHODS IN THE FOUNDATIONS OF SCIENCE

Thesis by

David Lawrence Randall

In Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1970

(Submitted October 8, 1969)

ACKNOWLEDGEMENTS

This thesis is the result of prolonged study under the guidance of two men. It was through the readiness of Professor Gilbert D. McCann to involve me in his own experimental research that I was able to see the absolute dependence of the scientist on his experimental reality and his willingness to change and adapt theory and concept to that reality. It was through the patient and continual guidance of Professor Frederick B. Thompson that I came to realize that one could reasonably seek deeper understandings of the scientific process through the development of formal methods.

ABSTRACT

The intent of this study is to provide formal apparatus which facilitates the investigation of problems in the methodology of science. The introduction contains several examples of such problems and motivates the subsequent formalism.

A general definition of a formal language is presented, and this definition is used to characterize an individual's view of the world around him. A notion of empirical observation is developed which is independent of language. The interplay of formal language and observation is taken as the central theme. The process of science is conceived as the finding of that formal language that best expresses the available experimental evidence.

To characterize the manner in which a formal language imposes structure on its universe of discourse, the fundamental concepts of elements and states of a formal language are introduced. Using these, the notion of a basis for a formal language is developed as a collection of minimal states distinguishable within the language. The relation of these concepts to those of model theory is discussed.

An a priori probability defined on sets of observations is postulated as a reflection of an individual's ontology. This probability, in conjunction with a formal language and a basis for that language, induces a subjective probability describing an individual's conceptual view of admissible configurations of the universe. As a function of this subjective probability, and consequently of language, a measure of the

informativeness of empirical observations is introduced and is shown to be intuitively plausible - particularly in the case of scientific experimentation.

The developed formalism is then systematically applied to the general problems presented in the introduction. The relationship of scientific theories to empirical observations is discussed and the need for certain tacit, unstatable knowledge is shown to be necessary to fully comprehend the meaning of realistic theories. The idea that many common concepts can be specified only by drawing on knowledge obtained from an infinite number of observations is presented, and the problems of reductionism are examined in this context.

A definition of when one formal language can be considered to be more expressive than another is presented, and the change in the informativeness of an observation as language changes is investigated. In this regard it is shown that the information inherent in an observation may decrease for a more expressive language.

The general problem of induction and its relation to the scientific method are discussed. Two hypotheses concerning an individual's selection of an optimal language for a particular domain of discourse are presented and specific examples from the introduction are examined.

TABLE OF CONTENTS

I. INTRODUCTION

The tremendous scientific advances of the last several centuries have resulted in a variety of new disciplines and a greater degree of specialization in existing ones. In spite of this rapid growth and diversification, there remains a very firm belief on the part of many researchers in the unity of science. This view is aptly summarized in the following statement by Lee A. DuBridge [13, p. 6].

> "...science has reached a new level of attainment. Men now do comprehend the basic laws that enable them to interpret, and even in a large measure to predict, the behavior of the physical world—the world of the earth, the sun, the planets, the stars; of electricity and gravitation and the forces that hold atoms and molecules together. We even know that the processes of life depend on these same physical laws; that life is, in a very deep sense, the ultimate proof of the quantum theory. Hence, man's eternal quest for understanding of the world about him has led finally to the beginning of understanding the world within him."

Although many philosophers and scientific historians have contributed greatly to our understanding of science, there remain many difficult problems, and much needs to be done to fully explicate the scientific method and to characterize the relationship of the scientist to his science. In particular, the means by which a scientist becomes better informed, either by the gathering of experimental data or by communication with other researchers, is poorly understood. A prime task of information science should be the formulation of concepts which are useful in understanding these processes. This thesis is directed to that task.

The formal apparatus which is developed in subsequent chapters would be barren if it did not provide insight into basic informational problems and, in particular, into fundamental problems in the development and understanding of science. Therefore, in order to give perspective to these later quantitative chapters, we will begin by sketching several examples of such problems. We complete this introductory chapter by identifying two, more general philosophical issues, namely the problem of reductionism, and the question of the relationships between scientific theory and empirical observations.

The central portion of the thesis develops a mathematical apparatus by means of which these issues raised in the introduction can be formally treated. The later part of the thesis then returns to the problems of the introduction, viewing them from the vantage point of the developed apparatus.

Consider the relationship of the individual and his science as reflected in the language he uses to describe his theories and observations. As Polanyi points out, we often find that the words or concepts in this language shift in meaning on the basis of new observational evidence [26, p. 111].

> "When heavy hydrogen (deuterium) was discovered by Urey in 1932, it was described by him as a new isotope of hydrogen. At a discussion held by the Royal Society in 1934 the discoverer of isotopy, Frederic Soddy, objected to this on the grounds that he had originally defined the isotopes of an element as chemically inseparable from each other, and heavy hydrogen was chemically separable from light hydrogen. No attention was paid to this protest and a new meaning of the term 'isotope' was tacitly accepted instead. The new meaning allowed heavy hydrogen to be included among the

isotopes of hydrogen, in spite of its unprecedented property of being chemically separable from its fellow isotopes. Thus the statement 'There exists an element deuterium which is an isotope of hydrogen' was accepted in a sense which re-defined the term isotope, so that this statement, which otherwise would be false, became true. The new conception abandoned a previously accepted criterion of isotopy as superficial, and relied instead only on the identity of nuclear charges in isotopes.

"Our identification of deuterium as an isotope of hydrogen thus affirms two things: (1) that there exists in the case of hydrogen and deuterium an instance of a new kind of chemical separability, pertaining to two elements of equal nuclear charge, (2) that these elements are to be regarded as isotopes in spite of their separability, merely on the grounds of their equal nuclear charge. The new observations referred to in (1) necessitated the conceptual and linguistic reforms decreed in (2)."

This example illustrates that scientific terms tend to be redefined from time to time. One obvious question is why this process of re-interpretation takes place at all. What combination of circumstances necessitates such a change in language? Even more important, when words do change their meanings, what becomes of the theories which made use of them, and what is the weight of old observations on new enunciations of theory?

A closely related problem concerns two observers who use the same term, but one of them has a more detailed or ramified view of the concept denoted by the term than the other. For instance, an electrical circuit designer might visualize a "transistor" as a current-controlled device having three terminals and functional behavior described by a certain set of input-output equations. A solid state physicist, however, might very well characterize a "transistor" as a piece of semi-conducting material with a certain distribution of impurities which determine how the electrons and "holes" will move through the material when a potential difference exists. He may also implicitly associate such concepts as the mean free path of an electron, or the avalanche breakdown voltage of a junction with his notion of transistor.

There are clearly many circumstances when these men might find it advantageous or essential to converse on subjects which presuppose an understanding of what a transistor is. Even if we assume that their views are not fundamentally conflicting, we must admit that they implicitly disagree on what the relevant properties of a transistor are. Nevertheless, communication does take place, and presumably information is exchanged. But if these men both observe a transistor behaving in a certain novel fashion, are they equally informed by this observation? Will they both describe the observation in similar terms?

Another interesting phenomenon is illustrated by a comparison between what might be called the explicit scope and implicit scope of a scientific paper. Consider a biologist who describes the response characteristics of some neuron in a particular nerve ganglion of an insect, say the wolf spider. The writer is generally very careful to indicate that his results apply only under a certain set of experimental conditions. Sometimes the experimental animals are mutants, or many generations of them have been bred in the laboratory. And perhaps the stimulus is one which the animal would not normally encounter. Restrictions such as these, which are obviously necessary for practical reasons, tend to confine the stated results to rather specific cases.

What is the character of the information conveyed by such a paper? If the article is published in a reputable journal, is this because the editors believe that the restricted results obtained for the wolf spider are of overwhelming interest? It would not seem so. Rather, it must be that the results are supposed to have implications of a more general nature. Indeed, it would appear that these often unstated generalizations of the conclusions in the paper constitute much of its real significance. Writing in The American Psychologist, F. A. Beach emphasizes a slightly different aspect of this point in relation to the field of comparative psychology. [1, p. 119].

> "Perhaps it would be appropriate to change the title of our journal to read 'The Journal of Rat Learning', but there are many who would object to this procedure because they appear to believe that in studying the rat they are studying all or nearly all that is important in behavior. At least I suspect this is the case. How else can one explain the fact that Professor Tolman's book, 'Purposeful Behavior in Animals and Men', deals primarily with learning and is dedicated to the white rat, 'where, perhaps, most of all, the final credit or discredit belongs'. And how else are we to interpret Professor Skinner's 457-page opus which is based exclusively upon the performance of rats in bar-pressing situations but is entitled simply 'The Behavior of Organisms'?"

These three examples, from many that could be mentioned, illustrate problems of detail in understanding the current scientific scheme. We discuss now two problems of a more general nature which have loomed large in the consideration of the philosophy of science. The first of these is reductionism, which concerns the relationships between one theory and another, and the links connecting a theory with the observational data supporting it. Historically, the ideas of reductionism have been of paramount importance in influencing the attitudes of scientists on these questions. There are two problems, each of which has gone under the name reductionism, which should be distinguished although they are closely related. With regard to theory and observations, the purest form of reductionism is summarized by the view that every theoretical term can be explicitly defined in terms of observable quantities. That is, theories were considered to be simply disguised references to facts. These assumptions were quite prevalent until the early part of this century. As an example, Russell maintained the following. [33, p. 146].

> "Physics cannot be regarded as validly based upon empirical data until [light] waves have been expressed as functions of the colours and other sense-data."

The theory of relativity was instrumental in the demise of this very restrictive viewpoint. Statements such as "Space is curved in the neighborhood of the sun", were easily seen to be observationally unverifiable, at least directly. More recently, this form of reductionism was modified to allow such statements to be indirectly verified, for instance, by observing that light passing near the sun was deflected. However, this raises other problems since the original statement cannot be tested if no light is passing by the sun. Nevertheless, vestiges of these concepts are still evident today, particularly in those areas of scientific endeavor which have not proven amenable to the methods of formal mathematics. Thus, there are some researchers, in fields such as biology, who contend that the gathering of experimental data is informative in itself. That is, the mere existence of knowledge about an organism conveys information, even in the absence of any theoretical framework. Examples such as this demonstrate that the role of abstract theory as it relates to the acquisition of scientific knowledge remains open to question. In order to speak sensibly about information, therefore, it is necessary to examine how observational evidence and theories interact. They certainly appear to be independent in some respects, and yet our intuition indicates that they normally combine to support one another.

A second aspect of the reductionist philosophy deals with the problems of reducing one theory to another more basic one. For example, in the nineteenth century, classical mechanics was thought to be basic to much of science, and new theories such as thermodynamics were justified in part by demonstrating that they could be restated in the framework of classical mechanics. Many scientists went so far as to propose that all scientific phenomena should be reducible to some universal physical science. However, biologists, for example, generally refuted this view by claiming that there were characteristics of living organisms which could not be explicated in terms of atomic particles or any similar notions. These questions certainly bear directly on metaphysical assumptions concerning the unity of science. For a detailed historical discussion of these phenomena, see Nagel [25, chap. 11].

At the present time, very little effort is being expended to formally reduce one scientific theory to another, although there have been recent attempts to partially axiomatize biology, for example [46]. Instead, reducibility is often tacitly assumed, at least in an informal sense. Thus, even though physical chemistry and quantum mechanics are not explicitly linked by systems of axioms, they are treated as though intimately related. Furthermore, recent advances in the biological sciences have revitalized the reductionist viewpoint in this area, as the following quotation from Sinsheimer indicates [38, p. 5].

> "As we have penetrated the processes of the living cell, as the domains of mystery have receded, it has become ever more clear that all the properties of life can be understood to be simply inherent in the material properties of the complex molecules which comprise a cell. And thus that seemingly qualitative gap—self-evident to the most naïve—between the living and the non-living has in our time been bridged. Life is but a property of matter in a certain state of organization, and, given an organization which can reproduce itself, then adaptation and natural selection and, consequently, evolution will be just as inevitable a process as is the action of the second law of thermodynamics."

We now turn to the general question of the relationships between an abstract scientific theory and the empirical observational evidence supporting it. In this connection we inquire under what conditions a theory may be considered to be false. One obvious instance in which we would say that a given theory is false is if it is logically inconsistent. That is, the statements which characterize the theory contradict one another. Suppose, however, that the theory we are considering is not inconsistent. In this case, there will be some model for which the statements constituting the theory are all true. This model thus represents a possible reality, although it is not necessarily a plausible one from our standpoint. Therefore, given some set of statements which we call a theory, if these statements are logically consistent, we can postulate some collection of objects and relationships among them together with some interpretation of the sentences of our theory so that the theory becomes true of this domain. As we implied, the interpretation necessary to make the statements of the theory true may not agree with our common sense or our notion of reality. The pertinent question here is how one is able to make a distinction between what is plausible and what is implausible. Our intuition suggests that this must depend upon our observational experience, and that even though an abstract theory which is logically consistent has some model, this model may not conform to our empirical observations. We can therefore conclude that a consistent theory can not be said to be false without reference to observations or some means of characterizing the domain to which the theory refers.

As a simple example of this, suppose that we consider the statement "1/x is a rational number" as a theory. If this statement is about the positive integers it is certainly true, but if we are referring to the real numbers it is not universally true. Perhaps we are speaking about all of the integers and not just the positive ones; then what is the situation when x = 0? We might regard the statement as false in this case, or we could say that it is meaningless or undefined. Notice, therefore, that we must be able to specifically delimit the objects to which a theory refers in order to be able to assert its truth or falsity. This corresponds to interpreting the words of the abstract statements into particular entities, just as we have been concerned with the meaning of x in the example. These entities could be determined, in simple cases, by pointing to them one by one, or they might be characterized by some property which they have in common; but either directly or indirectly, certain observational evidence will be involved.

Now assume that we have some set of observations; can we then say definitely when a given theory is false of these observations? Here again we are faced with the dilemma of what a proper interpretation of the statements of the theory is to be. Thus, if we had an interpretation such that the statements of the theory were true of our observational experience, what if we then made some observation which the theory does not adequately explain? One possibility is that we may simply make the interpretation of the concepts in the theory less restrictive, thereby causing the theory to encompass the new observation. Consequently, without changing the statement of the theory we are nevertheless able to extend it to accommodate new evidence. Even though this may be technically possible, it is often not esthetically pleasing, and rather than extending the old theory by reinterpretation, it is discarded and a new theoretical framework is introduced in its place. Under what circumstances is a theory discarded and no longer considered to be scientific, and conversely, when can it be extended to explain some new phenomenon? Describing the anomalies confronting the scientific historian, Kuhn [22, p. 2] makes the following comment on this difficult problem.

"Simultaneously, these same historians confront growing difficulties in distinguishing the 'scientific' component of past observation and belief from what their predecessors have readily labeled 'error' and 'superstition'. The more carefully they study, say, Aristotelian dynamics, phlogistic chemistry, or caloric thermodynamics, the more certain they feel that those once current views of nature were, as a whole, neither less scientific nor more the product of human idiosyncrasy than those current today. If these out-of-date beliefs are to be called myths, then myths can be produced by the same sorts of methods and held for the same sorts of reasons that now lead to scientific knowledge."

Thus, the question of what is science and what is merely superstition is irrevocably bound to the notion of what constitutes a valid interpretation of an abstract scientific theory and thereby to the observations held to be relevant to it.

To say this, however, is only half of the problem. The plausibility or implausibility of an interpretation of some theory presumably depends directly on the individual's metaphysical assumptions about the nature of reality. Quine [30, p. 17] says the following about the relationship between an individual's ontology and scientific theory.

"Our acceptance of an ontology is, I think, similar in principle to our acceptance of a scientific theory, say a system of physics: we adopt, at least insofar as we are reasonable, the simplest conceptual scheme into which the disordered fragments of a raw experience can be fitted and arranged. Our ontology is determined once we have fixed upon the over-all conceptual scheme which is to accommodate science in the broadest sense; and the considerations which determine a reasonable construction of any part of that conceptual scheme, for example, the biological or the physical part, are not different in kind from the considerations which determine a reasonable construction of the whole. To whatever extent the adoption of any system of scientific theory may be said to be a matter of language, the same—but no more—may be said of the adoption of an ontology."

II. FORMAL LANGUAGES

In the introduction, we broadly outlined the problem area by means of a number of examples. This chapter will establish a specific framework within which the questions we have raised can be meaningfully discussed. In order to do this, we shall present a formalized theory. The question is, what type of theory is appropriate to the study and understanding of information processes. Since we are interested primarily in a description of how an individual, such as a scientific researcher, gains knowledge and structures the raw data he gathers, we should like to be able to characterize his personal view of his science. Therefore, the techniques used in well-established domains such as automata theory or information theory would seem to be unsuitable, at least intuitively, because they do not provide an apparatus for dealing with the subjective, differentiating aspects of an individual's understanding. Also, since communication by means of the spoken or written word is a basic process by which information is conveyed, our selection of a theoretical framework should be guided by the theory's ability to explicate this phenomenon. These considerations suggest that an individual's language is of prime importance in determining how he becomes informed [44, p. 207]. Therefore, we shall formalize the notion of an individual's language as it might relate to some restricted domain of discourse.

Our definition of a language will be firmly based on the notion of constructivity. Certain classes of constructive procedures have been studied in great detail. For instance, the methods of recursive function theory [20], Turing machines [43], the λ-calculus of Church [10], and Post production systems [28] have all been used to characterize what we mean by an effective process. In spite of the fact that the formalism underlying these methods differs widely, their intent is similar, and, indeed they have been shown to be mathematically equivalent. We shall use the term constructive to mean that given certain primitives and rules for operating on them, there is a well-defined procedure for deriving some result. In particular, the structures associated with our languages will have this property. We shall assume that any of the above formulations of constructivity coincides with our intuitive meaning of this notion, a view enunciated by Church [10, sec. 62].

We shall also be concerned with relating our definition of a language to modern structural linguistics. In this regard, the following statement of Chomsky's is of prime importance [7, p. 1].

> "The central fact to which any significant linguistic theory must address itself is this: a mature speaker can produce a new sentence of his language on the appropriate occasion, and other speakers can understand it immediately, though it is equally new to them."

The ideas expressed in this passage will be shown to relate directly to the question of constructivity. In addition, we shall make use of many of the linguistic concepts pertaining to the syntax and semantics of languages, particularly those dealing with the formal properties of grammars.

At this point, we should remark that we are not interested in language per se, at least in the normal sense of languages which are spoken or written. Rather we are interested in language as a vehicle embodying certain structure, and we shall regard it as a formal apparatus for explicating what is going on internal to a man's understanding. Therefore, this chapter will be devoted primarily to a characterization of the structural properties of language.

Now that we have indicated the suitability of the linguistic approach to the understanding of information processes, we shall describe precisely what we mean by a formal language. Such a language is to be distinguished from the natural languages, such as English or German, which do not have sufficiently well-defined rules for sentence formation, nor do they have a fixed vocabulary. However, as indicated by the above discussion, our definition of a formal language will provide a reasonable approximation to certain limited kinds of natural language communication. Mathematicians and logicians have used the notion of a formal language for a number of years. Most often they are referring to a rigid and highly stylized language within which a mathematical theory is expressed, the first-order predicate calculus, for example. We should like to take a somewhat broader viewpoint and examine the minimum conditions characterizing a formal language.

First of all, such a language should have a well-defined vocabulary or set of symbols which can be combined into meaningful strings of the language. Thus, the vocabulary of a formal language should be fixed, although it may be extremely large. Secondly, the sentences of the formal language should be certain finite strings of vocabulary symbols. And finally, given a string of symbols which constitutes a sentence, we should be able to specify a process for determining the meaning of that sentence.

Obviously, in order to be able to describe a formal language, we shall be forced to adopt some language, known as the meta-language, within which we embed our description. Furthermore, to determine the meaning of a sentence in a formal language, we shall need some model or structure which expresses the actual or possible interrelationships among the objects in the universe of discourse of the language. Thus, a sentence would be true of a model, if the objects referred to by the sentence have the structure within the model that the sentence requires.

The concepts developed in the area of model theory are quite similar to those we are seeking for this purpose [31]. However, in model theory, the models used are specific to the language being considered, the first-order predicate calculus, for example. These models are formulated in terms of set theory, and, indeed, set-theoretic models comprehend those of model theory. Since we wish to deal with a variety of languages and to characterize a model independently of them, our notion of model will be based on set theory.

The decision to structure our models in this way carries with it certain basic metaphysical assumptions. In essence, we are implying that Aristotelian logic and the concept of extensionality are valid for the universe of discourse of formal languages. That is, we are assuming that what is relates to an existing external world, in contrast to the phenomenalistic view that physical objects are merely convenient myths. Extensionality dictates that distinct objects or things possess different properties and are sensibly differentiable. These assumptions seem reasonable, and they clearly underlie all of modern science. We therefore make the basic ontological assumption that the universe of discourse of a formal language is set-theoretic in nature. By this we mean that objects and relationships denoted by the words of the language can be modeled by abstract sets and are expressible within set theory. This, of course, coincides with the practices of mathematical science today. Hence, the models for our formal languages will be models of axiomatic set theory.

A model of set theory will be interpreted as a set S of objects in the universe and an associated ε, which is a binary relation on S satisfying the axioms of set theory. We shall take as axioms those of Zermelo-Frankel set theory since we shall not need the class-set distinction of the Hilbert-Bernays system [11]. We assume the consistency of the axioms of set theory, thereby guaranteeing the existence of a model. Notice that there will be many models satisfying the conditions we have specified. Therefore, for a given S, in addition to the model where ε is the "natural ε", i. e. the standard model, and other models on S isomorphic to some standard model, we are also allowing all non-standard models on S. The natural ε of set theory will be denoted by $\in$, and the collection of all models of set theory by $\mathfrak{M}$, where

$$\mathfrak{M} = \{M \mid M = \langle S, \varepsilon \rangle \text{ is a model of set theory.}\} \quad \text{S is fixed.}$$

Fixing the set of objects S may seem arbitrarily restrictive, but this is not so if we assume that S is large enough to contain a set corresponding to every object in the universe of discourse. The fact that the universe of discourse can never encompass everything is irrelevant to this thesis and will not be further discussed. S is guaranteed to be at least countably infinite by the set-theoretic axiom of infinity. In general, of course, a given formal language will refer only to some subset of S.

In order to be precise in our formulations, we shall make use of a special language to be used as a descriptive aid in characterizing the semantics of formal languages. This language will be referred to as the language of set theory, and it comprises the standard formulation of the lower predicate calculus with an identity symbol and a single binary predicate ε. In addition, we shall refer to the axioms of set theory as expressions within this framework. Furthermore, we will augment this language occasionally with a denumerable number of names corresponding to particular elements of S . Both this language and the formal languages we shall define are considered to be in the domain of discourse of our meta-language. That is, they are object languages. The meta-language itself is assumed to be normal English together with an embodiment of the notion of a set. Thus, the meta-language speaks primarily about sets and is concerned with characterizing the relationships among the language of set theory, other formal languages, and the objects and relations to which they refer.

Using the language of set theory, we shall formulate the notions of meaning for a formal language and completely specify the semantics of such a language. It should be clearly understood that when we say that we assume the universe is set-theoretic in nature, we do not mean that a speaker of a formal language thinks about the universe as a collection of abstract sets. Rather, we mean that some hypothetical omniscient being, who "speaks" the meta-language, could analyze such a speaker's responses to questions in terms of set theory. On the one hand, we should like to make our ontology as strong as possible, because to do so lends additional underlying structure to our characterization of formal languages. On the other hand, if our ontology is too powerful, the allowable models may not include some possibilities which we intuitively feel are reasonable. In this case, we would preempt some of the aspects of understanding we are trying to explicate. Thus, our set-theoretic ontology has been selected as one which has sufficient structure to permit formalization of the semantics of a formal language without excluding models or configurations of the universe which are obviously plausible.

In view of the preceding remarks, we can now specify more precisely the minimal constituents of a formal language:

(i) A recursively enumerable set of symbols T.

(ii) A set of sentences $\mathcal{S} \subseteq T^*$. (The notation T^* designates the set of all finite strings of elements of T.)

(iii) For each model $M = \langle S, \varepsilon \rangle$ of set theory, and each sentence $\gamma \in \mathcal{S}$, $\varphi(M, \gamma)$ is a function whose value is the truth value of the sentence γ for the model M. φ is called an interpretation [14].

These three conditions are certainly necessary for a language, but are there other criteria which a formal language should satisfy? Suppose, for instance, that we are given a string of symbols which is a sentence of the language. We feel intuitively that there must be some finite process by which we can determine this fact. [7, p. 1]. That is, there should exist some algorithmic method for parsing the string and thus recognizing it as a sentence. If this were not the case, we would clearly be unable to ascertain the meaning of a sentence in a finite amount of time. Notice that we are not requiring the somewhat stronger condition that any string of symbols can be classified either as a sentence or a non-sentence in a finite number of steps. In summary, this additional restriction means that the set of sentences of a formal language must be recursively enumerable. The set of rules determining the process by means of which a sentence can be analyzed or generated is known as the grammar of the language. Properties of various types of grammars have been intensively studied by modern structural linguists. Our requirement that the set of sentences must be able to be enumerated recursively by the grammar is generally accepted as the weakest possible condition for a formal grammar. Chomsky designates such grammars as Type 0 and discusses features of these and other more restrictive grammars in [6].

Now consider the notion of the logical consequences of some set of sentences. Following Tarski [39, chap. 16], we will say that a sentence γ is a logical consequence of some set of sentences $\mathcal{E}$, if whenever all of the sentences of $\mathcal{E}$ are true, then γ is also true. We mean truth here in the sense of condition (iii) above, that is, in terms of the interpretation φ which is defined for any sentence γ and each model M of set theory. We remark that the concept of logical consequence as defined above does not necessarily embody a notion of proof. Hence, knowing that γ is a logical consequence of $\mathcal{E}$ does not mean that γ is provable from $\mathcal{E}$. A more thorough discussion of this distinction will be given in chapter IV.

Suppose that we have a set of sentences $\mathcal{E}$ and some sentence γ such that for every configuration of the universe, i. e. model, for which all of the sentences of $\mathcal{E}$ are true, γ is also true. Thus γ is a logical consequence of $\mathcal{E}$. Notice that this relationship is independent of specific knowledge about the universe. For instance, we presumably cannot decide the truth of the sentence "All men are mortal" because this depends upon knowledge about the world which we are incapable of obtaining. But if we are told that "All men are mortal" and that "George is a man", we conclude that the sentence "George is mortal" is a logical consequence of the preceding two sentences. Hence, our conclusion is based not upon the specific conditions for the truth of each sentence, but rather on a relationship among their truth conditions.

To require that we are able to decide in a finite number of steps whether or not a given sentence is true is certainly too restrictive. Such a position is tantamount to the strong verifiability theory of meaning of the logical positivists, and, as mentioned above, this has been refuted. But if we are given some set of sentences $\mathcal{E}$ and a logical consequence γ of $\mathcal{E}$, then we should have some finitary process for deciding that this is the case. That is, there must be a constructive procedure for enumerating the logical consequences of a set of sentences $\mathcal{E}$. We now state these two additional conditions formally.

Conditions of Adequacy for a Formal Language:

(i) The set of sentences $\mathcal{S}$ is recursively enumerable over T.

(ii) For any set of sentences $\mathcal{E} \subseteq \mathcal{S}$, the logical consequences of $\mathcal{E}$ are recursively enumerable over $\mathcal{E}$.

At this point we shall introduce a definition of formal language, and later, we will show it satisfies the conditions of adequacy.

Definition: A language $\mathcal{L} = \langle \text{Syn, Sem, Rn} \rangle$ is a formal language if:

(1) $\text{Syn} = \langle T, P, L, G, \sigma \rangle$ is a syntax for $\mathcal{L}$, that is

(i) T is a finite or countably infinite set, the referent words or terminal vocabulary of $\mathcal{L}$.

(ii) P is a finite set and $P \cap T = \emptyset$. P is the set of parts of speech or the non-terminal vocabulary.

(iii) L is a finite set of rules of the form $\alpha \rightarrow \beta$, where $\alpha \in P$, $\beta \in T$. L is the lexicon.

(iv) G is a finite set of rules of the form $\alpha \rightarrow \beta$, where $\alpha, \beta \in P^*$. G is the set of grammar rules.

(v) $\sigma \in P$. σ is a preferred part of speech, considered to be the part of speech of sentences.

(2) $\text{Sem} = \langle C, \tau \rangle$ is a semantics for $\mathcal{L}$, that is

(i) C assigns to each part of speech $\alpha \in P$, a formula $F(x)$ of the language of set theory, and C assigns to $\sigma \in P$ the formula $x = 0 \vee x = 1$. We will typically write $C_\alpha(x)$ rather than $F(x)$.

C_α will be called a semantic category.

(ii) τ assigns to each rule of grammar $R: \alpha_1\alpha_2 \cdots \alpha_n \rightarrow \beta_1\beta_2 \cdots \beta_m \in G$, a formula of the language of set theory $F(x_1, \cdots, x_m, y_1, \cdots, y_n)$ such that $F(x_1, x_2, \cdots, x_m, y_1, y_2, \cdots y_n)$ is of such a nature that it implies $C_{\beta_1}(x_1) \wedge \cdots \wedge C_{\beta_m}(x_m) \wedge C_{\alpha_1}(y_1) \wedge \cdots \wedge C_{\alpha_n}(y_n)$. τ_R will be called a semantic transformation.

(3) $Rn = < \mathfrak{m}, \Phi >$ is a realization space for $\mathcal{L}$, that is

(i) $\mathfrak{m} = \{ M | M = < S, \varepsilon >$ is a model of the language of set theory$\}$, where S is fixed. $\mathfrak{m}$ represents all possible configurations of the objects in the universe.

(ii) $\Phi = \{\varphi | \varphi$ maps T into $S\}$. Φ is the set of interpretations of $\mathcal{L}$.

A number of comments need to be made to aid in the interpretation of the preceding definition. Our notion of a syntax, for instance, is closely related to the representation of a syntax in algebraic linguistics. In terms common in that field we would say that $< T \cup P, T, L \cup G, \sigma >$ is a syntax[15, p. 8]. Also, the grammar G is an arbitrary Post production grammar or general rewrite rule grammar. There could obviously be special cases of a formal language where the grammar was context-free or some other particular form.

However, we shall not be too deeply concerned with the details of the syntax, since the properties of grammars such as this have received a great deal of attention in the literature, for example [6, 8, 15]. Notice that there may be referent words, i.e. elements of T, that do not participate in any meaningful strings of a language, since they may not occur in any rule in L. Thus, by taking a suitably large alphabet and all finite strings over this alphabet as elements of T, we could presumably obtain a common T for a large class of formal languages.

The most significant requirement in our definition of formal language concerns the interaction between the syntax and the semantics. We have assumed that the grammatical analysis of a sentence is not independent of the sentence's meaning. Therefore, in parsing a sentence, when some grammar rule R applies, we can also apply the corresponding semantic transformation τ_R to determine the meaning of the resultant of R. In English, for example, the syntactic entities which we call phrases generally are meaningful strings of words. Thus, "the boy on the red bicycle" is a noun phrase and presumably denotes some specific individual; whereas the string "boy on" which occurs within the preceding phrase, but is not itself a phrase, seems to be meaningless. Another way of looking at this condition involves recognizing that there are typically many different grammars which will generate a particular set of strings. We are insisting that the grammar chosen for a formal language be well-behaved in the sense that the intermediate strings generated by it, which are not sentences, must have a sensible interpretation. The fact that a speaker of a natural language is able to analyze and ascribe a meaning to a sentence he has never heard before lends intuitive support to an analogous view for formal languages, since even though the complete sentence is novel, presumably the words are known, and they are combined in a manner conforming to certain familiar rules [42].

It is very important to realize that the semantic categories and semantic transformations of a formal language have been specified to be structural in nature, i. e. independent of particular objects in the universe. Thus, the set of objects belonging to a given semantic category is a subset of all of the objects of S which share certain common properties. For instance, we might have a part of speech BR which signifies that the members of the corresponding semantic category are all to be binary relations. Thus the formula associated with BR would be $F(x) \equiv \forall y \left(y \varepsilon x \rightarrow \exists u \, \exists v \, (y = \langle u, v \rangle) \right)$, and examples of grammatical strings having part of speech BR might be, "<" or "is taller than". These strings could be embedded in sentences such as, "3 < 5", or "John is taller than Bob". In any particular model, the objects belonging to the semantic category C_{BR} would be given by

$$C_{BR}^{M} = \{ x \mid F_M(x) \} ,$$

where $F_M(x)$ denotes the relativization of the formula $F(x)$ to the model M . Thus, the formula $F(x)$ is expressed in terms of the predicate ε of the language of set theory, and $F_M(x)$ indicates that this predicate is to be interpreted as the particular ε associated with M .

The semantic category C_σ, corresponding to the distinguished part of speech σ, is uniquely specified for any formal language. That is, by convention, we always associate $F(x) \equiv x = 0 \vee x = 1$ with the part of speech corresponding to sentence, and we identify 0 with "false" and 1 with "true". Consequently, following Frege [14], a sentence of formal language denotes, or has as its extensional meaning, its truth value. If the set of denotations of the sentence is:

(i) $\{0\}$, the sentence is false;

(ii) $\{1\}$, the sentence is true;

(iii) $\{0, 1\}$, the sentence is ambiguous;

(iv) ϕ, the sentence is meaningless.

We shall define the manner in which a sentence denotes its truth value and discuss the questions of ambiguity and meaninglessness later in this chapter.

The semantic transformations of a formal language prescribe relations which must exist among elements of S , and the domain of a semantic transformation is restricted to those sets having certain structural properties. Thus, the semantic transformation $\tau_R(x_1, \cdots, x_m, y_1, \cdots, y_n)$, corresponding to the rule of grammar $R: \alpha_1\alpha_2 \cdots \alpha_n \rightarrow \beta_1\beta_2 \cdots \beta_m$, is such that, for any model M ,

$$\left\{ < x_1, \cdots, x_m, y_1, \cdots, y_n > \middle| \tau_R^M(x_1, \cdots, x_m, y_1, \cdots, y_n) \right\}$$
$$\subseteq C_{\beta_1}^M \times C_{\beta_2}^M \times \cdots \times C_{\beta_m}^M \times C_{\alpha_1}^M \times \cdots \times C_{\alpha_n}^M \quad .$$

We will occasionally write $< x_1, x_2, \cdots, x_m, y_1, \cdots, y_n > \in \tau_R^M$ to indicate that the elements of S identified with the x's and y's satisfy the semantic transformation τ_R in the model M. Therefore, the sets operated on by a semantic transformation must belong to the semantic categories specified by the parts of speech participating in the corresponding rule of grammar. If, in the course of parsing some sentence, we have certain elements of S identified with the variables $y_1, y_2, \cdots, y_n$ in the transformation τ_R, and there is no collection of elements of S corresponding to the $x_1, \cdots, x_m$ which satisfies the conditions of the transformation, then that particular parsing of the sentence fails on semantic grounds. A special case of this is often called a vacuous description, as in "the present king of France", which, although it is a grammatical English phrase, fails to describe any individual. Similarly, semantic ambiguity results if more than one collection of elements corresponds to the variables $x_1, \cdots, x_m$, as might occur in the phrase "the wife of Henry VIII".

The syntax and semantics of a formal language are both abstract mathematical entities. Only when they are associated with some space of realizations do the meanings of strings in the language become concrete. Thus, the models $\mathfrak{M}$, which we have previously discussed, represent various configurations of the actual objects in the language's universe of discourse, and the set of maps Φ characterizes the association of these objects with words in the language. A particular $\varphi \in \Phi$ will therefore assign to a word $\omega \in T$, some set or sets in S. Notice that the properties of these sets will generally vary from model to model, since each model has a distinct ε. In much of the following material, we shall assume a fixed interpretation φ for a given formal language and hence refer to the interpretation φ.

Now that we have described some of the salient features of a formal language qualitatively, we shall state more formally the mechanism which associates meaning with a phrase or sentence. To do this, it is necessary to define the notion of a parse.

<u>Definition:</u> For $\gamma \in T^*$, $p = < \alpha_0, \alpha_1, \cdots, \alpha_n >$ is a parse of γ, denoted by $< \gamma, p >$, if:

(i) $\alpha_0 = \beta_1\beta_2 \cdots \beta_k$ where $\beta_i \rightarrow \gamma_i \in L$ for $i = 1, \cdots, k$ and $\gamma = \gamma_1\gamma_2 \cdots \gamma_k$.

(ii) For each $i = 1, \cdots, n$ there are $\mu_{i1}, \mu_{i2}, \mu_{i3}, \mu_{i4} \in P^*$ such that: $\alpha_{i-1} = \mu_{i1}\mu_{i2}\mu_{i3}$, $\alpha_i = \mu_{i1}\mu_{i4}\mu_{i3}$, and $\mu_{i4} \rightarrow \mu_{i2} \in G$.

Utilizing this definition, we can now specify the interpretation to be assigned to a particular parse of some string γ. Therefore, let

$$\gamma = \gamma_1\gamma_2 \cdots \gamma_k \quad \text{where } \gamma_i \in T ,$$

and let

$$p = < \alpha_0 , \alpha_1 , \cdots , \alpha_n >$$

be a parse of γ , $\varphi \in \Phi$ and $M \in \mathfrak{M}$.

<u>Definition:</u> The interpretation φ^*_M of $< \gamma , p >$, denoted by $\varphi^*_M(\gamma , p)$ is given by

(i) if $p = < \alpha_0 > = << \beta_1, \beta_2, \cdots, \beta_k >>$, and if

$\varphi(\gamma_i) \in C^M_{\beta_i}$ for $i = 1, \cdots, k$, then $\varphi^*_M(\gamma, p) =$

$\{<\varphi(\gamma_1), \varphi(\gamma_2), \cdots, \varphi(\gamma_k)>\}$,

(ii) if $p = < \alpha_0, \alpha_1, \cdots, \alpha_h, \alpha_{h+1} >$, $h+1 \leq n$, and

$\alpha_h = \lambda_1\lambda_2 \cdots \lambda_m \mu_1 \cdots \mu_r \nu_1 \nu_2 \cdots \nu_t$

$\alpha_{h+1} = \lambda_1\lambda_2 \cdots \lambda_m \mu'_1 \cdots \mu'_s \nu_1 \nu_2 \cdots \nu_t$

$R : \mu'_1 \cdots \mu'_s \rightarrow \mu_1 \cdots \mu_r \in G$ and

$< x_1, \cdots, x_m, y_1, \cdots, y_r, z_1, \cdots, z_t > \in$

$\varphi^*_M(\gamma, <\alpha_0, \cdots, \alpha_h>)$ and $<y_1, y_2, \cdots, y_r, y'_1, \cdots, y'_s> \in \tau^M_R$,

then $< x_1, \cdots, x_m , y'_1, \cdots, y'_s , z_1, \cdots, z_t > \in \varphi^*_M(\gamma, p)$,

(iii) otherwise $\varphi^*_M(\gamma, p) = \phi$.

We now show that it is possible to associate a unique formula of the language of set theory with an interpretation $\varphi^*_M(\gamma, p)$.

<u>Lemma</u>: Let $\gamma = \gamma_1\gamma_2 \cdots \gamma_h \in T^*$ and $p = < \alpha_0, \cdots, \alpha_n >$ a parse of γ, $\alpha_n = \beta_1\beta_2 \cdots \beta_k$, $M \in \mathfrak{M}$. Then there is a formula $F(x_1, \cdots, x_h, y_1, \cdots, y_k)$ of the language of set theory such that: there are $< a_1, \cdots, a_k > \in \varphi^*_M(\gamma, p)$ if and only if $F(x_1, \cdots, x_h, y_1, \cdots, y_k)$ holds for $< \varphi(\gamma_1), \cdots, \varphi(\gamma_h), a_1, \cdots, a_k >$ in M.

Proof: (i) if $p = < \alpha_0 >$, then $h = k$ and $\beta_i \rightarrow \gamma_i \in L$. Hence,

$\varphi^*_M(\gamma, p) = \{ < \varphi(\gamma_1), \cdots, \varphi(\gamma_h) > \}$ if and only if

$\varphi(\gamma_i) \in C^M_{\beta_i} = \{ x \mid F^{\beta_i}_M(x) \}$, by the definition of φ^*_M.

Now, consider the formula $F(x_1, \cdots, x_h, y_1, \cdots, y_h) =$

$F^{\beta_1}(x_1) \wedge F^{\beta_2}(x_2) \wedge \cdots \wedge F^{\beta_h}(x_h) \wedge (x_1 = y_1) \wedge \cdots \wedge (x_h = y_h)$.

Clearly, $<a_1, \cdots, a_h > \in \varphi^*_M(\gamma, p)$ if and only if $F_M(x_1, \cdots, x_h, y_1, \cdots, y_h)$ holds for $< \varphi(\gamma_1), \cdots, \varphi(\gamma_h), a_1, \cdots, a_h >$.

(ii) if $p = <\alpha_0, \cdots, \alpha_i, \alpha_{i+1} >$, and $\alpha_i = \overline{\lambda}\,\overline{\mu}\,\overline{\nu}$, $\alpha_{i+1} = \overline{\lambda}\,\overline{\mu'}\,\overline{\nu}$ where $\overline{\lambda}, \overline{\mu}, \overline{\mu'}, \overline{\nu} \in P^*$. R: $\overline{\mu'} \mapsto \mu \in G$ and $H(\overline{x}, \overline{u}, \overline{y}, \overline{v})$ is the formula of set theory associated with $\varphi^*_M(\gamma, <\alpha_0, \cdots, \alpha_i>)$, i. e. $< \overline{a}, \overline{b}, \overline{c} > \in \varphi^*_M(\gamma, < \alpha_0, \cdots, \alpha_i >)$ if and only if $H_M(\overline{x}, \overline{u}, \overline{y}, \overline{v})$ holds for $< \overline{\varphi(\gamma)}, \overline{a}, \overline{b}, \overline{c} >$.

$\overline{\varphi(\gamma)}$ denotes $\varphi(\gamma), \varphi(\gamma), \cdots, \varphi(\gamma_h)$. Suppose $G(\overline{z}, \overline{y})$ is the formula of set theory corresponding to τ_R, i. e. $\tau^M_R = \{ < \overline{z}, \overline{y} > \mid G_m(\overline{z}, \overline{y}) \}$. Now consider the formula $F(\overline{x}, \overline{u}, \overline{z}, \overline{v}) = \exists \overline{y} \left(H(\overline{x}, \overline{u}, \overline{y}, \overline{v}) \wedge G(\overline{z}, \overline{y})\right)$ and note that $< \overline{a}, \overline{d}, \overline{c} > \in \varphi^*_M(\gamma, p)$ if and only if there are $\overline{b}$ such that $< \overline{a}, \overline{b}, \overline{c} > \in \varphi^*_M(\gamma, < \alpha_0, \cdots, \alpha_i >)$ and $< \overline{d}, \overline{b} > \in \tau^M_R$.

But this is true if and only if there are $\overline{b}$ such that $H_M(\overline{x}, \overline{u}, \overline{y}, \overline{v}) \wedge G_M(\overline{z}, \overline{y})$ holds when $\overline{x} = \overline{\varphi(\gamma)}$, $\overline{u} = \overline{a}$, $\overline{y} = \overline{b}$, $\overline{v} = \overline{c}$, and $\overline{z} = \overline{d}$. And by definition this holds if and only if $F_M(\overline{x}, \overline{u}, \overline{z}, \overline{v})$ holds for $< \overline{\varphi(\gamma)}, \overline{a}, \overline{d}, \overline{c} >$. Therefore F is the desired formula, and the lemma is proved.

<u>Theorem</u>: Let $\gamma = \gamma_1\gamma_2 \cdots \gamma_h \in T^*$, and p a parse of γ to σ, i.e. $p = < \alpha_0, \alpha_1, \cdots, \alpha_n >$ where $\alpha_n = < \sigma >$. Then there is a formula $F(x_1, x_2, \cdots, x_h, y)$ of the language of set theory such that:

$a \in \varphi^*_M(\gamma, p)$ if and only if $F_M(x_1, x_2, \cdots, x_h, y)$ holds for $< \varphi(\gamma_1), \cdots, \varphi(\gamma_h), a >$ where $a = 0_M$ or $a = 1_M$.

Proof: Follows directly from the preceding lemma and the fact that $C^M_\sigma = \{ 0_M, 1_M \}$, where 0_M and 1_M denote the "0" and "1" of the model M of set theory.

We shall now characterize the conditions under which a sentence is ambiguous, either syntactically or semantically. Suppose $\gamma \in T^*$, and p_i, $i \in I$ is a parse of γ to σ, i.e. γ is a sentence of the formal language.

<u>Definition</u>: The sentence γ is semantically ambiguous on a given model $M \in \mathfrak{m}$, if for some p_i, $i \in I$

$$\varphi^*_M(\gamma, p_i) = \{0_M, 1_M\} .$$

Definition: The sentence γ is syntactically ambiguous for a given model M, if for some p_i, p_j ; $i, j \in I$

$$\varphi^*_M(\gamma, p_i) = \{ 0 \} \text{ and } \varphi^*_M(\gamma, p_j) = \{ 1 \} .$$

Examples of sentences which are either syntactically or semantically ambiguous are easy to find. Thus, in English, "They are flying planes" is an often used illustration of a sentence which is syntactically ambiguous. Similarly, the statement $2 = \sqrt{4}$ can be considered to be semantically ambiguous since we do not know whether $\sqrt{4}$ denotes $+2$ or -2. When we say that the above two statements are ambiguous, we really mean that they are ambiguous from the standpoint of most English speakers. That is, they are ambiguous with respect to models or configurations of the universe admitted by these speakers. However, it should be obvious that a given sentence γ may be ambiguous in one formal language and unambiguous in another. This is certainly the case in natural languages, also, since ambiguity may depend on factors such as the speaker's context or environment, or his previous experience.

In the following discussion, we shall temporarily ignore the question of ambiguity in order to simplify the explanations somewhat. Under this restriction, we have $\varphi^*_M(\gamma, p) = \varphi^*_M(\gamma)$, which may assume any one of the values $\{0\}$, $\{1\}$, or $\emptyset$. Thus, given a sentence γ and a model $M \in \mathfrak{m}$, we have the following cases:

(i) γ is true for M if and only if $\varphi^*_M(\gamma) = \{1\}$, which is equivalent to $F_M(x_1, \ldots, x_h, y)$ holds for $< \varphi(\gamma_1), \ldots, \varphi(\gamma_h), 1 >$.

(ii) γ is false for M if and only if $\varphi^*_M(\gamma) = \{0\}$, which is equivalent to $F_M(x_1, \cdots, x_h, y)$ holds for $< \varphi(\gamma_1), \cdots, \varphi(\gamma_h), 0 >$.

(iii) γ is meaningless for M, if $\varphi^*_M(\gamma) = \phi$, which is equivalent to $\neg \exists y F_M(x_1, \cdots, x_h, y)$ holds for $< \varphi(\gamma_1), \cdots, \varphi(\gamma_h) >$. This points out a useful corollary to the preceding theorem.

Corollary: For a sentence $\gamma = \gamma_1\gamma_2 \cdots \gamma_h$ and a model M, there is a formula $G(x_1, \cdots, x_h)$ such that $\varphi^*_M(\gamma) \neq \phi$ if and only if $G_M(x_1, \cdots, x_h)$ holds for $< \varphi(\gamma_1), \cdots, \varphi(\gamma_h) >$.

Proof: $G(x_1, \cdots, x_h) = \exists y F(x_1, \cdots, x_h, y)$.

The formula $G_M(x_1, \cdots, x_h)$ in the corollary holds for the sets denoted by $\varphi(\gamma_1), \cdots, \varphi(\gamma_h)$ exactly in the case where γ is a meaningful sentence. But if γ is not a meaningful sentence, what then? The general problems caused by admitting that some grammatical strings of a language may fail to convey any meaning have troubled many logicians and linguists. As a consequence, a variety of explanations and solutions have been proposed. In almost all cases, however, meaningless sentences are treated in a rather ad hoc fashion. It is our contention that such sentences may be incorporated within the framework of a formal language in a much more natural manner. Further, it is quite necessary to do so in order to explicate the issues raised in the introduction. The implications of our approach will become more apparent in chapter IV when we discuss the question of probability.

Consider the sentence, "Green ideas sleep furiously.", an example proposed by Chomsky [5, p. 15] of a sentence which, although grammatical in English, would generally be said to be meaningless. Chomsky's method of handling such a sentence is to require that the grammar of the language be sufficiently powerful to exclude this as a grammatical string. [7, chap. 2]. This seems to be undesirable for several reasons. First of all, to attempt to eliminate all meaningless sentences on syntactic grounds alone will necessarily cause the grammar of the language to become increasingly complex, since such strings must fail to parse completely. Furthermore, the syntax of a language may be unable to accomplish this because the exclusion of such strings may depend upon the specific meanings of particular words and not simply on grosser structural classes to which they belong. Thus, in the limit, we could conceivably be forced to assign a separate part of speech to every referent word, thereby rendering the syntax useless as a general set of rules for sentence formation. One form of syntactic analysis which inherently involves this problem is co-occurrence analysis [18]. This involves examining grammatical strings of the language to determine when one word may be substituted in a string for another without destroying grammaticalness of the string.

This brief discussion does point out one particularly interesting fact about the structures of languages in general, and our formal languages in particular. That is, there will typically be a trade-off between the syntactic and semantic components of a language. We may be able to make one more powerful while letting the other become weaker and still preserve essentially the same set of meaningful sentences. Within certain limits, this is certainly possible, but in the extreme, the language will lose its intuitive relation to reality. Thus, the semantic categories of a formal language are intended to correspond to generic classes of perceivable entities, and the characterization of these classes may depend upon non-linguistic considerations involving the speaker's experience. Therefore, although this trade-off may be theoretically arbitrary, in realistic situations, it would not seem to be so.

In English, for example, the distinctions between an animate noun and an inanimate noun are extremely difficult to articulate. We must draw on a great deal of our knowledge about the world to characterize these two classes of entities. If, in a formal language, we attempt to discriminate these on a syntactic basis, then the associated semantic categories will be characterized by extremely complex set-theoretic relations, or possibly these concepts are not even structurally expressible. In the latter case, they would not be appropriate notions for a formal language, since we demand that the syntax and semantics should be clearly delineated and that the syntactic aspects be purely structural. It is important to realize that the semantic categories of a language are fundamental to the entire semantic structure of the language. They represent atomic classes of entities which share certain generic structural properties, and the semantic transformations can be considered to be explicitly developed or constructed on them. Thus the semantic categories embody the deeper aspects of the semantics, and since they are in a one-to-one correspondence with the parts of speech, their number and complexity directly affects the grammar. Additional ramification of semantic categories may result in a simplification of the grammar, but this occurs because certain relationships which were previously explicit in the grammar, now have become implicit in the deep structure of the semantics.

Another circumstance in which a sentence may be meaningless is when it contains a vacuous description, as we mentioned previously, for example, "The present king of France is bald". Russell [32] takes the view that such a sentence should be considered meaningful but false. The difficulty with this analysis is that it is not clear just how negation is to be handled. To say that the above sentence is false is to say either that there is no present king of France or that he is not bald. This strains normal English usage and by so doing confounds the very task we have set ourselves. Similar remarks hold for other proposed solutions that make such sentences "meaningful". In our scheme, however, if a sentence is true of some model, then the negation of that sentence is false of that model, provided that the negation of the sentence is expressible within the formal language. In a similar fashion, the negation of a meaningless sentence will also be meaningless. Recall that the extended interpretation φ^*_M of a string which is not meaningful is null, i.e. denotes the empty set of the model of set theory M . This is related to the approaches adopted by Quine, Frege, and Carnap. For a discussion of their viewpoints on this problem, see Carnap [4, chap. I].

One very important distinction inherent in our method is that we explicitly deal with a collection of models $\mathfrak{M}$. Thus a sentence may be meaningless for one model but meaningful for another. For instance, in the previous example, if it is not "known" whether or not France currently has a king then there will presumably be admissible models for which the sentence asserting that the king is bald is meaningless and other models for which the sentence is true or false. This question of "knowing" some fact will be explicated by our notion of observation which is discussed in the following chapter. However, it is sufficient at this point to realize that sentences of a formal language will generally be true of some models, false of others, and meaningless or ambiguous on still others. A further difference is that if a description or phrase of the language does not denote a unique entity, we consider the sentence in which the description is embedded as ambiguous, whereas Carnap, for example, says that such a sentence is meaningless. Thus for a model possessing two kings of France both of whom were bald, we would say that the sentence "The present king of France is bald" is true of the model, assuming the sentence has only two distinct interpretations.

The use of quantification in the statements of a language raises some rather subtle problems, as indicated by Quine [30, p. 13] .

For instance, if $\forall x P(x)$ is a statement of the predicate calculus, then the range of the bound variable x is assumed to be over all objects in the universe of discourse. In this case, the world which the language speaks about is essentially homogeneous, i.e. all objects belong to the same structural class. In our formal languages, as in all natural languages, the total collection of entities referred to does not have this property. In a formal language, the semantic categories partition the universe of discourse into distinct classes. Consider the statement "There is an x such that x is red". What objects are potentially represented by x ? Certainly not every element of the universe of discourse, since things like "ideas" are meaningless in this context. How are such things excluded? By the definitions of semantic categories and transformations for a formal language, the range of a quantifier is over exactly those entities in some semantic category. Thus, in the preceding example, x might denote any object whose part of speech was concrete noun. Observe that there may still be many objects in the semantic category which are not in the range of φ , i.e. are not named in the language, although they possess the generic structural properties of the objects which do have names. A similar situation exists for negative statements such as "That is not a book". Here again, the object referred to is not completely arbitrary but must belong to the same semantic category as books do. Thus many of the problems associated with the range of quantifiers are handled simply by formalizing the notion that the universe of discourse of a language is not homogeneous.

We mentioned briefly in the preceding discussion that a language might not be able to express both a sentence and its negation. That is, the language may not include any symbol or string of symbols which are always interpreted as "not". In general, the set of referent words T of a language may be quite arbitrarily mapped by an interpretation φ into the elements of S. Therefore, if the symbol "$\neg$" appears in T, there may be interpretations φ which do not associate the customary meaning of "not" with this symbol. In the case of this particular symbol, however, we have already implicitly presumed an understanding of its meaning in our basic ontological assumptions. The reason for this is that we formulated the axioms of set theory, which define the admissible models in terms of the first-order predicate calculus. Thus, symbols such as "$\neg$", "$\wedge$", "$\vee$", "$\forall$", and "$\exists$" appear in the underlying axioms of our system of models, and we have tacitly assumed that, for example, in the set-theoretic formula $F(x) \equiv G(x) \wedge H(x)$, the symbol "$\wedge$" has a well-defined interpretation. As a consequence, if we have a formal language which is incapable of expressing these basic logical concepts, then this language would seem to be somewhat pathological. Of course, a formal language which does include a logical notion such as "not" will not necessarily associate this with the symbol "$\neg$". In English, for example, we can generally negate a sentence γ by saying "It is not the case that γ".

One other minor point regarding our formal language definition is that we have assumed that the entire terminal vocabulary T consists of referent words. Note, however, that the entities denoted by the interpretations of the words are not required to participate meaningfully in the semantic transformations. For instance, the elements associated with words such as "the" or "an" could act essentially as dummy arguments, and their semantic categories would possess some correspondingly trivial properties. Words of this character are sometimes called function words and treated in a special fashion, but there is no reason for doing this in our formalization.

At this time, we return to the conditions of adequacy for a formal language, which were stated earlier in this chapter, and show that our concept of a formal language satisfies them.

<u>Theorem</u>: A formal language $\mathcal{L}$ satisfies the conditions of adequacy:

(i) The set of sentences $\mathcal{S}$ of $\mathcal{L}$ is recursively enumerable over T .

(ii) For any set of sentences $\mathcal{E}$ of $\mathcal{L}$, $\mathcal{E} \subseteq \mathcal{S}$, the logical consequences of $\mathcal{E}$ are recursively enumerable over $\mathcal{E}$.

Proof: (i) The quadruple $< T \cup P, T, L \cup G, \sigma >$ is a syntax for $\mathcal{L}$. It is equivalent to an unrestricted rewriting system and therefore the set of strings of sentences generated by it is recursively enumerable. [6, p. 358].

(ii) Suppose we have a set of sentences $\mathcal{E}$ of $\mathcal{L}$ and a sentence γ which is a logical consequence of $\mathcal{E}$, i.e. for every $M \in \mathfrak{M}$, if $\varphi^*_M(\gamma') = \{1\}$ for $\gamma' \in \mathcal{E}$, then $\varphi^*_M(\gamma) = \{1\}$. We must show that there is a constructive procedure by which we can prove γ starting with the set of sentences $\mathcal{E}$. We have previously demonstrated that for any sentence γ, there is a formula F^γ of the language of set theory which holds if and only if γ is true, i.e. when $\varphi^*_M(\gamma) = \{1\}$. A careful review of that proof will establish that the passage from γ to F^γ is constructive, indeed depending entirely on the constructive character of the grammar. Let $\mathcal{F}$ be the set of such formulas corresponding to $\mathcal{E}$, and let F^γ be the formula corresponding to γ. By the definition of logical consequence, for every $M \in \mathfrak{M}$, if $F^{\gamma'}_M$ holds for $F^{\gamma'} \in \mathcal{F}$, then F^γ_M holds. Therefore by virtue of the Gödel completeness theorem for the first order predicate calculus, F^γ is provable from $\mathcal{F}$. (See Lyndon [23, p. 56], for an appropriate form of this theorem.) Consequently, there is a constructive procedure which determines that γ is a logical consequence of $\mathcal{E}$, and the logical consequences of $\mathcal{E}$ are thus recursively enumerable with respect to $\mathcal{E}$.

Notice that the preceding theorem does not yield a method of proof for the language $\mathcal{L}$ in general, since the procedure for enumerating the logical consequences of a set of sentences $\mathcal{E}$ may depend upon the nature of $\mathcal{E}$. The following corollary shows that we can obtain a procedure which is independent of $\mathcal{E}$.

Corollary: For any formal language $\mathcal{L}$, there is a method of proof for $\mathcal{L}$, i.e. a procedure such that for any set of sentences $\mathcal{E}$, the logical consequences of $\mathcal{E}$ are recursively enumerable with respect to $\mathcal{E}$, using this procedure.

Proof: The requirement here is that the enumeration procedure obtained in the preceding theorem be uniformly recursive over the set of sentences $\mathcal{S}$, in the sense of Kleene [20, p. 233]. To see this, note that the method for deriving the set of formulas $\mathcal{F}$ from $\mathcal{E}$ is independent of the constituents of $\mathcal{E}$. That is, for a string $\gamma \in T$, the derivation of $\varphi^*(\gamma)$ and hence F^γ is constructive and does not depend upon the properties of $\mathcal{E}$. Furthermore, if a sentence γ is a logical consequence of $\mathcal{E}$, i.e. F^γ is provable from $\mathcal{F}$ by our theorem, then this proof procedure uses only the rules of inference of the predicate calculus, and consequently is independent of the nature of $\mathcal{F}$ and F^γ. Therefore, the complete procedure for making the passage from $\mathcal{E}$ and γ to $\mathcal{F}$ and F^γ, and the subsequent proof of F^γ from $\mathcal{F}$ is independent of $\mathcal{E}$. Hence, the enumeration procedure is uniformly recursive over $\mathcal{S}$ and thus the logical consequences of a set of sentences $\mathcal{E}$ are recursively enumerable by it.

Even though a method of proof for a language exists, there may be no constructive procedure for determining the actual rules of interference which it comprises. That our proof does not give a constructive way to find such procedures results from the use of the axiom of choice in the proof of Gödel's completeness theorem. Furthermore, the method need not involve any of the commonly accepted rules, such as modus ponens, which are associated with ordinary logic. Nevertheless the notions of provability and logical consequence become synonymous, even though in any particular case one may not be able to find them.

III. OBSERVATIONS

A basic concept of science is that of an observation. Observations are the means by which scientific theories are either confirmed or refuted; they provide the connection between the reality of the world around us and the abstract statements of our theories. In science, we often think of observations as resulting from the performance of some experiment. More generally, however, individuals make observations of the events and phenomena they continually perceive, not necessarily in connection with some well-defined experiment. We shall be concerned with characterizing observations in this broader sense, and in this chapter we will develop a precise notion of observation and consider its relationship to formal languages.

Suppose, for example, that an individual points to some collection of physical objects and exclaims, "Look, the cup is on the table". Is this what we mean by an observation? Although it may be tempting to say yes, more careful examination indicates that this is merely the observer's interpretation of the relationship existing among certain objects which he perceives. The words the observer uses to express what he sees are determined by his language. We shall treat observations as extra-linguistic phenomena, and therefore they will be independent of any particular formal language. Notice that this permits two individuals to observe the same event and each describe the occurrence differently.

We have made the basic ontological assumption that the universe is set-theoretic in nature, and we will make further use of this assumption in formulating our definition of observation. Thus, we shall presume that any perceivable relation among objects in the universe can be expressed within the language of set theory. This is not to say that an individual translates what he sees, either consciously or unconsciously, into set-theoretic notions, but rather that his perception can be expressed in these terms at the meta-level. Now, assuming that observations can be defined somehow within the framework of set theory, are there any restrictions on the character of this definition? One reasonable constraint is to require that a single observation include only a finite number of objects. Notice that this does not presuppose anything about the complexity of the relationships that may be perceived among these objects.

Before presenting our definition of observation, let us briefly examine some of the conditions which have traditionally been accepted as restrictions on what is actually observable. Russell [34], for example, maintains that it is impossible to perceive that "one of these roads leads to Rome", or that "either John or Bob is over six feet tall". The problem here is the disjunction implicit in these observations; thus, in the first case, we can only perceive that some specific road leads to Rome. Similarly, quantification and negation appear to introduce problems. To observe that "the book is not red", for instance, would not be possible since in order to do this we must perceive that the book is, say, blue. Also, the observation "all men are mortal" would be ruled out on the grounds that we can never actually observe all men. We do not wish to take direct issue with these arguments of Russell's, since they are intuitively plausible, at least in the case of normal human observers. On the other hand, they are not germain to the analysis we shall present. Therefore, we shall not be bound by these restrictions, and we will consider observations as a somewhat more general phenomenon. By so doing, we leave open the possibility that an observer could make an "observation" by receiving a communication from some individual whom he considers to be very reliable. Also, we might visualize some non-human observer such as a computer whose sensory inputs and perceptive capabilities are different from those of humans.

The following definition associates a formula and a collection of objects with an observation, and these, in conjunction, act to characterize a set of models.

<u>Definition</u>: An observation $O = \langle F(x_1, \cdots, x_n), a_1, \cdots, a_n \rangle$ where $F(x_1, \cdots, x_n)$ is a formula of the language of set theory and $a_1, \cdots, a_n$ are objects, i.e., $a_i \in S$.

Intuitively we think of the observation $\langle F(x_1, \cdots, x_n), a_1, \cdots, a_n \rangle$ as the perception that the objects $a_1, \cdots, a_n$ are in the relationship specified by the formula $F(x_1, \cdots, x_n)$. For a given model or configuration of the universe, either the objects $a_1, \cdots, a_n$

will have the structure required by $F(x_1, \cdots, x_n)$ or they will not, according as the relativized formula $F_M(x_1, \cdots, x_n)$ holds or does not hold for $a_1, \cdots, a_n$. Therefore, let

$$\mathfrak{M}_O = \{M \mid F_M(x_1, \cdots, x_n) \text{ holds for } a_1, \cdots, a_n\}$$

be the set of models, $\mathfrak{M}_O \subseteq \mathfrak{M}$, associated with the observation O. Note particularly that the formula $F(x_1, \cdots, x_n)$ may contain other variables $y_1, \cdots, y_k$ which are not free in F but are bound by quantifiers. Thus, the x's correspond explicitly to specific objects, while the y's are implicitly associated with related objects, not explicitly perceived. We emphasize that $\mathfrak{M}_O$ is not identical to the set of models satisfying the formula $\exists x_1 \exists x_2 \cdots \exists x_n F(x_1, \cdots, x_n)$. For any model, this formula merely asserts the existence of some arbitrary collection of objects in that model related in the manner specified by $F(x_1, \cdots, x_n)$. The models in $\mathfrak{M}_O$, on the other hand, are required to have this relationship among particular objects, which themselves may not be able to be characterized set-theoretically. Thus, it is only the relationship that is observed to exist among objects which is set-theoretically specified. The objects actually observed in that relationship are a quite separate aspect of the observation.

Suppose $M_0 \in \mathfrak{M}$ is the true model, i.e., M_0 represents the actual configuration of the objects in S. Normally, we would expect that for an observation O, $M_0 \in \mathfrak{M}_O$; however, this is not guaranteed in any way by our definition since we have not required that every observation be true of the real world. We are thus free to speak of possible observations, as for example in considering the possible outcomes of an experiment.

If we have several observations, then under what circumstances would we say that they are consistent? Intuitively we feel that this is so if there are possible configurations of the universe for which every observation is true. We state this condition formally.

Definition: A set of observations Ω is consistent if

$$\bigcap_{O \in \Omega} \mathfrak{m}_O \neq \phi \quad .$$

Suppose that an observer makes a very large number of observations which are consistent. Each new observation he makes, which is consistent with his previous experience, delimits a smaller set of models or possible configurations of the universe. Eventually, at least in this hypothetical case, he arrives at a single model $M \in \mathfrak{m}$. If this is the true model M_0, then he "knows" the complete structure of S.

Theorem: A maximal set of consistent observations Ω defines a unique model $M \in \mathfrak{m}$.

Proof: Suppose this is not the case, i.e., let Ω define $\mathfrak{n} \subseteq \mathfrak{m}$ where $M, M' \in \mathfrak{n}$. Thus, $M = < S, \varepsilon >$ and $M' = < S, \varepsilon' >$, where the binary relations ε and ε' are not identical. Hence, there are objects a, $b \in S$ such that "$a \; \varepsilon \; b$" and "$\neg a \; \varepsilon' \; b$". Therefore the observation

$O = \langle F(x_1, x_2), a, b \rangle$ where $F(x_1, x_2) \equiv x_1 \varepsilon x_2$ is such that $M \in \mathfrak{M}_O$ and $M' \in \mathfrak{M}_O$. Clearly, the set of observations $\Omega \cup \{O\}$ is consistent. This is a contradiction, therefore Ω defines a unique model.

If we have two sets of observations, perhaps associated with different observers, we would like to know under what conditions these two sets of observations can be considered to be related. The concept of relatedness is independent of consistency and only depends upon whether the two sets of observations have in common some specific objects, together with particular structural relations among them.

Definition: Two sets of observations Ω_1, Ω_2, are related if either

(i) The set of observations $\Omega_1 \cup \Omega_2$ is not consistent.

(ii) There are non-empty Ω_1', Ω_2', such that $\Omega_1' \subseteq \Omega_1$, $\Omega_2' \subseteq \Omega_2$ and

$$\bigcap_{O_1 \in \Omega_1'} \mathfrak{M}_{O_1} \subseteq \bigcap_{O_2 \in \Omega_2'} \mathfrak{M}_{O_2} ,$$

or vice-versa.

Condition (i) of this definition states that two sets of observations are related if they are in basic conflict. This seems intuitively reasonable since they will not disagree unless one requires the existence of a structure among some set of objects which the other refutes. Thus, they will be related by virtue of the common objects they reference, even though they may not agree on the structural interrelationships among them. Condition (ii) says that if two sets of observations are consistent, they are related only if some subset of one necessarily implies that certain of the observations of the other are valid. Notice that this concept is very similar to that of logical consequence. However, since we know nothing about whether these observations are expressible in a language, that terminology is inapplicable. Suppose we observed that "grass is green" and that "snow is white". These two observations are clearly consistent but are unrelated and might be said to be independent of one another. This suggests that independence of sets of observations is the converse of relatedness. It is worth mentioning that if the sets of observations each consist of many distinct elements, then they may either be related in a trivial way or a large percentage of the individual observations may be correlated.

IV. PROBABILITY AND INFORMATION

Using the concept of an observation, together with our definition of formal language, we shall now develop an information measure. As previously indicated, a formal language will be considered to characterize an individual's view of some particular domain, and thus the information measure we obtain will be subjective in nature. This notion of a non-objective definition of information is in sharp contrast with the classical information theoretic approach in which the purpose is to develop methods for studying the problems of communication networks and the coding of signals [37, 19]. Thus, in information theory, the informational content of a sentence from today's newspaper might be equal to that of a sentence stating Fermat's last theorem on the basis that the relative probability of occurrence of the groups of letters contained in each was the same. Nevertheless, we shall make use of many of the criteria established by information theory in formulating our definition of a subjective measure of information. The fundamental difference in our approach involves the characterization of the probability. We shall be concerned with probability primarily in the sense of degree of belief, and we will demonstrate that such a probability arises naturally from considerations of the manner in which a formal language structures its universe of discourse.

We have previously introduced the notion that $\mathfrak{M}$, the collection of models of set theory, represents all possible configurations of the universe. That is, given only the basic ontology of set theory, each model $M \in \mathfrak{M}$ expresses a possible structure among the objects of the universe, i.e. the elements of S. However, a given formal language will impose additional structure on $\mathfrak{M}$ in a manner reflecting the logic of that language. Suppose, for example, that two models $M_1, M_2 \in \mathfrak{M}$ differ only with regard to the properties of some object b. Then if $\mathcal{L}$ is a formal language which does not refer to b, either explicitly or implicitly, we expect that M_1 and M_2 will be identified as equivalent configurations of S relative to $\mathcal{L}$.

At this point, it would perhaps be useful to clarify the notion of the "logic" of a language and its relation to provability. A fundamental concept in this regard is that of implication. Unfortunately, however, this seemingly intuitive idea has been modified, extended, and restricted in a myriad of differing ways in the literature. Part of the confusion is due to the fact that we are dealing with several languages—namely the meta-language, the language of set theory, and our object languages or formal languages—each of which may possess a different kind of implication. Suppose, for example, that the symbol "$\supset$" occurs in the object language. That is, $A \supset B$ is a legitimate sentence of the object language, where A and B are also sentences. Then, "$\supset$" is material implication if the sentence $A \supset B$ is true if and only if either B is true or A is false. Another possibility is that "$\supset$" is strict implication. This notion is not well defined in the literature but is generally associated with modal logic.

It is usually construed to mean that the interpretation of $A \supset B$ involved some substantive relation between the interpretations of A and B. Thus, one can not decide the interpretation of $A \supset B$ by examining either A or B alone. In English, we might say that, "Grass is green" materially implies "Snow is white", but this would not be an example of strict implication since these are presumably unrelated truths.

Now consider the meta-linguistic statement A "implies" B, where A and B are sentences of the object language. We stress that this in no way presupposes the existence of an object language symbol for implication of any kind. Again, there are several possible interpretations of this statement. One of the most common is analogous to our definition of logical consequence; that is, B is true of all models where A is true. This is frequently written as $A \models B$. However, we might also mean that B is somehow provable from A. The notion of proof is generally assumed to be syntactic in nature and consequently only involves some process of symbol manipulation. The proof procedure must be specified by some set of inference rules describing how a proof may be constructed, but these rules need not be based on implication, or such familiar inference rules as <u>modus ponens</u>. The statement A "implies" B in this sense of proof is often written as $A \vdash B$. Our second condition of adequacy for a formal language essentially requires that these two notions of implication coincide relative to the meta-language. Hence, for our formal languages, $A \models B$ if and only if $A \vdash B$.

Since these two notions are thus equivalent for formal languages, one can characterize the logic of a formal language either semantically or syntactically. Notice that a rich interrelationship structure may exist among the sentences of a language independent of which possible state of the universe entails. Implication of the type $A \models B$ embodies these relationships. Formal languages then are precisely those languages where this underlying logic is constructively definable in terms of symbol manipulation, or formal syntaxes.

We now wish to investigate in detail how a formal language imposes structure on its universe of discourse. The fundamental idea here is that the syntax and semantics of the language together with the interpretation φ establish a correspondence between sentences of the language and certain sets of models contained in $\mathfrak{M}$. The result of this is a partitioning of $\mathfrak{M}$ into disjoint subsets each of which is definable by some collection of formulas of set theory. These partition sets, however, do not have equivalent status with respect to the formal language, even though they are all equally well-defined at the meta-level. To explicate this distinction, we introduce the two basic notions of an element and a state of a formal language. Our definitions are given in terms of the meaning or truth value of a sentence on a model $M \in \mathfrak{M}$. Recall that given a sentence $\gamma \in \mathcal{S}$, then for any model $M \in \mathfrak{M}$, γ is either true, false or meaningless on M.

Definition: Given a set $\mathcal{S}' \subseteq \mathcal{S}$ of sentences of $\mathcal{L}$, the element $\mathfrak{n}$ determined by $\mathcal{S}'$ is the maximal set of models such that for $M_1, M_2 \in \mathfrak{n}$, $\gamma \in \mathcal{S}'$ either:

(1) γ is true for both M_1 and M_2,

(2) γ is false for both M_1 and M_2,

(3) γ is meaningless for both M_1 and M_2.

Definition: Given a set $\mathcal{S}' \subseteq \mathcal{S}$ of sentences of $\mathcal{L}$, the state $\mathfrak{n}$ determined by $\mathcal{S}'$ is the maximal set of models such that for M_1, $M_2 \in \mathfrak{n}$, $\gamma \in \mathcal{S}'$ either:

(1) γ is true for both M_1 and M_2,

(2) γ is false for both M_1 and M_2

We will say that the set of sentences $\mathcal{S}'$ defines the state $\mathfrak{n}$.

It follows immediately from these definitions that each state is also an element, but in general not every element is a state. This fundamental distinction will be discussed in detail subsequently, but first we identify two common instances of a state.

Lemma: For any language $\mathcal{L}$, the entire set of models $\mathfrak{M}$ is a state, and if negation is expressible within $\mathcal{L}$, the empty set of models ϕ is also a state.

Proof: In the definition of a state let $\mathcal{S}' = \phi$, the null set of sentences. $\mathcal{S}'$ clearly defines $\mathfrak{m}$. Hence $\mathfrak{m}$ is a state. Now assume that if γ is a sentence of $\mathcal{L}$, then $\neg\gamma$ (the negation of γ) is also a sentence of $\mathcal{L}$. Let $\mathcal{S}' = \{\gamma, \neg\gamma\}$. Obviously, $\mathcal{S}'$ defines the empty state ϕ .

The concepts of element and state are fundamental to an understanding of the material that follows. We will therefore attempt to convey their intuitive meanings. The speaker of a formal language deals exclusively with states. A state is a collection of models which can be directly characterized or defined by enunciating some set of sentences of the language. These sentences thereby specify a configuration of the universe which can be recognized or described within the speaker's language, and for any given model $M \in \mathfrak{m}$, either the structure specified by the sentences exists in that model or it does not. Consider the set consisting of all the states of a language; it defines every configuration of the universe which the language can express.

Notice that it is possible that two distinct sets of sentences, $\mathcal{S}'$ and $\mathcal{S}''$ define the same state $\mathfrak{n}$. In order to characterize this case conveniently, we introduce the notion of logical equivalence. Thus, two sets of sentences $\mathcal{S}'$ and $\mathcal{S}''$ will be said to be logically equivalent if they define the same state. Recalling our previous discussion of implication, this clearly means that every sentence of $\mathcal{S}'$ is a logical consequence of $\mathcal{S}''$, and vice-versa. Furthermore, our second condition of adequacy for a formal language requires that the sentences of either set be provable from the sentences of the other.

An element of a language is an extra-linguistic concept, since in general the speaker of the language has no precise way of characterizing it. This is because some of the sentences which are used to define the element may be meaningless on all of the models which constitute the element. At the meta-level, of course elements and states are both defined by formulas of extended set theory and are therefore conceptually the same. But the speaker of a formal language can only assert the truth or falsity of a sentence, not that a sentence is meaningless. We shall make use of the definition of the elements of a language to simplify the notions of probability and information, but it is important to bear in mind that elements are not generally describable within the formal language.

Suppose we now consider two states of some formal language $\mathfrak{L}$. That is, we enunciate the sentences $\mathcal{S}'$ and $\mathcal{S}''$, where $\mathcal{S}'$ defines the state $\mathfrak{h}'$ and $\mathcal{S}''$ defines the state $\mathfrak{h}''$. We then ask a speaker of the language $\mathfrak{L}$ which of these two configurations of the universe, or states, he considers most likely to describe the actual situation. His answer to this question should reflect his subjective view of the relative probabilities of the two states. If we imagine that we could continue this process for all pairs of states describable in his formal language, it seems that we would then have characterized, at least roughly, his view of the world relative to the formal language. Savage [35] approaches the question of probability in a similar fashion and defines what he calls a personalistic probability, which is used to formulate a theory of decision making. Now what is the basis for our speaker's answers? Certainly they are affected by previous experience or observations, but we shall be most interested in his answer to questions for which he has no directly applicable observational data. His assignment of a degree of belief or probability to two states which can not be discriminated on the basis of previous observations will be indicative of his metaphysical assumptions about the world.

In general we can conceive of a set of observations which would be true on exactly those models which are defined by some set of sentences of a formal language. That is, the models satisfying the observations coincide with some state. It therefore seems quite natural to assume that an observer can assign some probability to an observation. This can be interpreted in the sense of expectation, i. e. the likelihood of observing a certain structure among objects in the universe. We have stressed that an observation is independent of language. Hence, the assignment of a probability to an observation is not directly related to the observer's ability to express that observation in some formal language. Rather, it specifies the observer's metaphysics and could be considered to be a part of his ontology. We shall now state in precise terms our assumption that an observer can assign some probability to an observation.

Let $\mathcal{A}$ be the class of all observations, that is, the class of all sets definable by formulas of extended set theory. For $\mathfrak{m}_O \in \mathcal{A}$, $\mathfrak{m}_O = \{M \mid$ for some formula $F(x_1, \cdots, x_n)$ of extended set theory and some collection of objects $a_1, \cdots, a_n \in S$, $F_M(x_1, \cdots, x_n)$ holds $\}$.

Given the class of sets $\mathcal{A}$, consider the σ-algebra generated by $\mathcal{A}$, denoted by $\overline{\mathcal{A}}$. $\overline{\mathcal{A}}$ is closed under the formation of countable unions and complements, and therefore is also closed under the formation of countable intersections. $\overline{\mathcal{A}}$ is a σ-algebra on $\mathfrak{m}$, since $\mathfrak{m} \in \mathcal{A}$ by taking as the formula F some theorem of set theory. We then postulate the probability P as a measure on the σ-algebra $\overline{\mathcal{A}}$, such that $P(\mathfrak{m}) = 1$. The requirement that P be a measure means that:

(1) $P(\phi) = 0$,

(2) $P(\mathfrak{n}_1) \leq P(\mathfrak{n}_1 \cup \mathfrak{n}_2)$ for $\mathfrak{n}_1, \mathfrak{n}_2 \in \overline{\mathcal{A}}$,

(3) For any sequence of sets $\{\mathfrak{n}_i\}$ such that

$$\mathfrak{n}_j \cap \mathfrak{n}_k = \phi \text{ for } j \neq k, \quad P\left(\bigcup_{i=1}^{\infty} \mathfrak{n}_i\right) = \sum_{i=1}^{\infty} P(\mathfrak{n}_i).$$

Thus, P satisfies the axioms of classical probability theory. When we first introduced the notion of a probability, we referred to the probability assigned to a state of some formal language. In the course of defining this probability, we will make use of the probability P, which will be called the observer's a priori probability. At a minimum, we would like to know that any state is P-measurable, or measurable in the Carathéodory sense. Every member of $\overline{\mathcal{A}}$ is measurable in this sense, therefore it is sufficient to ensure that any state, for an arbitrary formal language, belongs to $\overline{\mathcal{A}}$.

<u>Lemma:</u> Given a formal language $\mathcal{L}$ and a state $\mathfrak{h}$ of $\mathcal{L}$, $\mathfrak{h} \in \overline{\mathcal{A}}$ and hence is P-measurable.

Proof: Let $\mathcal{S}' \subseteq \mathcal{S}$ be the set of sentences defining $\mathfrak{h}$. $\mathcal{S}'$ is clearly countable. For each $\gamma \in \mathcal{S}'$, let $F^{\gamma}(\overline{x}, y)$ be the formula of extended set theory corresponding to γ, where $\overline{x} = \overline{a}$, the set of elements of S denoted by the referent words of γ, and $y = 0$ if γ is false of every $M \in \mathfrak{h}$; $y = 1$ if γ is true of every $M \in \mathfrak{h}$. Let

$$\mathfrak{m}_F\gamma = \{M \mid F^{\gamma}_M(\overline{x}, y) \text{ holds for } \overline{x} = \overline{a},\ y \text{ as above}\}.$$

By definition, $\mathfrak{m}_F\gamma \in \overline{\mathcal{A}}$ for each $\gamma \in \mathcal{S}'$, and therefore $\bigcap_{\gamma \in \mathcal{S}'} \mathfrak{m}_F\gamma \in \overline{\mathcal{A}}$. But $\mathfrak{h} = \bigcap_{\gamma \in \mathcal{S}'} \mathfrak{m}_F\gamma$ by the definition of state. Hence $\mathfrak{h} \in \overline{\mathcal{A}}$ and is P-measurable.

Having established the concept of an observer's a priori probability, we now wish to define a relation which is fundamental to an understanding of the structure imposed by a formal language on its universe of discourse. The question is, under what conditions can a language distinguish between two different configurations of the universe? That is, any two distinct models actually represent different configurations of the universe, but it need not be the case that a given formal language can articulate this difference.

<u>Definition</u>: Given a formal language $\mathfrak{L}$, and sets of models $\mathfrak{n}_1$, $\mathfrak{n}_2 \subseteq \mathfrak{m}$, $\mathfrak{n}_1$ and $\mathfrak{n}_2$ are distinguishable $(\mathfrak{n}_1/\mathfrak{n}_2)$ if there is a sentence $\gamma \in \mathfrak{S}$ such that γ is true of $\mathfrak{n}_1$ and false of $\mathfrak{n}_2$. Otherwise, $\mathfrak{n}_1$ and $\mathfrak{n}_2$ are indistinguishable $(\mathfrak{n}_1 /\!\!/ \mathfrak{n}_2)$.

Thus, two sets of models are indistinguishable if every sentence of the language either has the same set of truth values on both of them, or is true (or false) of one but meaningless on the other. In the same sense that a single model $M \in \mathfrak{m}$ represents a configuration of the universe, an arbitrary set of models $\mathfrak{n} \subseteq \mathfrak{m}$ represents a partially specified configuration, or a group of configurations having in common exactly those structures which occur in every $M \in \mathfrak{n}$. Distinguishability requires that a sentence be able to express a difference between two partially specified configurations. We now state without proof some simple properties of distinguishability and indistinguishability.

(1) If $\mathfrak{n}_1/\mathfrak{n}_2$ and $\mathfrak{n}_1' \subseteq \mathfrak{n}_1$, then $\mathfrak{n}_1'/\mathfrak{n}_2$,

(2) If $\mathfrak{n}_1 /\!\!/ \mathfrak{n}_2$ and $\mathfrak{n}_1 \subseteq \mathfrak{n}_1'$, then $\mathfrak{n}_1' /\!\!/ \mathfrak{n}_2$,

(3) For any $\mathfrak{n} \neq \emptyset$, $\mathfrak{n} /\!\!/ \mathfrak{m}$,

(4) For any $\mathfrak{h}$, $\mathfrak{h}/\phi$,

(5) If $\mathfrak{h}_1 \cap \mathfrak{h}_2 \neq \phi$, then $\mathfrak{h}_1 /\!/ \mathfrak{h}_2$.

Perhaps contrary to expectations, the relation of indistinguishability is not an equivalence relation. It is clearly reflexive—$\mathfrak{h} /\!/ \mathfrak{h}$ by property (5)—and just as obviously symmetric. It is not, however, guaranteed to be transitive. That is, if $\mathfrak{h}_1 /\!/ \mathfrak{h}_2$ and $\mathfrak{h}_2 /\!/ \mathfrak{h}_3$, then it may be the case that $\mathfrak{h}_1 / \mathfrak{h}_3$. To give a simple example, let $\mathfrak{h}_2 = \mathfrak{m}$. By property (3), it is certain that $\mathfrak{h}_1 /\!/ \mathfrak{h}_2$ and $\mathfrak{h}_2 /\!/ \mathfrak{h}_3$, for any $\mathfrak{h}_1$ and $\mathfrak{h}_3$. Thus if indistinguishability were a transitive relation, then any two sets of models (denoted here by $\mathfrak{h}_1$ and $\mathfrak{h}_3$) would be indistinguishable, which is patently false.

We have said previously that the collection of all of the states of a language characterizes every configuration of the universe expressible within the language. However, these configurations (and the associated states) are not independent of one another. We would like to establish an independent set of states such that the conditions postulated by a given state preclude the possibility of being in any other state. To this end we make the following definition.

<u>Definition:</u> Let $\mathfrak{S}$ be the set of all states of a language $\mathcal{L}$. A state $\mathfrak{h} \in \mathfrak{S}$ is a minimal state if there is no other state $\mathfrak{h}' \in \mathfrak{S}$, $\mathfrak{h}' \neq \phi$, such that $\mathfrak{h}' \subseteq \mathfrak{h}$.

<u>Definition:</u> Given sets of models $\mathfrak{h}_1$, $\mathfrak{h}_2$ and a set of sentences $\mathcal{E}$, $\mathcal{E}$ agrees on $\mathfrak{h}_1$, $\mathfrak{h}_2$ if for any $\gamma \in \mathcal{E}$, and for any $M_1 \in \mathfrak{h}_1$, $M_2 \in \mathfrak{h}_2$, γ is meaningful on M_1 and M_2, and $\varphi^*_{M_1}(\gamma) = \varphi^*_{M_2}(\gamma)$.

We now prove that given any consistent set of sentences, there is a minimal state on which those sentences have the same truth value.

<u>Theorem</u>: Given a language $\mathcal{L}$, and a set of sentences $\mathcal{H} \subseteq \mathcal{S}$, each of which is meaningful for some model $M \in \mathfrak{M}$, then there is a minimal state $\mathfrak{h}$, such that $\mathcal{H}$ agrees on $\{M\}$, $\mathfrak{h}$.

Proof: Let $t = \{\mathcal{G} \mid \mathcal{H} \subseteq \mathcal{G} \subseteq \mathcal{S}$ and there exists an $M' \in \mathfrak{M}$ such that $\varphi^*_{M'}(\gamma) \neq \phi$ for $\gamma \in \mathcal{G}$, and $\varphi^*_M(\gamma) = \varphi^*_{M'}(\gamma)$ for $\gamma \in \mathcal{H}\}$:

(1) $t \neq \phi$ since $\mathcal{H} \in t$,

(2) Let C be a chain of elements of t under set inclusion. Consider $\bigcup_{\mathcal{G} \in C} \mathcal{G} = \mathcal{G}'$. Clearly $\mathcal{G}' \subseteq \mathcal{S}$.

We must show that there is an $M \in \mathfrak{M}$ such that $\varphi^*_M(\gamma) \neq \phi$ for $\gamma \in \mathcal{G}'$.

For $\gamma \in \mathcal{G}'$, $\gamma = \gamma_1\gamma_2 \cdots \gamma_h$, there is a formula $F^\gamma(\overline{x})$ such that $<\varphi_M(\gamma_i), \cdots, \varphi_M(\gamma_h)>$ satisfies $F^\gamma_M(x_1, \cdots, x_h)$ if and only if $\varphi^*_M(\gamma) \neq \phi$, by a preceding corollary.

Now, we have previously assumed that within the language of set theory, we have a "proper name" for each element of S in the range of φ, i.e. for each element $\{\varphi(\gamma) \mid \gamma \in T\}$. Let a_γ "name" $\varphi(\gamma)$, and

$$E^{\gamma}(a_{\gamma_1}, \cdots, a_{\gamma_h}) \equiv F^{\gamma}(a_{\gamma_1}, \cdots, a_{\gamma_h}) \wedge a_{\gamma_1} \Delta a_{\gamma_2} \wedge a_{\gamma_1} \Delta a_{\gamma_3}$$

$$\wedge \cdots \wedge a_{\gamma_{h-2}} \Delta a_{\gamma_h} \wedge a_{\gamma_{h-1}} \Delta a_{\gamma_h} ,$$

where $a_{\gamma_i} \Delta a_{\gamma_j}$ is $\begin{cases} a_{\gamma_i} = a_{\gamma_j} & \text{if} \quad \varphi(\gamma_i) = \varphi(\gamma_j) \\ a_{\gamma_i} \neq a_{\gamma_j} & \text{if} \quad \varphi(\gamma_i) \neq \varphi(\gamma_j) \end{cases}$.

Thus $E^{\gamma}_M(a_{\gamma_1}, \cdots, a_{\gamma_h})$ holds if and only if $< \varphi_M(\gamma_1), \cdots, \varphi_M(\gamma_h) >$ satisfies $F^{\gamma}_M(x_1, \cdots, x_h)$. Now, let $\mathcal{E}$ be the following set of formulas of extended set theory.

$$\mathcal{E} = \{E^{\gamma}(a_{\gamma_1}, \cdots, a_{\gamma_h}) \mid \gamma \in \mathcal{J}'\} .$$

Suppose there were no model which satisfies all of the sentences (of the language of set theory) of $\mathcal{E}$. Then by the completeness theorem for set theory, there would be a finite subset of $\mathcal{E}$, $E^{\gamma_1}, \cdots, E^{\gamma_k}$ with no model. But $\gamma_i \in \mathcal{J}^i \in C$, $i = 1, \cdots, k$ and C is simply ordered. Therefore, for some $\mathcal{J}'' \in C$, $\gamma_i \in \mathcal{J}''$, $i = 1, \cdots, k$ and hence there is a model M' such that $\varphi^*_{M'}(\gamma) \neq \phi$ for $\gamma \in \mathcal{J}''$. Therefore, $E^{\gamma_i}(a_{\gamma_{i1}}, a_{\gamma_{i2}}, \cdots, a_{\gamma_{ih_i}})$ $i = 1, \cdots, k$ holds in M' . Thus with the interpretation of a_{γ} as $\varphi(\gamma)$, these $E^{\gamma_1}, \cdots, E^{\gamma_k}$ have a model M' . This is a contradiction. Therefore, $\mathcal{E}$ has some model, say M''.

Now, $\mathcal{E}$ has a model M'', but we cannot guarantee that the interpretations of the a_{γ}'s are $\varphi(\gamma)$'s in this model. That is, we are only assured of this for the a_γ's occurring in $E^{\gamma_1}, \cdots, E^{\gamma_k}$. However, we can get a permutation ρ of S which carries the interpretations of all a_γ's in M'' into $\varphi(\gamma)$'s. Then let ρ induce an ε corresponding to ε'' of M''; the resulting model $M = < S, \varepsilon >$ is the desired one such that, for each $\gamma \in \mathcal{G}'$, $E^\gamma(\bar{a}_\gamma)$ holds in M, and hence $<\varphi_M(\gamma_1), \cdots, \varphi_M(\gamma_h)>$ satisfies $F^\gamma_M(x_1, \cdots, x_h)$. Therefore $\varphi^*_M(\gamma) \neq \phi$ for $\gamma \in \mathcal{G}'$.

Thus, $\mathcal{G}'$ is an upper bound in t for the chain C, and by Zorn's lemma, t has a maximal element $\mathcal{G}$. We now claim that

$$\mathfrak{h} = \{M \mid \varphi^*_{M'}(\gamma) = \varphi^*_M(\gamma), \text{ for } \gamma \in \mathcal{G}\}$$

is a minimal state of $\mathcal{L}$, i.e. there is no non-empty state $\mathfrak{h}'$ of $\mathcal{L}$ such that $\mathfrak{h}' \subseteq \mathfrak{h}$ and $\mathfrak{h}' \neq \mathfrak{h}$. Suppose $\mathfrak{h}$ is not a minimal state. By the definition of state $\mathfrak{h}$ is clearly a state; therefore $\mathfrak{h}$ must not be minimal. Hence, there must be a $\gamma_o \notin \mathcal{G}$ such that the set

$$\left\{ M' \mid \varphi^*_{M'}(\gamma) = \varphi^*_M(\gamma) \text{ for } \gamma \in \mathcal{G} \cup \{\gamma_o\} \right\}$$

is not empty. Therefore, $\mathcal{G} \cup \{\gamma_o\}$ has a model, contradicting the hypothesis that $\mathcal{G}$ is a maximal element of t. Hence, $\mathfrak{h}$ is a minimal state of $\mathcal{L}$, and since $\mathcal{H} \subseteq \mathcal{G}$, $\mathcal{H}$ agrees on $\{M\}$, $\mathfrak{h}$.

Since the set of sentences $\mathbb{S}$ of a formal language may be countably infinite, the collection of minimal states is in general non-denumerably infinite. This is because any subset of $\mathbb{S}$ potentially defines a unique state. The minimal states of a formal language represent the most precisely specified configurations of the universe expressible within the language. However, it is not the case that every minimal state is necessarily distinguishable from every other. Consequently, the collection of minimal states is not the independent set of states we are seeking. In order to obtain an independent set of minimal or atomic states, we introduce the notion of a basis for a formal language.

Definition: A set of minimal states $\mathfrak{B}$ of a formal language $\mathfrak{L}$ is a basis for $\mathfrak{L}$ if for any $\mathfrak{n}_1, \mathfrak{n}_2 \in \mathfrak{B}$, $\mathfrak{n}_1 / \mathfrak{n}_2$ and for any element $\mathcal{E}$ of $\mathfrak{L}$, there is some $\mathfrak{n} \in \mathfrak{B}$ such that $\mathfrak{n} // \mathcal{E}$.

Thus, the basis is a set of minimal states such that any collection of models $\mathfrak{n} \subseteq \mathfrak{m}$ is indistinguishable from at least one member of the basis. We now prove that it is always possible to choose a basis for any formal language $\mathfrak{L}$.

Theorem: Given a formal language $\mathfrak{L}$, for any set of minimal states $\mathfrak{S}$ of $\mathfrak{L}$ which are pairwise distinguishable, there is a basis $\mathfrak{B}$ for $\mathfrak{L}$ such that $\mathfrak{S} \subseteq \mathfrak{B}$.

Proof: Let $\mathcal{K}$ be the set of minimal states of $\mathfrak{L}$. $\mathcal{K} \neq \emptyset$ by the preceding theorem. Then let

$$p = \{ \mathcal{G} \mid \mathfrak{S} \subseteq \mathcal{G} \subseteq \mathcal{K}, \text{ and for } \mathfrak{n}', \mathfrak{n}'' \in \mathcal{G},\ \mathfrak{n}'/\mathfrak{n}'' \} .$$

(1) $p \neq \emptyset$ since $\mathfrak{S} \in p$.

(2) Let C be a chain of elements of p under set inclusion, and consider $\bigcup_{\mathcal{G} \in C} \mathcal{G} = \mathcal{G}^+$. Clearly $\mathcal{G}^+ \subseteq \mathcal{K}$. We must now show that for $\mathfrak{h}', \mathfrak{h}'' \in \mathcal{G}^+$, $\mathfrak{h}'/\mathfrak{h}''$. Suppose $\mathfrak{h}' /\!/ \mathfrak{h}''$. Now $\mathfrak{h}' \in \mathcal{G}^i \in C$ and $\mathfrak{h}'' \in \mathcal{G}^j \in C$. Hence, either $\mathcal{G}^i \subseteq \mathcal{G}^j$ or $\mathcal{G}^j \subseteq \mathcal{G}^i$. Assume $\mathcal{G}^i \subseteq \mathcal{G}^j$. Therefore $\mathfrak{h}', \mathfrak{h}'' \in \mathcal{G}^j$ and $\mathfrak{h}'/\mathfrak{h}''$. This is a contradiction. Therefore, $\mathcal{G}^+$ is an upper bound in p for C, and by Zorn's lemma p has a maximal element $\mathfrak{B}$.

We claim that $\mathfrak{B}$ is a basis for $\mathfrak{L}$. We must show that for any element $\mathcal{E}$ of $\mathfrak{L}$, $\mathcal{E}$ is indistinguishable from some $\mathfrak{h}' \in \mathfrak{B}$. Suppose this is not the case, i.e. $\mathcal{E}$ is distinguishable from every $\mathfrak{h}' \in \mathfrak{B}$. Let

$$\mathcal{H} = \{\gamma \mid \gamma \in \mathfrak{S}, \text{ for some } \mathfrak{h}' \in \mathfrak{B}, \ \mathfrak{h}'/\mathcal{E} \text{ by } \gamma\}.$$

By the definition of distinguishability, every sentence in $\mathcal{H}$ is meaningful for any $M \in \mathcal{E}$. By the previous theorem, there is a minimal state $\mathfrak{h}$ such that $\mathcal{H}$ agrees on $\{M\}$ and $\mathfrak{h}$. Hence, $\mathfrak{h}$ is a minimal state which is distinguishable from every $\mathfrak{h}' \in \mathfrak{B}$. This is a contradiction, and therefore any element $\mathcal{E}$ is indistinguishable from some $\mathfrak{h}' \in \mathfrak{B}$. Thus, $\mathfrak{B}$ is a basis for $\mathfrak{L}$ and $\mathfrak{S} \subseteq \mathfrak{B}$.

At this point it is appropriate to indicate some general properties of a basis for a formal language, and to compare them with similar concepts in the literature. First of all, a formal language is not guaranteed to have a unique basis. Actually, this will only occur in certain special cases. More will be said about the implications of this fact for probability and information in a subsequent section.

Corollary: A formal language $\mathfrak{L}$ has a unique basis $\mathfrak{B}$ if and only if the set of all minimal states $\mathcal{K}$ is pairwise distinguishable.

Proof: "if" - By the preceding theorem $\mathfrak{L}$ has a basis $\mathfrak{B} \supseteq \mathcal{K}$. Since $\mathcal{K}$ includes all minimal states $\mathfrak{B} = \mathcal{K}$. "only if" - Suppose there are minimal states $\mathfrak{h}'$, $\mathfrak{h}''$ such that $\mathfrak{h}' /\!/ \mathfrak{h}''$. In the preceding theorem let $\mathfrak{S} = \{\mathfrak{h}'\}$. Thus $\mathfrak{L}$ has a basis containing $\mathfrak{h}'$, but not $\mathfrak{h}''$ by definition. Similarly, $\mathfrak{L}$ has a basis containing $\mathfrak{h}''$ but not $\mathfrak{h}'$. These are clearly distinct bases for $\mathfrak{L}$. Since this is a contradiction, for every $\mathfrak{h}', \mathfrak{h}'' \in \mathcal{K}$, $\mathfrak{h}' / \mathfrak{h}''$.

Several other important properties of a basis for a formal language are directly related to the notion of meaninglessness, as is shown by the following theorem and its corollaries.

Theorem: Let $\mathfrak{L}$ be a formal language. If a set of models $\mathfrak{m}_o \subseteq \mathfrak{m}$ is such that for every sentence $\gamma \in \mathfrak{S}$, γ is meaningful for each $M \in \mathfrak{m}_o$, then for any basis $\mathfrak{B}$ of $\mathfrak{L}$, $\mathfrak{m}_o \subseteq \bigcup_{B \in \mathfrak{B}} B$.

Proof: Consider any basis $\mathfrak{B}$ and some $M \in \mathfrak{M}_o$. Suppose that $M \notin \bigcup_{B \in \mathfrak{B}} B$. Therefore, by the definition of basis, there is some minimal state $\mathfrak{h} \in \mathfrak{B}$ such that $\{M\} // \mathfrak{h}$. Let $\mathcal{S}'$ be the set (possibly empty) of all sentences which are meaningless on $\mathfrak{h}$, i.e. meaningless on some $M' \in \mathfrak{h}$. Now, since $\mathcal{S}'' = \mathcal{S} - \mathcal{S}'$ is a set of sentences each of which is meaningful on both M and $\mathfrak{h}$, $\mathcal{S}''$ must clearly agree on $\{M\}$ and $\mathfrak{h}$ if $\mathfrak{h} // \{M\}$, and since $\mathcal{S}''$ consists of all of the sentences which are meaningful on $\mathfrak{h}$, there is some truth assignment to the sentences of $\mathcal{S}''$ such that $\mathcal{S}''$ defines $\mathfrak{h}$. But $\mathcal{S}''$ agrees on $\{M\}$ and $\mathfrak{h}$; therefore $M \in \mathfrak{h}$ by the definition of a state. This is a contradiction; hence $\mathfrak{M}_o \subseteq \bigcup_{B \in \mathfrak{B}} B$ for any basis $\mathfrak{B}$.

<u>Corollary</u>: Let $\mathfrak{L}$ be a formal language with a basis $\mathfrak{B}_o$. If for every sentence $\gamma \in \mathcal{S}$, γ is meaningful for every $M \in \bigcup_{B \in \mathfrak{B}_o} B$, then $\mathfrak{B}_o$ is a unique basis for $\mathfrak{L}$.

Proof: In the preceding theorem, let $\mathfrak{M}_o = \bigcup_{B \in \mathfrak{B}_o} B$. Therefore, $\bigcup_{B \in \mathfrak{B}_o} B \subseteq \bigcup_{B \in \mathfrak{B}} B$ for any basis $\mathfrak{B}$, and since any $\mathfrak{B}$ is a maximal set of pairwise distinguishable minimal states, $\bigcup_{B \in \mathfrak{B}_o} B = \bigcup_{B \in \mathfrak{B}} B$. Further, since the set of minimal states contained in any set of models is unique, $\mathfrak{B}_o$ is a unique basis for $\mathfrak{L}$.

<u>Corollary:</u> If $\mathcal{L}$ is a formal language such that for every sentence $\gamma \in \mathcal{S}$, γ is meaningful for each $M \in \mathfrak{M}$, then $\mathcal{L}$ has a unique basis $\mathfrak{B}$ such that $\bigcup_{B \in \mathfrak{B}} B = \mathfrak{M}$.

Proof: Immediate from the theorem and the preceding corollary.

Let us examine the condition that every sentence of a formal language be meaningful on each model M. One possibility is to demand that no sentence contain any description such as "the present king of France" which is vacuous. This appears to be unreasonable since there will generally be plausible configurations of the universe in which France has no king or perhaps has more than one king. Of course, it is possible to define a formal language such that every grammatical string of the language is meaningful, but languages of this type will generally be quite primitive and will not provide suitable approximations to the natural language phenomena we wish to explicate. For example, even the commonly used computer languages, which tend to have very simple grammars and are certainly formal languages in our sense, generally admit grammatical but meaningless sentences. Thus, in most versions of FORTRAN we may write the statement Y = X even though the variable X appears nowhere else in our program. The result of this is that Y will be set to some completely arbitrary and unknown value, thus making this statement meaningless. In this case, X can be considered to be a vacuous description since it does not denote a well-defined entity.

Some logicians have proposed technical solutions to the problem of vacuous or non-unique descriptions. Perhaps the most satisfactory of these have been suggested by Quine [29], Frege [14], and Carnap [4]. Their methods are similar in that they assign to any non-unique or vacuous description some particular entity in the model. This has the result that sentences containing such descriptions become meaningful, and thus all grammatical sentences have some associated meaning. Unfortunately, however, the specific meaning assigned to a sentence containing a vacuous description, for example, then depends upon the properties of the entity taken as the denotation of such descriptions. Since the properties of this distinguished entity will vary from one model to another, the truth conditions for a sentence are not structural. That is, the procedure for deriving the truth value of a sentence becomes dependent on factual information, namely the particular attributes of the distinguished entity relative to the model in question.

In our method, the truth conditions remain structural since we explicitly allow grammatical sentences to be semantically meaningless. Indeed, as we have previously mentioned, this appears to be an essential characteristic of most commonly used languages. By referring to our definition of semantic category, we can get a slightly different view of this problem. If we did require that every sentence in our formal languages was to be meaningful on each model, then the semantic categories of the language would necessarily be devoid of structure, since if they were not, the interpretation function φ might assign an entity of the wrong structural class to some referent words, at least in some models. To make the semantic categories this weak is tantamount to assuming that the universe of discourse is homogeneous with respect to the formal language, which contradicts the idea that there are implicit structural distinctions among various classes of objects.

In Carnap's approach, he further requires that a language in his sense be such that the atomic sentences are independent, i.e. that no set of atomic sentences is contradictory. In our case, however, this is much too restrictive since the non-null states of the language exactly reflect the observer's view of consistent sets of sentences within his language. Thus, the fact that some set of sentences defines the empty state implies an inconsistency relative only to the observer. In a different formal language, these same strings of words may be consistent and have a model. Thus, in one formal language the two sentences, "It is freezing" and "It is not cold" may be consistent, while in another formal language they could be inconsistent by virtue of the logic of the language, as reflected in the semantic transformations.

Also important with regard to the literature is the relationship of our formal languages to the concepts of model theory [21, 31]. For example, suppose we have a formal language with referent words $p_1^{n_1}, \cdots, p_m^{n_m}$, where $p_i^{n_i}$ is an n_i-ary predicate symbol. Then if for some model $M = \langle S, \varepsilon \rangle$ of set theory, the interpretation of these symbols is such that

$$\varphi(p_i^{n_i}) \in \left\{ y \,\middle|\, \forall x \left(x \in y \rightarrow \exists u_1 \exists u_2 \cdots \exists u_{n_i} (x = \langle u_1, \cdots, u_{n_i} \rangle) \right) \right\},$$

then $\langle S, \varphi(p_1^{n_1}), \cdots, \varphi(p_m^{n_m}) \rangle$ is a model in the sense of model theory since it comprises some domain of objects or individuals and a set of relations on this domain. Thus, we can easily construct particular formal languages which are equivalent to the languages of model theory—the first-order predicate calculus, for instance—and we can then choose an interpretation φ so that some model of set theory embodies the necessary characteristics of a model of model theory.

Suppose that we do take the first order predicate calculus with predicate symbols $p_1^{n_1}, \cdots, p_m^{n_m}$ as a formal language. Then if we consider $\mathfrak{m}_o$, where the interpretation function φ satisfies the above condition for each model $M \in \mathfrak{m}_o$, the sentences of this language will all be meaningful on every model of $\mathfrak{m}_o$. As shown in a previous corollary, this implies that $\mathfrak{m}_o$ is contained in the union of any basis for the language. Furthermore, as in model theory, any distinct set of atomic sentences will define some non-empty state, and consequently every minimal state will be contained in $\mathfrak{m}_o$. Thus, the basis will be unique and its union will be equal to $\mathfrak{m}_o$, and the minimal states are then the elementary equivalence classes of models exactly in the sense of model theory [31, p. 55].

Although languages of this type are not our primary concern, this reduction to model theory is significant in view of the many powerful results in this area which have contributed to our understanding of the foundations of mathematics.

Having developed the idea of a basis for a formal language, we shall now make use of this in defining a probability for such a language. In terms of classical probability theory, the event space will be the collection of all possible sets of observations and thus will include the states of any given formal language. We wish to be able to assign to any observation a numerical value indicating its probability relative to a formal language. That is, given a language we can determine the set of associated states, and this, together with the a priori probability, should be sufficient to specify the observer's expectation of making some observation O .

Definition: Given a formal language $\mathfrak{L}$ having a basis $\mathfrak{B}$ and an observation O , the *-probability of O is given by

$$P^*(\mathfrak{m}_O) = \frac{\overline{P}\left(\bigcup\limits_{\substack{B \in \mathfrak{B} \\ B /\!/ \mathfrak{m}_O}} B\right)}{\overline{P}\left(\bigcup\limits_{B \in \mathfrak{B}} B\right)}$$

where $\overline{P}$ is the outer measure induced by the observer's a priori probability P .

Note that we have used the outer measure $\overline{P}$, rather than the measure P, to define the *-probability. The reason for this is that if the basis is composed of a non-denumerable collection of basis elements, we can not in general prove that the union of all basis elements is P-measurable. Use of the outer measure $\overline{P}$ simply ensures that the *-probability will be well-defined. For many languages, however, $\overline{P}$ and P will be identical on basis elements. Thus, if we consider the first-order predicate calculus or other similar languages common in model theory, where all sentences are meaningful on the basis, it can easily be shown that the union of all basis elements is P-measurable. It is also easily established that if $\bigcup_{B \in \mathfrak{B}} B$ is P-measurable, then so is $\bigcup_{\substack{B \in \mathfrak{B} \\ B // \mathfrak{m}_O}} B$, where O is any observation.

We see immediately that P^* depends directly upon the notion of indistinguishability. Thus, the *-probability of an observation is directly proportional to the measure of those basis elements from which it is indistinguishable. Recall that the basis can be considered as a collection of independent atomic states of the language, and if an observation can not be distinguished from some basis element, it becomes identified with the configuration expressed by the sentences defining that basis element. The following are easily proved properties of the *-probability.

(1) $P^*\left(\bigcup_{B \in \mathfrak{B}} B\right) = P^*(\mathfrak{m}) = 1$.

(2) $P^*(\mathfrak{m}_{O_1}) \leq P^*(\mathfrak{m}_{O_1} \cup \mathfrak{m}_{O_2})$.

(3) $P^*(\emptyset) = 0$.

However, it is not the case that for $\mathfrak{m}_{O_1} \cap \mathfrak{m}_{O_2} = \emptyset$,

$$P^*(\mathfrak{m}_{O_1} \cup \mathfrak{m}_{O_2}) = P^*(\mathfrak{m}_{O_1}) + P^*(\mathfrak{m}_{O_2}) ,$$

and thus, in general P^* does not satisfy the traditional axioms of probability. Although this is at first disturbing, it is closely related to established results in the literature. Consider the case of languages based upon the lower predicate calculus. If we restrict the notion of state to only those sets of models which are definable by a finite number of sentences, then P^* is additive on those states. However, in the case of states defined by an infinite number of sentences, states that are not "finitely axiomatizable", it is known [39, chap. 12] that their complement is not a state. Thus, we would not expect additivity in these situations. Therefore, even in the case of the lower predicate calculus, when there are more than a finite number of states, P^* is not a probability, since we are lacking the additivity property on the space of models in the sense of model theory. In general, this non-additivity is a direct result of the fact that indistinguishability is not an equivalence relation.

The possibility that $\overline{P}\left(\bigcup_{B \in \mathfrak{B}} B\right) = 0$, and hence that P^* is undefined, has been ignored on the grounds that such a language possesses a set of minimal states which in no way reflects the observer's expectations. The definition of the *-probability was given in terms of a single observation O. In general, however, some set of observations Ω will have been made prior to observing O. Providing that the resulting set of observations, $\Omega \cup \{O\}$, is consistent, we simply take the conditional *-probability to determine the likelihood of observing O given the set of observations Ω, where this conditional probability is given by

$$P^*(\mathfrak{m}_O \mid \mathfrak{m}_\Omega) = \frac{P^*(\mathfrak{m}_O \cap \mathfrak{m}_\Omega)}{P^*(\mathfrak{m}_\Omega)}$$

where $\mathfrak{m}_\Omega$ denotes $\bigcap_{O' \in \Omega} \mathfrak{m}_{O'}$.

Very generally speaking, the *-probability of some observation O may be interpreted as being inversely proportional to the observer's estimate of the rarity of O. Thus, if $P^*(\mathfrak{m}_O) = 1$, the observation O is a certainty in terms of the observer's previous experience and his assumptions concerning plausible configurations of the universe. On the other hand, if the *-probability of O is very small, then O represents to the observer a significant discovery relative to his current beliefs.

Prior to this point, we have considered an observer's formal language to be static in nature. However, it is very likely that as an observer gains experience his formal language will change in some way which is dictated by this experience. That is, new alternatives or states may be introduced and old ones either eliminated or made less probable. Thus, the processes of observation and language change are strongly coupled. The mechanism of this change and the general problem of the dynamics of language are not explicitly treated in this thesis since many additional concepts relating to the rate of acquisition of observational evidence and the temporal aspects of language would be necessary. Nevertheless, at any given time, a formal language can be used to characterize the observer's view of reality, and we shall later indicate that, in many instances, this formal language is inherently quite inflexible and resistant to change.

We shall now turn to the question of information, or what in our sense might more appropriately be called informativeness to reflect its subjective nature. Suppose we have some observer and a formal language which expresses his current viewpoint on some domain of interest to him. In addition, we naturally presume that he has some previous experience as characterized by a set of observations Ω. He now makes a new observation, perhaps as the result of performing some experiment. We wish to know if this observation increases his information, and if it does, by how much in comparison with some other hypothetical observation. We have stressed that the *-probability completely specifies his expectations with respect to all observations. Hence, the information gained should be a function only of this probability, and we define it in a manner analogous to information theory.

Definition: Given a formal language $\mathcal{L}$, and an associated probability P^*, the information gained on observing O given the prior set of observations Ω is

$$I(O|\Omega) = \log P^*(\mathfrak{m}_\Omega) - \log P^*(\mathfrak{m}_\Omega \cap \mathfrak{m}_O) = \log \frac{P^*(\mathfrak{m}_\Omega)}{P^*(\mathfrak{m}_\Omega \cap \mathfrak{m}_O)} .$$

Before discussing the implications of this definition, a few comments are necessary. First, since we are interested only in information in a relative sense, we shall not specify the base of the logarithm. Secondly, we are assuming that the new observation O is consistent with the previous experience Ω, and hence that $\mathfrak{m}_\Omega \cap \mathfrak{m}_O \neq \emptyset$. Finally, we note that by taking $\mathfrak{m}_\Omega = \mathfrak{m}$, i.e. no previous experience, we obtain a measure of the "absolute" information in O relative to $\mathcal{L}$.

The information measure has the following properties:

(1) If $\mathfrak{m}_{O_1} \subseteq \mathfrak{m}_{O_2}$, $I(O_1|\Omega) \geq I(O_2|\Omega)$,

(2) If $\mathfrak{m}_{O_1} \cap \mathfrak{m}_{O_2} = \mathfrak{m}_{O_3}$, then

$$I(O_3|\Omega) = I(O_1|\Omega) + I(O_2|\Omega \cup \{O_1\}) .$$

Property (1) says that if O_1 is a more precise observation than O_2, then the information to be gained by observing O_1 is greater than that gained by observing O_2. As used here, the terminology "more precise" means that the structural relations among the objects of the universe as expressed by O_1, necessarily entails the existence of the structure expressed by O_2. Property (2) is a condition on the additivity of the information. That is, observing O_1 followed by O_2 yields the same gain in information as the simultaneous observation of O_1 and O_2. Taken together, these two properties ensure that our notion of information conforms to the information theoretic definition of the à fortiori information, or the informational gain resulting from the selection of some particular alternative [36, p. 12].

Extending the analogy with information theory somewhat further, suppose we have some finite set of observations $O_1, \cdots, O_n$ such that $\mathfrak{m}_{O_i} \cap \mathfrak{m}_{O_j} = \emptyset$ for $i, j = 1, \cdots, n$ and $i \neq j$. Thus the observations are mutually exclusive or inconsistent. Also, assume that $\bigcup_{i=1}^{n} \mathfrak{m}_{O_i} = \mathfrak{m}$; these observations might therefore represent the n possible outcomes of some experiment. We can then express the informational gain which the observer expects will result from actually performing the experiment.

<u>Definition</u>: Let $\mathcal{E}$ be an experiment with possible outcomes $O_1, \cdots, O_n$ such that $\bigcup_{i=1}^{n} \mathfrak{m}_{O_i} = \mathfrak{m}$ and $\mathfrak{m}_{O_i} \cap \mathfrak{m}_{O_j} = \emptyset$ for $i, j = 1, \cdots, n$ and $i \neq j$. Then the expected gain in information for an observer with formal language $\mathcal{L}$ and probability P^* to perform the experiment $\mathcal{E}$ is

$$I(\mathcal{E} \mid \Omega) = -\sum_{i=1}^{n} P^*(\mathfrak{m}_{O_i} \mid \mathfrak{m}_{\Omega}) \log P^*(\mathfrak{m}_{O_i} \mid \mathfrak{m}_{\Omega})$$

where Ω is the set of prior observations.

In order to facilitate a clearer understanding of the formal apparatus developed in this chapter, we shall now interpret some of the more important concepts diagrammatically. Note that the diagrams we shall present are intended as heuristic aids only and do not constitute an adequate method for dealing with many of the more complex aspects of the problem. We begin by assuming that the space of models $\mathfrak{m}$ of the universe of discourse is arranged on our diagrams in such a way that the a priori probability P of any observation O is proportional to the area encompassed by the set of models $\mathfrak{m}_O$ associated with that observation. Figure 1 illustrates two strictly consistent observations O_1 and O_2 such that $P(\mathfrak{m}_{O_1}) > P(\mathfrak{m}_{O_2})$, i.e. O_1 is more probable than O_2 based on the observer's a priori probability. The shaded area, $\mathfrak{m}_{O_1} \cap \mathfrak{m}_{O_2}$, represents the set of models for which both observations hold, and all models contained within it cannot be discriminated from one another until further observational evidence is obtained.

In a similar fashion, we can illustrate the way in which a sentence of a formal language partitions the space of models. Therefore, suppose that some sentence γ is true of some models, false of others, and meaningless on still others; the space $\mathfrak{m}$ might possibly be partitioned as in Figure 2. Again, interpreting the areas as being proportional to the a priori probability, this particular sentence is most likely to be false, in the observer's view. This diagram shows the general case for a sentence γ, but any of these three areas could be null for some particular sentence. For instance, there might be

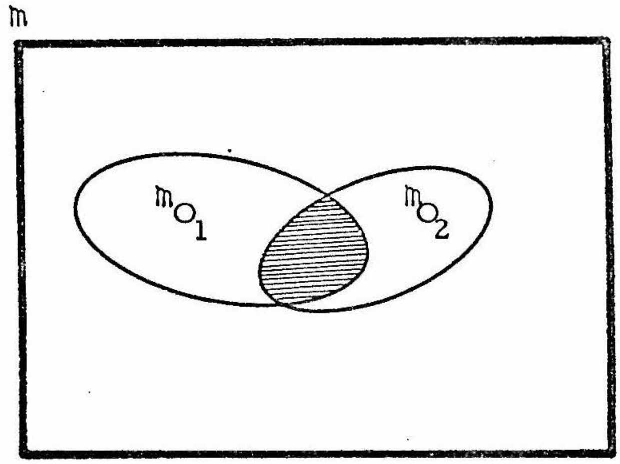


Figure 1. Two consistent observations.

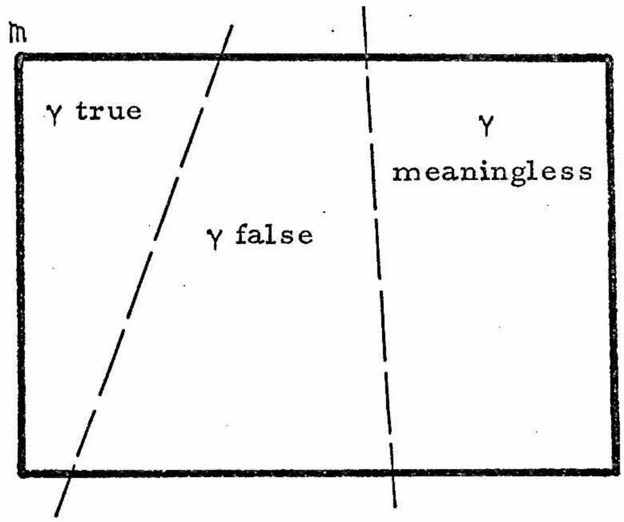


Figure 2. Partitions induced by a sentence γ.

no models where the sentence γ was false; such a sentence would be a tautology with respect to the language since whenever it is meaningful, it is also true. Notice that we have characterized an observation by a closed curve and a sentence by lines bisecting the space $\mathfrak{m}$. This is an artificial distinction for the sake of visual convenience only, since in either case the sets of models are defined by formulas of set theory.

Now consider the superposition of Figures 1 and 2 as shown in Figure 3. The observer has made the two observations O_1 and O_2, and we are interested in how the sentence γ relates to them. Suppose he makes observation O_1 first; the sentence γ is ambiguous for this observation since there are models contained in $\mathfrak{m}_{O_1}$ for which γ is true and models for which γ is false. Therefore, γ does not aid in the characterization of O_1. If he now makes observation O_2, then, as previously pointed out, the shaded area will represent the only models compatible with both observations. The diagram shows that γ is false of all of these models, and thus γ can be used to express what has been observed. Notice that γ is not ideal for this purpose since there are many models for which γ is false that are not contained in $\mathfrak{m}_{O_1} \cap \mathfrak{m}_{O_2}$. That is, the observer's observational experience is more refined or precise than the sentence γ can express.

Suppose, on the other hand, that O_2 was the first observation; the sentence γ is clearly not true of O_2, but neither is it false,

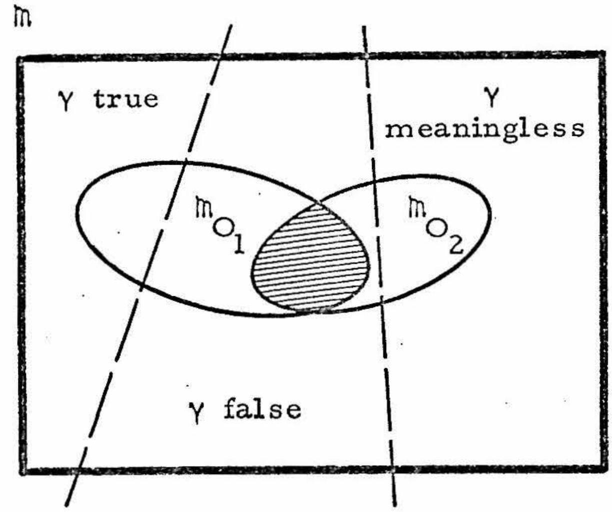


Figure 3. Relationship of the sentence γ to the observations O_1 and O_2. (Superposition of Figures 1 and 2)

since O_2 admits models for which γ is meaningless. The models contained within $\mathfrak{M}_{O_2}$ are equivalent to one another on the basis of the current observational evidence, and therefore γ is not useful in describing O_2 since the observer can only assert the truth or falsity of a sentence. Referring to our definitions of elements and states of a formal language, we conclude that the set of models for which γ is true is a state, as is the set of models for which γ is false, but the set of models where γ is meaningless is not, although it is an element.

In order to clarify the notion of a basis somewhat more, consider the following simple example; assume that we have a language whose only two sentences, γ_1 and γ_2, are:

γ_1 : "Spiders are responsive to light" ,

γ_2 : "Spiders eat ants" .

These sentences are supposed to have been generated by some syntax and their meanings to be expressed by an associated semantics. Figure 4 illustrates how these two sentences might partition the model space. To see how each of the sentences could be either true, false, or meaningless, independent of the other, consider the word "spiders". This word will be assigned some part of speech and thereby be associated with a semantic category, perhaps the class of animate objects. The formula defining this semantic category will express the generic structural properties which are required of animate objects. However, it may be that in some possible configurations of the universe, the

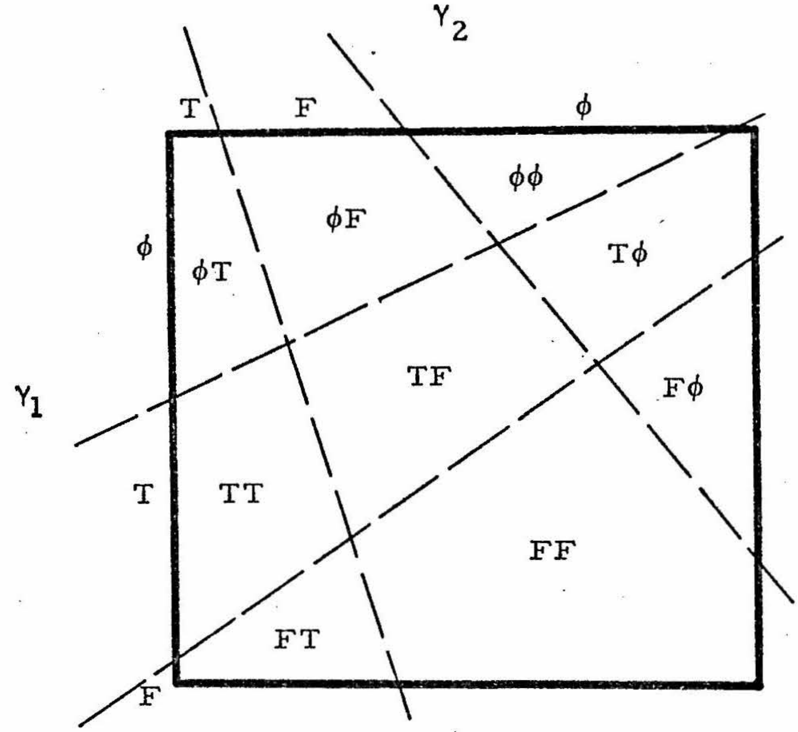


Figure 4. State diagram for a simple language having two sentences.

class of entities denoted by the word "spiders" fails to satisfy these requirements, i. e. the interpretation φ of "spiders" does not belong to the appropriate semantic category. For instance, in a given model, the class of "spiders" might include a spider which had just been stepped on and killed. Does this entity still satisfy the structural requirements of "spiderness"? It might, if for example these requirements involved only the molecular composition of the object in question, or it might not, in which case any sentences speaking about spiders would be meaningless for this model. An analogous situation holds for the other referent words in these two sentences, and thus, in general, the implicit structure required of the objects in the domain of the language may be violated in some plausible configurations of the universe. If this is not the case, i. e. if the sentences are meaningful, there are certainly valid interpretations of the words in both sentences which render them either true or false. As we have previously mentioned, the deep structural aspects of a language, which are in part embodied by the semantic categories, are essential to the constructive generation and analysis of strings of words into understandable sentences. On the other hand, the very imposition of this structure dictates that these sentences are not universally applicable to every conceivable reality.

Since the two sentences of our example language are independent in the manner discussed above, Figure 4 shows a non-empty element corresponding to each of the nine possible truth assignments, and the symbols indicate the particular truth assignments defining the elements. For instance, the element Tϕ is the set of models where γ_1 is true and γ_2 is meaningless. The states of this language are defined by the following sentence/truth value combinations: T-, F-, -T, -F, TT, TF, FT, FF, where a "-" indicates that the corresponding sentence is not used to define the state involved. In addition to the states defined by these sets of sentences, $\mathfrak{m}$ and the empty set of models are also states. The minimal states are those defined by TT, TF, FT, and FF, and any one of these is distinguishable from each of the others. As we have shown, in this case the basis is unique and comprises the entire collection of minimal states.

As a final example, we illustrate a situation where the basis is not unique. Figure 5 shows the state diagram for a hypothetical language having three sentences γ_1, γ_2, and γ_3. Each element of the language is characterized by an assignment of truth values to γ_1, γ_2, and γ_3; these values are specified by the sequences of three symbols shown on the diagram, which will be considered to name the corresponding elements. There are seven minimal states which are defined by the following sets of truth values:

TTT, TFT, FTT, FFT, FFF, -TF, T-F ,

where "-" indicates that the correponding sentence is unnecessary to define that state. Notice that the two minimal states TϕF and ϕTF are indistinguishable, i.e. there is no sentence which is true of one and false of the other. As a consequence of this, there are two distinct bases for this language, namely,

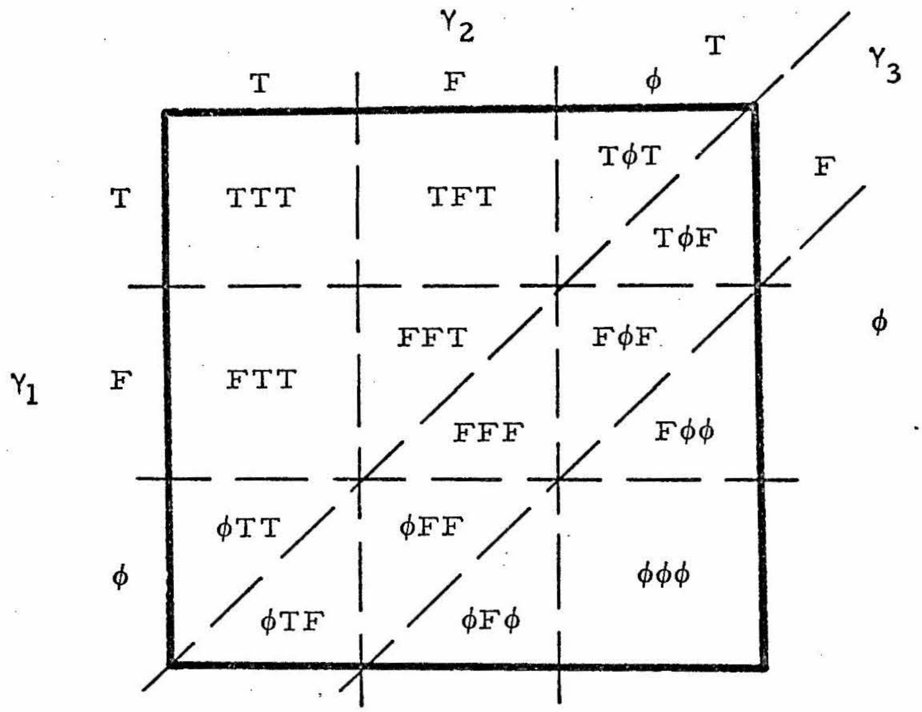


Figure 5. State diagram for a simple language having a non-unique basis.

$$\mathfrak{B}_1 = \{TTT,\ TFT,\ FTT,\ FFT,\ FFF,\ \emptyset TF\}$$

and

$$\mathfrak{B}_2 = \{TTT,\ TFT,\ FTT,\ FFT,\ FFF,\ T\emptyset F\}.$$

Also significant is the fact that the logic of this simple language precludes the existence of certain states. For instance, the state defined by TFF is empty, indicating that this is an inconsistent assignment of truth values relative to the semantics of the language. If the meanings of these three sentences were completely independent of one another, then there would be $3^3 = 27$ non-empty elements, nine minimal states, and the basis would be unique, However, we emphasize again that an essential characteristic of nearly all formal languages is that they embody some non-trivial logic, and the sentences of these languages are consequently not independent.

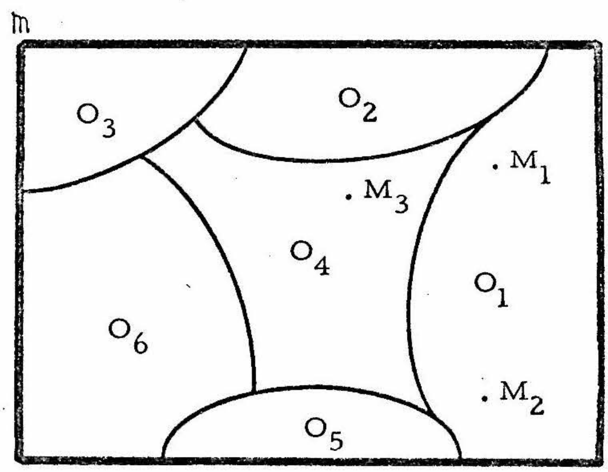
V. EXPERIMENTAL OBSERVATIONS AND SCIENTIFIC THEORY

In this chapter, we shall utilize our formal definitions of probability, information, and language to interpret some of the problems confronting a scientific researcher. Therefore, suppose that we are considering a scientist who is about to perform some experiment. Let us imagine that this man is an experimental biologist who is attempting to study the processing of visual information in certain classes of insects. Specifically, we will assume that he is interested primarily in examining the behavior of certain types of neurons and their functional relationship to external visual stimuli. In order to accomplish this, of course, he must have available a certain amount of experimental equipment, and he chooses some appropriate portion of this equipment to aid him in any particular experiment. This apparatus constitutes the experimental environment and is instrumental in determining exactly what quantities will be measurable in the experiment. Thus, for example, he may be planning to insert tiny microelectrodes into one of the insect's optic lobes and then to record the electrical signals from nearby nerve cells. It could be that he is then interested in examining the detailed electrical waveform of a spike discharge from these cells, but we shall suppose that he is most concerned with the temporal behavior of the neurons in encoding the visual signals. Thus he will not find the analog wave shapes as relevant as the times of discharge of spikes from a cell, and consequently he views the neuron's output as a train of essentially zero width pulses and concentrates his attention on the intervals between these pulses. Of course, in order to determine exactly when a pulse occurs, i. e. when a cell discharges, he must adopt some criterion, such as the time when the voltage measured at the electrode exceeds some threshold. He might then record the neural signals, after they have been amplified and filtered suitably, on a device like a strip chart recorder, but since the firing times of the cells are the fundamental items of interest he can simply digitize these times and collect these data using a small computer system. We will suppose that he takes this latter course, thereby enabling him to collect a large amount of data and save it for subsequent analysis. Of course, he could also perform some analysis in "real time" as the experiment proceeds but this is not germain to our example.

Now, with regard to the actual experiment he is going to perform, he will perhaps insert two of these microelectrodes into certain general areas of interest within the insect's visual nervous system. He then plans to display to the insect, which is fixed in the center of a large globe, some pattern consisting of alternating dark and light stripes. This pattern will be turned on and off, and he will thus be able to examine both the transient and steady-state responses of the neurons near his probes and also the interrelationships among the two neurons. The computer system will record the firing times of the cells together with timing signals pertinent to the stimulus. Each phase of the experiment will be repeated a number of times to ensure a statistically sufficient sample. Subsequent to this he will analyze the data using various computer algorithms which he has previously found useful. As might be surmised, this experimental description is not hypothetical; for a more detailed discussion see McCann and Dill [24].

The purpose of the preceding discussion has been to establish a specific framework within which we may discuss various instances of more general phenomena. One of the questions of interest here concerns the possible outcomes of such an experiment. As we have indicated, the space of possible outcomes has been at least partially determined by the selection of the apparatus and the design of the overall experimental environment.

In terms of the diagrams introduced previously, an experiment may be characterized by a disjoint covering of the space $\mathfrak{M}$, where the members of the cover are sets of models associated with observable experimental outcomes. That is, the cover comprises a set of observations having the property that some one of them holds for any model $M \in \mathfrak{M}$.



The diagram illustrates the case of an experiment that has six possible observable outcomes. That is to say, the models contained in a given O_i are all those possible configurations of the world in which this experiment would be observed to have the same outcome. Suppose for a moment that M_1 were the actual configuration of the world, the "true" model. Then one could conclude as a result of this experiment that the "true" state was in O_1, and that we were not, for example, in the configuration M_3. However, it would remain an open question whether we were at M_1 or M_2.

If virtually any conceivable experiment could be performed, then every such disjoint cover would be representative of some possible experiment. What we are saying is that the experimental apparatus, once chosen, determines some sub-collection of the collection of all possible disjoint covers. Among the covers that are excluded will be those which have as members some observation involving a quantity which the equipment is incapable of measuring. Those which remain are indicative of the experimenter's freedom of choice with regard to the experimental parameters under his control. In our example, if the amplitude of the spikes gradually decreased with time, then because the data recorded make no direct allowance for this, it can not be an observable outcome of an experiment using this equipment. Consequently, the raw data collected already represent an abstraction from what is potentially available. In addition, it could be that other unmeasured quantities, such as the ambient air temperature, have some non-trivial effect, but again, correlations of this type are not possible observational outcomes. The selection of the stimulus itself clearly also limits what can be directly observed. Thus, we certainly can not observe what the response of the neurons in question is to the appearance of another insect, say, by using a pattern of stripes. Notice that we are speaking now about what is observable in this experiment, that is, what can be perceived. This is to be interpreted exactly in the sense that we have formally defined an observation. Consequently, the possible outcomes are independent of the observer's language and relate only to the configurations of the universe which might conceivably obtain. As yet we have said nothing about the conclusions he might draw from some given outcome, since this will be determined by his language.

The main point of the preceding is that the observable results of a scientific experiment are dictated in large measure by the specific details of the experimental set-up. Various outcomes are thus excluded a priori, and others become not only possible but highly probable. In many ways this point is obvious and we are certainly not saying that there is any practical method for greatly expanding the space of possible outcomes. For example, if a microelectrode could be placed adjacent to every neuron in the insect's visual system, the set of potential observables would be immense. But since there may be a million such nerve cells, this is clearly a technical impossibility, without even considering whether the data collected could be sensibly analyzed. Also, it would be naive to assume that the outcomes which are possible for a given experiment are arbitrary, since they are largely the result of the scientist's evaluation of what are meaningful or relevant quantities to measure. This evaluation then results in a determination of the actual experimental conditions, within practical limitations. However, these choices, once made, do strictly limit the scope or range of alternative outcomes.

Now, returning to our example, we can visualize that such an experiment has a very large number of possible results. That is, the set of all observations which the experimenter could conceivably make as a result of performing this type of experiment, even though restricted by the experimental conditions, still encompasses a virtually unlimited number of actual possibilities. These include many uninteresting cases where, for example, the equipment breaks down or the insect dies before the data can be gathered. Furthermore, the total collection is not a mutually exclusive set of observations. Thus, one possibility would be that the cells respond to the stimulus, whereas two other possibilities would be that they respond to the stimulus by increasing or decreasing their average rate of spike discharge. These are clearly not independent events; on the other hand the experimenter would presumably not observe the more general of these if his apparatus permitted the observation of one of the latter two. Therefore, we can assume that there is some maximal set of observations which are all mutually exclusive, where the elements of this set are determined by the quantities he is measuring, the precision of the experimental apparatus, and the means available for examining the results. Each possibility in this set thus represents a distinct experimental outcome, and the set itself defines a disjoint cover of the model space, as previously mentioned.

If we consider any such specific instance of a scientific experiment, then it seems intuitively clear that there are only a finite number of observable possible outcomes. That is, we have no reason to believe that the cover associated with the experiment has more than a finite number of cover sets. Furthermore, since there will always be an infinite number of relationships which are not specified by the experiment, in general every cover set will contain an infinite number of models among which the experiment does not distinguish.

Recalling our definition of observation, the set of observations which constitute a cover is inconsistent, and in addition, the union over the set of possible configurations of the universe of discourse associated with these observations includes all admissible configurations or models. Thus, it is certain that one and only one of these possibilities will be realized by performing the experiment. We emphasize again that the outcomes of the experiment correspond to observations and are not directly related to the observer's theories as expressed by his language. Of course, the various outcomes of the experiment are not equally likely, the probability of any one being given by the observer's a priori probability P which reflects his metaphysical assumptions about the domain he is studying.

We now turn to the question of how the experimenter discriminates these experimental outcomes. We have already said that certain outcomes are somehow more relevant than others. How is this matter of the relevance or importance of the result of the experiment to be explicated? Our first inclination is to say that this depends upon his current theory about the domain he is studying. That is, he would certainly like the actual outcome to be consistent with what he now believes. Furthermore, he hopes that the result will do more than simply confirm what he already knows. Ideally the new observation should further reduce the set of models satisfying his total observational experience and thereby contribute to refining his theory. In order to say what we mean by a theory in this sense, we utilize the notion of states of a formal language and the sets of sentences that define them. Prior to performing the experiment we have outlined, our experimenter has some body of data previously gathered about the domain, in this case data relevant to the insect visual system. These data comprise some set of observations which we presume to be consistent, and associated with these observations is some set of models satisfying all of them. Now, consider the smallest state of his formal language containing this set of models; the sentences defining this state are all either true or false of every observation he has previously made. These sentences act therefore as axioms for his theory. They might include such statements as, "The insects studied have some neurons which increase their rate of spike discharge for approximately two seconds after a 20^{o} spot of light is turned on in their field of view", or "The insects studied do not respond in any way to polarized light". In general, these axioms are subject to change since they are based only on what he has observed, and it is possible that some subsequent observation may contradict them.

Figure 6 shows a partial state diagram for the experimenter's formal language. The set of models denoted by Ω represents all those configurations of the universe which are consistent with his previous observational experience. The state h is assumed to be the smallest state containing Ω, and thus the sentences of the language that define h are axioms of the experimenter's current theory. Notice that the axioms do not precisely delimit Ω; there are models $M \in h$ which are not members of Ω. This simply means that the formal language he employs is not capable of expressing every detail of his previous experience and provides only some convenient approximation to it.

This should not, however, be viewed as necessarily negative. The lack of refinement results from ignoring aspects specific to the time of day, individual subject used, etc., which are of no relevance to the researcher. These he appropriately ignores if they indeed are irrelevant, a chance he must always in some way take.

In addition to the axioms, which are either wholly true or false of his observations, there will be other statements which represent speculations or hypotheses. Very often these will be generalizations of his previous experience: for instance, "all insects have

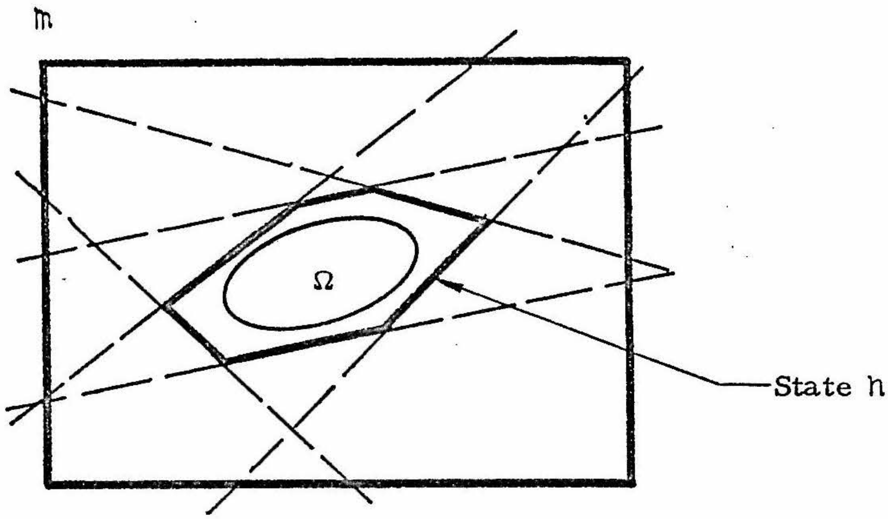


Figure 6. Partial state diagram showing experimental axioms and their relationship to prior observations.

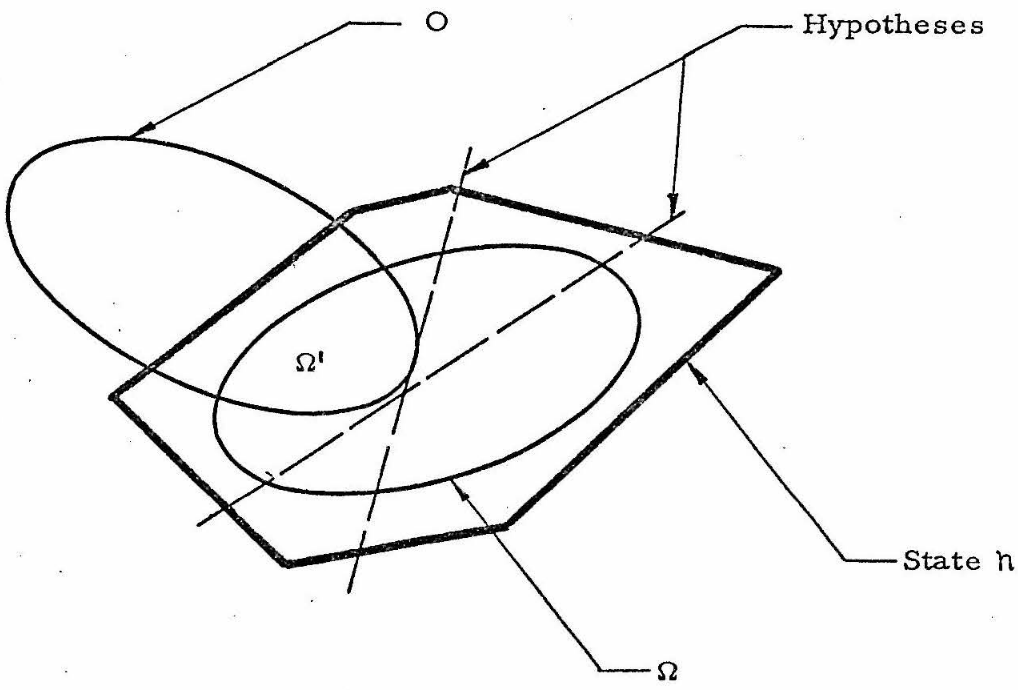


Figure 7. Expanded view of state h (from Figure 6) showing experimental hypotheses and a new observation O.

neurons which increase their rate of spike discharge in proportion to the ambient light level". Since he has clearly not tested this for every insect, this statement will be neither true nor false of all his observations. There may also be other hypotheses which are completely untested, in the sense that no previous observational evidence is related to them. These he presumably intends to either confirm or deny by subsequent experiments. The salient point here is that the conditional or hypothetical statements in his formal language act to partition the state defined by the axioms into a number of smaller states contained in it. Each of these states represents a refinement of his theory for which insufficient evidence currently exists. Moreover, the statements characterizing these states will most likely be experimentally verifiable, although to do so may involve procedures which are very complex or not technically feasible with his present equipment. In any case, these hypotheses act as a guide to further experimentation, since to perform an experiment whose possible outcomes can not be discriminated by some statement of his language would not be meaningful. Therefore, the experimental outcomes which are relevant to him are mirrored directly in the statements of his language.

An illustration of the effect of hypotheses on the partitioning of the model space is shown in Figure 7. The state h and the set of models Ω are the same as in Figure 6; the two dashed lines correspond to sentences of the formal language whose truth or falsity is not determined by the observations previously made. Now suppose that the experimenter can perform some experiment which has the observation O as a possible outcome. If he actually does this and observes O as the result of the experiment, then Ω' will characterize the set of models consistent with his total experience. Also, the two hypotheses will have been resolved, since they are now either confirmed or denied on the basis of the new evidence O . Consequently, observing O is informative relative to the experimenter's theory as expressed within his formal language.

Notice that if he has no hypotheses to be tested, that is, if every statement of his language is either completely true or false of all his observations, then he will do no further experiments. His theory is complete because he can visualize no sensible means of refining it. Needless to say, very few modern scientific theories have achieved such a status.

Now, with regard to the selection of a specific experiment to be performed, several remarks can be made in light of the preceding discussion. The choice of experiment will be based upon how informing the experimenter anticipates the outcome will be. Since he knows, in principle, exactly what the potential outcomes are, and since he also has various hypotheses which he wishes to confirm, the combination of these will dictate that certain experiments are more useful to him than others. This brings us back to our definition in the preceding chapter of the expected informational gain for an experiment. If there are several experiments which may be performed with equal ease, then the proper choice from the experimenter's standpoint is the one which has the highest expected informational payoff. In a manner analogous to information theory, this experiment is ideally one where the *-probabilities of the various outcomes are all equal, and since this probability is a function of the observer's language, the hypotheses he wishes to investigate are implicitly taken into account. Consequently, the informational gain referred to here is again in the subjective sense of informativeness and the experimenter therefore attempts to choose an experiment whose possible outcomes coincide well with the alternatives opened by certain of his unconfirmed hypotheses.

To illustrate this point, consider the following example which is represented graphically in Figure 8. Suppose that an experimenter has some body of experimental data; Ω is the set of models characterized by the corresponding observations. Further, assume for the sake of simplicity that the sentences of his formal language all result in a horizontal cut across the space of models. That is, using the experimental apparatus he has available together with all of the theory he feels is in any way applicable to the domain he is considering, the expressible concepts within his formal language divide up the possible worlds into collections of models like the rectangle delimited by α and β. Thus, in the diagram, α and β correspond to previously verified statements, and γ represents an untested hypothesis relative to Ω. Now imagine that he could perform one of two possible experiments, say, $\mathcal{E}_1$ or $\mathcal{E}_2$. $\mathcal{E}_1$ and $\mathcal{E}_2$ each have two outcomes which are delimited by the solid lines E_1 and E_2, respectively.

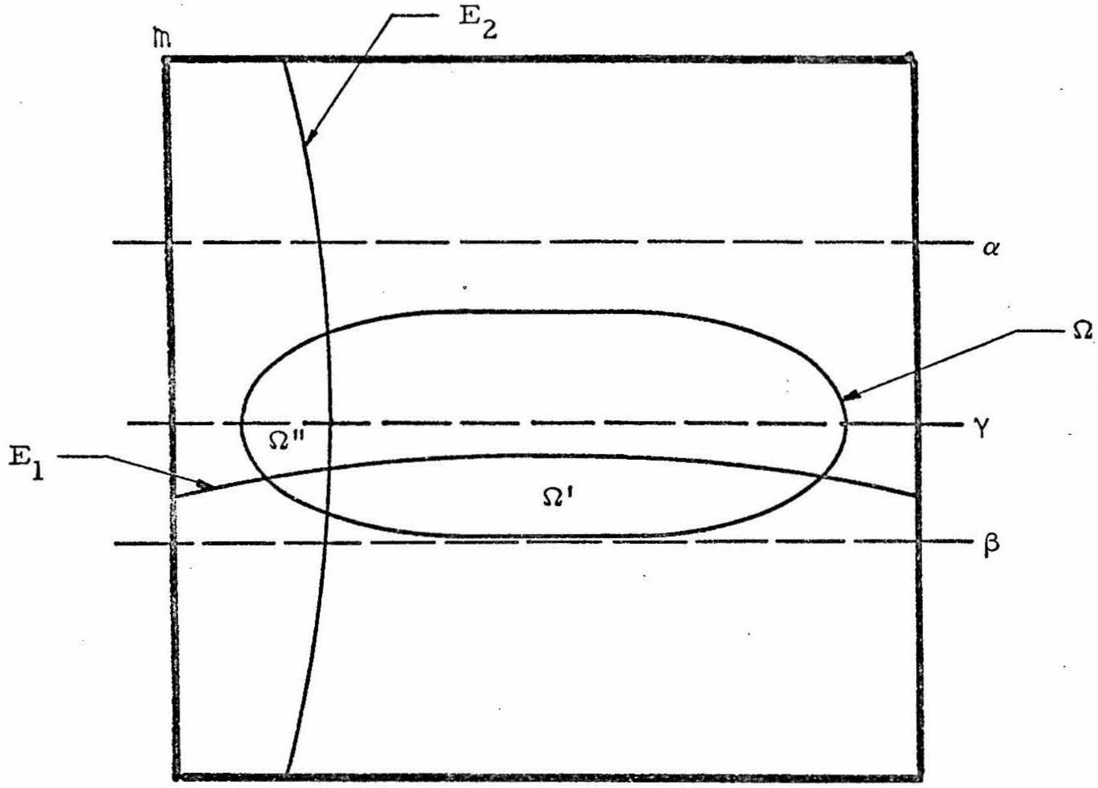


Figure 8. Hypothetical case showing three sentence partitions, α, β, and γ, belonging to some formal language, and two experimental partitions, E_1 and E_2. Ω is the set of models representing prior observational experience.

Which of these two experiments will he actually choose to perform? If he performs experiment $\mathcal{E}_1$ and thus determines whether the models below the line E_1 or the models above the line E_1 are consistent with what he observes, then he may be able to confirm or deny the hypothesis γ, depending on which of the results he obtains. On the other hand, if he performs experiment $\mathcal{E}_2$, neither of its outcomes will confirm or deny γ or any other hypothesis of his language since they would also partition the model space horizontally.

Assume, as we have previously, that the observer's a priori probability is proportional to area on the diagram. Now, if experiment $\mathcal{E}_1$ turns out successfully, the statements γ and β will then characterize, as well as possible, the resulting set of models Ω', but no matter what the outcome of $\mathcal{E}_2$ is, the best state description of the resulting set of models will still be given by the statements α and β. Consequently, the observer will gain no information by performing $\mathcal{E}_2$ but can reasonably expect to be better informed by performing $\mathcal{E}_1$. Notice, however, that if his language could express some statement which, for instance, coincided with the boundary E_2, then if $\mathcal{E}_2$ was completely successful, he could characterize the resulting set of models Ω'' by using α, β, and the statement coinciding with E_2. This state is clearly much smaller than the state defined by γ and β which is the most he can achieve by performing $\mathcal{E}_1$. Consequently, if his language embodied the concepts necessary to appropriately express the outcomes of $\mathcal{E}_2$, this experiment would be seen to have a greater potential informational payoff than $\mathcal{E}_1$. But his inability to grasp these concepts prevents him from conceiving of $\mathcal{E}_2$ as a reasonable experiment, and his choice is necessarily $\mathcal{E}_1$. This explicates in a precise way the channeling effect of observations and theory on further experimentation, and thus on the development of new theory. These notions are related to those of "normal science" by Kuhn [22] and to the Whorfian hypothesis [44, p. 213].

The counter-argument—namely that the scientist is not confined by his language/theory—depends on either the existence of a universal language or, what is tantamount to the same thing from the point of view of this paper, his ability to move cognitively with no conceptual boundaries. The fact that a universal language can not exist will be discussed subsequently. The dynamics of language change are beyond the scope of this thesis. However, we can not help but feel there are inertias involved which essentially substantiate the above analysis.

The relationship between the language structures and the possible experimental outcomes has another dimension which stems from rather different considerations. We have previously mentioned that the available experimental apparatus determines certain bounds on the set of experimental results, for example, by limiting the degree of precision of certain measurements. In other instances, the equipment may actually preclude observations of some physical quantities. Thus in our example experiment, the microelectrodes used may only be suitable for measuring potential changes external to a neuron but can not be employed to make similar measurements within such cells.

Therefore, practical limitations on the amount of available equipment and matters of technical feasibility may impose rather arbitrary restrictions on the potential experimental outcomes. Another way of looking at this is to say that the apparatus acts to determine the experimenter's "contact points" with the reality of his experiment. If we visualize the system he is studying as a network of "black boxes" connected in some fashion, then these "contact points" are identified with the inputs and outputs to the "black boxes", i. e. the set of measurable quantities.

Now, recognizing that the experimental apparatus plays an important role in determining the character of the observable experimental outcomes, we ask how this affects the observer's language. Since the information gained by the experimenter when he performs some experiment is directly related to his ability to distinguish the outcomes of that experiment in his language, it follows that the hypotheses embedded within the language are indirectly linked to the experimental apparatus. Thus, the experimental environment influences the observable outcomes and they in turn impose conditions upon the language, if it is to be an informative language for that experimental domain. This relationship is particularly important because the adaptability of the experimental environment is typically not great. Modern scientific practice indicates that the trend is to more and more sophisticated and complex experimental equipment and in many cases, once this equipment is adequately developed, it is used for long periods of time. The result of this is that the language of the experimenters also tends to stabilize for equally long periods, and consequently the language acquires a somewhat unnatural rigidity or resistance to change. Thus, although the formal language of an experimentalist may not be inherently fixed in nature, it is apt to become so by virtue of its necessary relationship to empirically observable phenomena.

The preceding sections of this chapter have dealt with the general relationships among formal languages, observations, and scientific experiments. We shall now consider, in greater detail than we have previously, the formulation and interpretation of scientific theories. Following this, we shall return to the specific examples presented in the introduction and investigate them in the light of these discussions. In order to get at the notion of a scientific theory, let us ask what we can say about the meaning of a sentence of a formal language. We know that to each such sentence there corresponds some formula of the language of set theory, having certain free variables. Precisely what does such a formula express? We have said that it characterizes a relationship among the objects which are taken as the values of its free variables.

Now imagine some specific sentence of a formal language and a particular model or configuration of the universe. Note that one can not determine whether the sentence is true of that configuration without knowing the interpretation of the referent words. Thus the meaning of the sentence involves the denotation of the referent words in addition to their structural properties, as specified by the semantic categories, and the interrelationships required to exist among them, as specified by the structure of the sentence itself.

At this point in our discussion we can not say what the referent words denote, only that once this is known we can describe the relationships existing among those things. Can this be strengthened? Is it possible that we can say within a formal language all that we mean or, on the other hand, is it necessary that there be some implicit or tacit knowledge?

What do we mean by tacit knowledge? We assume that a person speaking a given language understands the syntax of that language and the structural aspects of its semantics. Suppose, for example, that a person who speaks English hears the sentence "Bob is in Pettalle". One expects him to know that this is indeed a sentence and that it expresses the fact that someone is located in a certain place; we would not expect him to respond "I am not sure that what you said was a sentence, or the nature of the relationship it presumably expressed". He could sensibly respond "Who is Bob and where is Pettalle ?" The meanings of Bob and Pettalle, unless otherwise defined, must be tacitly known. Scientific theories that require no tacit knowledge, i.e. those in which all concepts are completely structural, are precisely the theories of pure mathematics. An empirical theory, however, is a theory where such tacit knowledge is necessary. This is an obvious statement that simply calls attention to the fact that any empirical theory has certain basic terms which can not be further reduced by definition. Observe that there is no need to debate whether tacit knowledge is required only in connection with the meanings of words. It can easily be seen that by adding new words to a language and giving them tacit meanings, one can reduce all tacit knowledge to implicit understanding of the meanings of individual words.

As the next step, we inquire into the legitimate meaning of the notion of scientific theory. Such a theory clearly concerns some set of sentences of a formal language. These sentences, say $\mathcal{E}$, are the axioms and hypotheses of the theory, together with their logical consequences. Furthermore, a scientific theory certainly purports to say what configurations of the universe actually obtain, i.e. something about the world around us. In view of the above discussion, however, we can not conclude that the meaning of the theory is the set of all possible models for which the sentences of $\mathcal{E}$ are true, simply because the truth or falsity of a sentence of a formal language depends also on the interpretations of the referent words which for an empirical theory must be essentially tacit. Thus, as a first hypothesis, a scientific theory is a set of sentences and an interpretation of the referent words of those sentences; and the meaning of the theory is the set of all models for which those sentences, under the given interpretation, are all true.

Of course it is essential to associate the notion of a scientific theory with observations, which we now do. Whereas a scientific theory depends upon an implicit understanding of the meanings of its words, that tacit knowledge should be tied to empirical observations. Suppose someone asks "What do you mean by α?" where α is a basic, tacitly understood word of our theory. We can not tell him. But we certainly should be able to show him, perhaps by pointing or suggesting to him that he perform certain experiments or examine certain objects. Thus the tacit meanings should not be mysterious but should arise directly from observational experience. This is indeed the function of student laboratories in a scientific education. One can translate this, through the definitions and theorems given previously, into the requirements that the set of models prescribing the meaning of a scientific theory should be definable in terms of observations in our sense. How are we to do this? Consider two possibilities:

(1) The set of models that constitute the meaning of the scientific theory must be associated with an observation, or a finite number of observations.

(2) Given a set of models $\mathfrak{h}$ which is the meaning of a scientific theory, then for any model $M \notin \mathfrak{h}$ there is some observation which holds for M but for no model in $\mathfrak{h}$. Loosely speaking, if the theory does not hold for some possible world, one can ascertain this fact by an appropriately designed experiment. For a similar view, see Popper [27, sec. 6].

Condition (1) is clearly too strong; a useful scientific theory obviously says more than that some finite set of objects are in a certain fixed relationship. When we say the condition is too strong, we do not mean to exclude scientific theories that satisfy it, but we note that the assumption of the condition for all theories seems to imply acceptance of a totally finite universe, finite in every way, including a finite number of discrete time quanta. Consider the simplistic theory consisting of the single sentence "Flies respond to visual stimuli". It seems unwarranted to assume there will be only a finite number of instants and/or stimuli where this could be either true or false. The second condition, (2), is tantamount to the assumption that the meaning of a scientific theory is given by the intersection of the sets of models associated with an infinite number of observations. To see this, let $\mathfrak{n}$ be the set of models which constitute the meaning of the theory. For any model $M \notin \mathfrak{n}$, let O_M be an observation which is valid for every $M' \in \mathfrak{n}$, but not for M. This corresponds to the negation of the observation stipulated in (2). $\mathfrak{n}$ is obviously the intersection of all such O_M's. Thus, this weaker condition seems more plausible, and we tentatively accept (2) as the relationship between a scientific theory and observations.

The following definition specifies the conditions which must hold if a language is to be able to precisely delimit some observation:

Definition: An observation O is describable in a formal language $\mathfrak{L}$ if $\mathfrak{m}_O = \mathfrak{n}$, where $\mathfrak{n}$ is a state of the language. If $\mathcal{E}_{\mathfrak{n}}$ is a set of sentences of $\mathfrak{L}$ defining $\mathfrak{n}$, then $\mathcal{E}_{\mathfrak{n}}$ is said to describe O.

Every possible observation O is describable in some formal language, and this can be accomplished by a single sentence. The argument to this end is trivial:

Lemma: For any observation O, there is a formal language $\mathcal{L}$ in which O is describable by a single sentence.

Proof: Let $O = < F(x_1, \cdots, x_n), a_1, \cdots, a_n >$ where $a_i \in S$. Then let $\mathcal{L}$ be a formal language with the sentence $\gamma = \gamma_1\gamma_2 \cdots \gamma_n$ such that $\varphi(\gamma_i) = a_i$. The lexicon L contains the rules $\beta_i \rightarrow \gamma_i$ for $i = 1, \cdots, n$. $C_{\beta_i} = S$. The grammar G contains the rule $R: \alpha \rightarrow \beta_1\beta_2 \cdots \beta_n$ and the associated semantic transformation is

$$\tau_R : F(x_1, x_2, \cdots, x_n) \equiv y = 1 .$$

Therefore, the formula of set theory associated with the sentence γ is

$$F^{\gamma}(\bar{x}, y) \equiv \left(F(x_1, \cdots, x_n) \equiv y = 1\right) .$$

Hence, γ is true of some model $M \in \mathfrak{M}$ if and only if $M \in \mathfrak{M}_O$. Therefore γ describes O.

Now, a finite set of sentences together with an interpretation of their referent words always characterizes a set of models which could be considered as the meaning of some scientific theory. In fact, the following lemma shows that these sentences describe some single observation.

Lemma: Any finite set of sentences $\mathcal{E}$ of a formal language $\mathcal{L}$, together with some truth assignment to the sentences, describes some observation O.

Proof: suppose $\mathcal{E}$ consists of the sentences $\gamma_1, \gamma_2, \cdots, \gamma_n$. Let $y_i = 1$ if γ_i is assigned the value true, $y_i = 0$ if γ_i is to be false, and let $F^{\gamma_i}(\overline{x}_i, y_i)$ be the formula of set theory corresponding to γ_i. Then let $F(x_1, \cdots, x_m) = F^{\gamma_1}(\overline{x}_1, y_1) \wedge F^{\gamma_2}(\overline{x}_2, y_2) \wedge \cdots \wedge F^{\gamma_n}(\overline{x}_n, y_n)$, where $x_1, \cdots, x_m$ are the accumulated free variables that stand for $\varphi(\delta_1), \cdots, \varphi(\delta_m)$. $\delta_1, \cdots, \delta_m$ are all the distinct referent words appearing in the γ's, and y_i takes the value 1 or 0 as specified by the truth assignment. Then $O = < F(x_1, \cdots, x_m), \varphi(\delta_1), \varphi(\delta), \cdots, \varphi(\delta_m) >$ is an observation described by $\mathcal{E}$ under the given truth assignment.

Note that this lemma does not imply that the models characterized by $\mathcal{E}$ represent the actual outcome of some experiment, only that the observation O is possible in principle. As we have previously discussed, there is one model M_0 which is the true model—the configuration which actually holds. To verify a scientific theory is to make the observations defining the set of models associated with it and to ascertain that this set of models contains M_0. Obviously verification in this manner is limited by experimental apparatus and technological possibilities. Thus, for example, given that we tacitly know the meanings of "life" and "Mars", it is a scientific theory that there is life on Mars. This particular theory is as yet unverifiable.

We have shown that a finite set of sentences describes some observation; what about an infinite set of sentences? The following theorem demonstrates that if the infinite set of sentences is recursive, then there is some formal language where a finite set of sentences will characterize exactly the same configurations of the universe, and therefore the infinite set describes some observation. By a recursive set of sentences, we mean that for some Gödel numbering of the referent words of a language, then the resulting Gödel numbers for the sentences of the language form a recursive set of integers.

Theorem: Let $\mathcal{L}$ be a formal language. If $\mathcal{E}$ is a recursive set of sentences of $\mathcal{L}$, then there is some formal language $\mathcal{L}'$ such that, for a finite set of sentences $\mathcal{E}'$ belonging to $\mathcal{L}'$, the set of models defined by $\mathcal{E}$ relative to $\mathcal{L}$ coincides with the set of models defined by $\mathcal{E}'$ relative to $\mathcal{L}'$. Hence $\mathcal{E}$ and $\mathcal{E}'$ both define some observation O.

Proof: $\mathcal{L}$ is a formal language. Both the syntax and the semantics of $\mathcal{L}$ are formalized within set theory. Thus it is easy to see that a meta-language for $\mathcal{L}$ could also be formalized within set theory, including a definition of the notion "true in language $\mathcal{L}$". Call this formal language $\mathcal{L}'$. (Of course, in general, we can not define "true in language $\mathcal{L}'$", within $\mathcal{L}'$ itself [39, chap. 8].)

Since the set of sentences $\mathcal{E}$ is recursive, we can give a recursive definition of $\mathcal{E}$ within $\mathcal{L}'$ in a finite way. Then one can express within $\mathcal{L}'$, using this definition, that the sentences of $\mathcal{E}$ are true, and to do this requires only a finite number of sentences of the meta-language $\mathcal{L}'$. As previously shown, this finite set of statements describes some possible observation O , and thus $\mathcal{E}$ also describes O , relative to $\mathcal{L}$. To carry out this argument in all its detail would require a great deal of space and time.

This theorem and the preceding lemma raise the possibility that our definition (1) of the meaning of a scientific theory may be adequate. Do we indeed know of any set of sentences which do not describe some observation? The answer is definitely yes; the Gödel incompleteness theorem tells us that the integers (in this case a theory in which no tacit knowledge is necessary since it is purely mathematical) can not be described in any formal language by a finite or recursive set of sentences. More precisely, given an object b , the set of models where b is the integers can not be characterized by a recursive set of sentences. Thus, the meaning of the theory of the integers does not correspond to any finite set of observations. This extremely strong form of the Gödel result is discussed in Kleene [20, sec. 60] .

Although for most practical scientific theories, it is not possible to definitely establish that definition (1) in fact does not hold, it is our feeling that this is often the case. For example, consider a theory which purports to be about living organisms; it is highly unlikely that a definition of the notion of living in terms of a finite number of observationally specifiable objects can establish precisely what we mean by this concept. Concepts like "living", "cell", "metabolic process", etc., are learned not through language but become tacitly understood through years of experience and long hours of directed laboratory familiarization. One everyday concept which we all take for granted is that of volume. It has been shown that it can not be prescribed in a finite number of observations. Indeed, a much stronger result holds. However, this matter will be taken up from a somewhat different point of view.

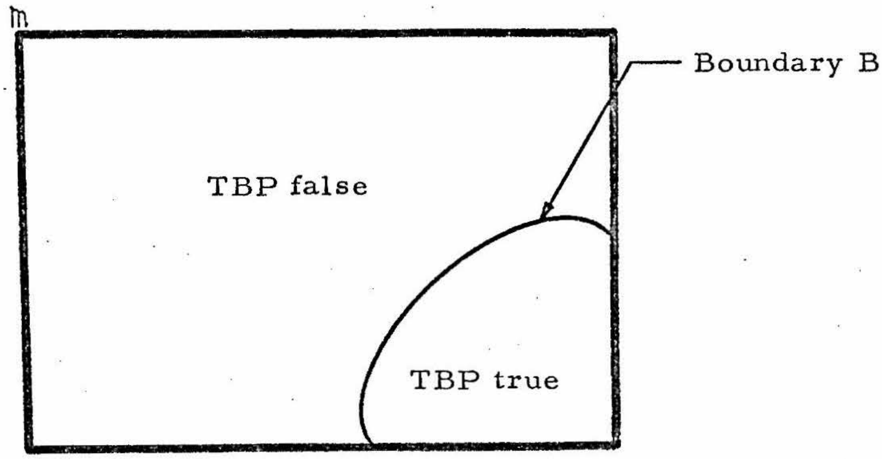
Let us turn to the related problem of reductionism. In chapter I, we identified two forms of this problem, the first of which will be discussed here. The second aspect of the reductionist philosophy, namely the reduction of one theory to another more basic one, will be treated subsequently. The immediate question is whether the truth or falsity of a theory can always be reduced to the outcome of some experiment. More precisely, is it legitimate to restrict the notion "valid theory" to those theories for which the possibility of verification or denial on the basis of experimental evidence exists?

Consider the following statement:

(TBP): "There is a way to cut a sphere into a finite number of pieces which can then be moved rigidly and fitted together without deformation to form two spheres each of exactly the same volume and size as the original."

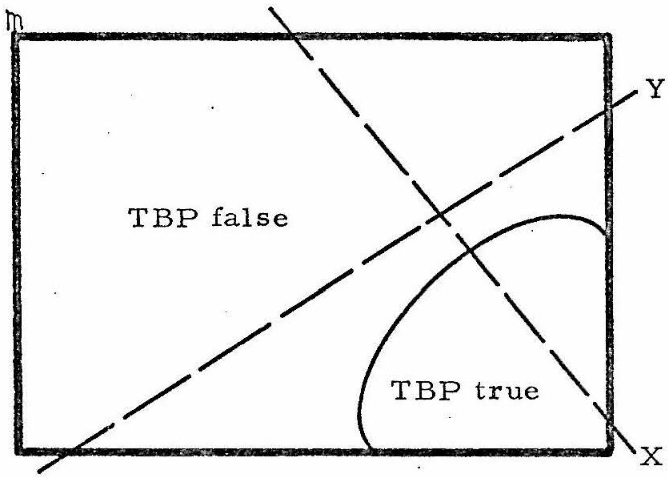
Suppose we have a formal language, for instance a segment of the language of mathematical physics, within which this statement can be given precise expression. That is, the language speaks about three-dimensional Euclidian space and the notion of volume as a measure invariant under rigid motions in that space. We then make the following three statements:

(1) The physicist believes that the statement TBP is false because he assigns a zero probability to the set of models for which it is true. This probability corresponds to the a priori probability P and indicates the physicist's certainty that he will never make an observation which implies the truth of the meta-statement TBP.



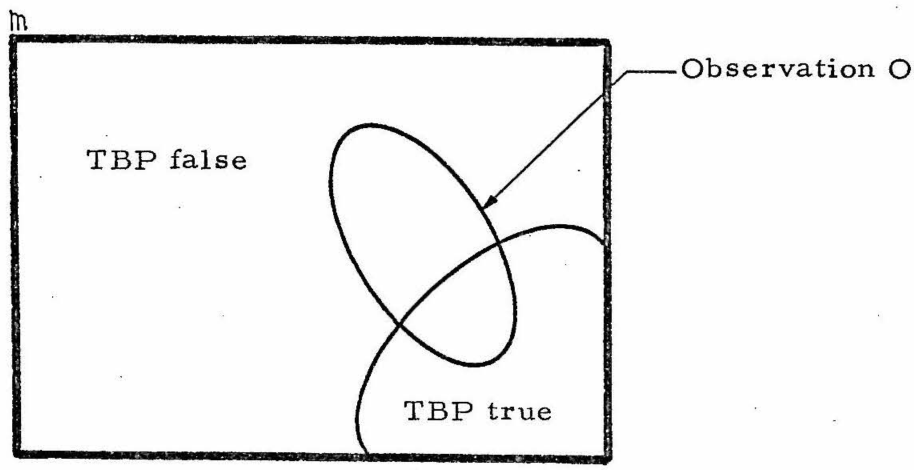
(2) On the other hand, the mathematician says that, in the given formal language, the physicist can not prescribe a set of sentences that defines the boundary B. The physicist's theory will include an object language counterpart of the statement TBP, but the mathematician's result, namely the Tarski-Banach Paradox [17, p. 51], [40, p. 244], shows that the physicist's theory must not be commensurate with what he really means. It must either (a) include models where the statement TBP (at the meta-level, of course) is true, or (b) exclude models where the statement is false, which the physicist has no a priori reason to exclude, and thus go

beyond what the physicist who believes in the first form of reductionism can accept. That is to say, the theory either does not express what the physicist believes, or it contains a metaphysical assumption not based on experimental evidence. In the diagram below the line labeled X corresponds to a sentence where the first of these is the case, and the line Y to a sentence where the second is the case.



(3) The situation would be quite different than it actually is if the physicist could conceive of an experiment whose partitions had boundaries coinciding with the boundary B . The fact is that the mathematician's proof of TBP depends upon a non-constructive argument. Clearly, it is impossible to show that the statement is not true, and to date no one has been able to cut a sphere in such a way as to demonstrate its truth. Thus, the physicist's a priori assumption can have no adverse effect on the course of his science, as guided by <u>experimental</u> evidence. Indeed, it is precisely the type

of metaphysical assumption that simplifies conceptual structure in a benign way. Graphically, the situation for realistically possible observations appears to be more restrictive than for sentences of the formal language. Thus, every observation partition will include models for which TBP is true and models for which TBP is false.



In the discussion on the meaning of scientific theories, we had left open the question of whether an empirical theory required tacit knowledge that could not be acquired by means of a finite number of observations. The above argument concerning the Tarski-Banach Paradox indicates that volume is exactly such a notion.

We have been talking about the condition where a set of sentences together with interpretations of the words constitutes a scientific theory, and what we would mean by such a theory. Now, returning to the example in the introduction concerning the relationships between abstract theory and empirical observations, we ask similar questions about what we mean by superstitions or myths. Again let us say that a mythology or body of superstitions partially consists of some set of sentences of a formal language. Certainly if such a set of sentences is strictly inconsistent, that is if the existential closure of its formulas has no model, then the associated state is empty. This corresponds to the case where no possible interpretation can make the sentences true and is clearly too strong a condition since for most superstitious beliefs, there is some conceivable configuration of the world in which these beliefs are true, although ludicrous. What is implied by stating that some set of sentences constitutes a superstition is that the sentences do not hold for the true state of the universe. But let us examine this more carefully. Suppose that we consider two sets of sentences which we will call $\mathcal{E}_{\text{science}}$ and $\mathcal{E}_{\text{superstition}}$. Envision a man who has verified through observation, at least to a reasonable degree, the theory expressed by $\mathcal{E}_{\text{science}}$. That is, he has a tacit understanding of the meanings of the referent words; let this be given by the interpretation φ. In terms of this, he has made a sufficient number of observations to convince himself of the truth of $\mathcal{E}_{\text{science}}$. We emphasize that he has carried out this verification of $\mathcal{E}_{\text{science}}$ in terms of his implicit knowledge, as embodied by φ. Now, holding this tacit understanding φ, the sentences $\mathcal{E}_{\text{superstition}}$ are indeed irrational.

On the other hand, it is certainly conceivable that another man at another time and in another place, having tacit understanding

φ' of the words of the language, quite different from that of our scientist, would indeed have verified by observation the truth of $\mathcal{E}_{superstition}$. Further, the tacitness of both φ and φ' imply that it may be impossible for either man to communicate to the other what he has in mind, nor would this be surprising if their experiences were vastly dissimilar. Thus, superstitions can not be claimed to be false, but merely untrue in terms of our current, tacit understanding of their referent words.

The second form of reductionism discussed in chapter I concerns the relationships among scientific theories; the important question in this regard is when can one scientific theory be reduced to another more fundamental theory? The analyses presented earlier in this chapter bear directly on this problem, and we will utilize several of the key concepts from them in commenting upon it.

In the introduction we mentioned that the theory of thermodynamics was demonstrated to be restatable in the framework of classical mechanics. We wish to examine the meaning of such a reduction, but we propose to consider a case of greater current interest. Suppose therefore that we are concerned with the reduction of biology to modern physics. We can assume that both of these theories are characterized by sets of sentences, say $\mathcal{E}_{biology}$ and $\mathcal{E}_{physics}$, together with appropriate interpretations of their referent words. Since these theories are both empirical in nature, we know that certain tacit knowledge, embodied in the interpretations, is necessary to fully understand what these theories are to mean. Let us go so far as to assume that some single scientist fully grasps the meanings of all the words germain to both domains; formally, this understanding is embodied in a single interpretation φ. This unrealistic assumption is made for the sake of convenience only and is not pertinent to our argument.

Now, consider actually reducing the theory of biology to physics. What does this mean? Well, if we are able to restate $\mathcal{E}_{\text{biology}}$ in terms of $\mathcal{E}_{\text{physics}}$ and φ then we will have succeeded. Thus, intuitively, we introduce a number of definitions of the referent words of $\mathcal{E}_{\text{biology}}$ characterized in terms of the words and relationships of $\mathcal{E}_{\text{physics}}$. Having done this, we can in principle then translate any sentence of $\mathcal{E}_{\text{biology}}$ into some set of sentences of $\mathcal{E}_{\text{physics}}$. But what will we have accomplished? We have reduced the tacit knowledge necessary to the theory of biology, but we are still left with the tacit knowledge inherent in physics. Therefore, even if this process can be carried out, we have not in any way reduced biology to absolutes as would be required by the first notion of reductionism. Can we reasonably expect to carry out this reduction of biology to physics? What we must do is to define all biological concepts in terms of the tacitly understood notions of physics, as specified by φ, and the laws and hypotheses expressed by $\mathcal{E}_{\text{physics}}$. As an example, consider the biological term "cell". We have previously indicated the possibility that the tacit nature of such concepts may require the use of an infinite, non-recursive set of sentences to characterize them. In this case, however, we do not need to characterize the meaning of "cell" absolutely, but only relative to the tacit knowledge of the theory of physics.

Because of the failure of the first form of reductionism, there are biological concepts, say for instance "cell", knowledge of which can not be gained from a finite number of observations without some prior tacit understanding. The second form of reductionism, for example the view that biology can be reduced to physics, implies that cells can be completely characterized by a finite number of observations, with the tacit knowledge limited to the notions of physics.

What makes this problem particularly difficult is that if one already has tacit knowledge of what a cell is, then statements of physics concerning the nature of cells can be highly informing, and one will respond to an appropriate physical description by saying "Yes, cells are like that". On the other hand, suppose one had never looked through a microscope or seen a drawing of a cell, nor had any knowledge of cytoplasm, cell membrane, metabolism, or the like. It seems extremely doubtful that a notion of cell could then be conveyed by a finite description. Certainly one could describe a somewhat more general notion, and it is tempting to say that a finite description could be constructed that would be adequate for all practical purposes. Long before a student had digested the physical description, however, it would seem that the temptation would be even stronger to send him to the microscope with the comment, "Once you've seen a few you'll understand".

Our point here is not to refute the second form of reductionism, but rather to make clear the nature of the argument against it. We have established the necessity of tacitly known concepts in science which can not be fully confined with fewer than an infinite number of observations. The existence of such entities clearly makes it plausible that certain concepts relevant to one science may only be reducible to the tacitly known concepts of a second science through an infinite, non-recursive description or an infinite number of observations. This would be especially so when the laboratory experiences are as disparate as those of biology vis-à-vis physics, or psychology vis-à-vis biology.

VI. INFORMATION AS RELATED TO LANGUAGE CHANGE

In the preceding chapters, we have primarily been concerned with the characterization of a single individual's view of some domain as reflected in the particular formal language he employs to describe it. Now, however, we wish to consider some of the problems inherent in the communication process. Since communication necessarily involves two or more individuals, we must deal simultaneously with more than one formal language. As indicated previously, even though two individuals may both consider a given string of words to be grammatical, they need not assign equivalent meanings to it. That is, the sentence may result in a totally different partitioning of the model space for the two individuals. Furthermore, even if they do agree completely on the meaning of the sentence, the fact that they may have different a priori probabilities could induce correspondingly different *-probabilities and thereby result in one of them believing more strongly than the other in the truth of the sentence.

In order to clarify the possible relationships among different formal languages, we will introduce several auxiliary concepts based on our previous definitions. One of the things we would like to know is when one formal language can be considered to be more expressive than another. This is related to the sets of states of the two languages.

Definition: Let $\mathfrak{S}_1$, $\mathfrak{S}_2$ be the sets of states of the formal languages $\mathcal{L}_1$, $\mathcal{L}_2$, and let $\mathfrak{B}_1$, $\mathfrak{B}_2$ be bases for $\mathcal{L}_1$, $\mathcal{L}_2$. Then $\mathcal{L}_2$ is at least as expressive as $\mathcal{L}_1 (\mathcal{L}_2 \geq \mathcal{L}_1)$ if $\mathfrak{S}_1 \subseteq \mathfrak{S}_2$ and if for every basis element $\mathfrak{h} \in \mathfrak{B}_1$, there is a basis element $\mathfrak{h}' \in \mathfrak{B}_2$ such that $\mathfrak{h}' \subseteq \mathfrak{h}$.

To properly motivate this definition, we must consider what the intuitive meaning of expressiveness is. As we will use the term, it is intimately linked to the commonly accepted notion of precision of expression. That is, if two individuals are both asked to describe some event which they both have witnessed, one of them might reply by enunciating the statements γ_1, γ_2, and γ_3. The other might then agree that those statements were indeed true of what they observed, but might also add that γ_4, γ_5, and γ_6 were also true, thereby presumably giving us a more refined picture of the actual circumstances. The implication here is that even though the two individuals may have made the same observations, one of them expresses details of these observations which the other considers irrelevant. We have previously emphasized that the individual's formal language directly reflects not only that which he is capable of expressing, but necessarily also that which he considers to be relevant.

Our notion of expressiveness, however, can be considered applicable in a somewhat broader context than the previous example indicates. For instance, suppose the two individuals described above consider themselves to be in fundamental disagreement about the event which occurred. That is, one of them either can make no sense of what the other says, or he feels that it is untrue of what happened. This certainly need not mean that one of them is wrong; it may simply be that certain strings of words have radically different interpretations relative to their respective formal languages. Thus, if we were sufficiently omniscient to be able to comprehend exactly what each observer means, then we might very well be able to conclude not only that there was no basic disagreement, but also that one individual had rendered a more precise description than the other.

In view of this discussion, let us now examine the specifics of our definition. The first requirement is that every state of the less expressive language must also be a state of the more expressive. Notice that this does not imply that the same strings of words are used to define these states in both of the languages, but in general the more expressive language will be capable of characterizing a greater number of states. The second condition pertains to the bases of the two languages. The requirement here is that for each basis element of the less expressive language there is a basis element of the more expressive language contained in it. However, the more expressive language may have additional basis elements which are not contained in any of the basis elements of the less expressive language. Recall that the basis is a set of minimal states which are pairwise distinguishable and that these states represent mutually exclusive configurations of the universe relative to the language. The basis elements thus form what might be called a kernel set of states for the language, and our requirement is that if one language is to be considered more expressive than another it must be capable of more precise expression with respect to this kernel.

The relation of expressiveness can easily be seen to be both reflexive and transitive, and thus it induces a pre-ordering on the set of all formal languages and their bases, and a partial ordering of their state-partitions on the set of models. We emphasize that expressiveness is related primarily to the semantics of formal languages since it is a condition on the state diagram, and it is only indirectly tied to the syntax and the specific vocabulary of the languages. Nevertheless, our expressiveness relation is quite strong since it demands that the state diagram for one language be a proper partitioning of the state diagram for the other. There may be somewhat weaker conditions under which one language could be said to be more expressive than another, but even our relatively restrictive definition has some surprising implications.

One intuitively feels that the more expressive a language is the more suitable it is as a descriptive tool. The question we wish to consider next is whether more expressive languages are more informative to the individual who employs them. In other words, if an individual makes some observation O which he can describe in a formal language $\mathcal{L}_1$, and the resulting information gained is I_1, then if he had some more expressive language $\mathcal{L}_2 \geq \mathcal{L}_1$, would the informativeness of the same observation O necessarily be equal to or greater than I_1 ? To aid in the investigation of this, we shall consider an example. Figures 9, 10, and 11 show the state diagrams for three hypothetical languages $\mathcal{L}_1$, $\mathcal{L}_2$, and $\mathcal{L}_3$. Each of these languages has been obtained from the preceding one by the addition of some new sentence. Thus, $\mathcal{L}_2$ is the same as $\mathcal{L}_1$ with addition of the single sentence γ_3 , and similarly $\mathcal{L}_3$ results from adding the sentence γ_4 to $\mathcal{L}_2$. This is one trivial way of obtaining a more expressive language, as our intuition would suggest. In each case, the basis for the language is shown by the elements with the heavy borders. We remark that $\mathcal{L}_1$ has a unique basis, whereas $\mathcal{L}_2$ and $\mathcal{L}_3$ do not; $\mathcal{L}_2$ has the same state diagram as the language shown in Figure 5. The fact that $\mathcal{L}_2$ and $\mathcal{L}_3$ do not possess unique bases is unimportant for the present, but we will comment on the implications of this later in the chapter. As before, we assume that the a priori probability P is proportional to area on the diagrams. Thus, for example, in Figure 9 the a priori probability of the minimal state TT, P(TT), is equal to 1/9, and in Figure 11, P(T∅F∅) = 1/36. Referring to our definition of expressiveness, we see immediately that $\mathcal{L}_3 \geq \mathcal{L}_2 \geq \mathcal{L}_1$.

On each diagram, we have illustrated three observations O_1 , O_2 and O_3 . These represent typical observations which might be made, but they are not presumed to be related to each other in any

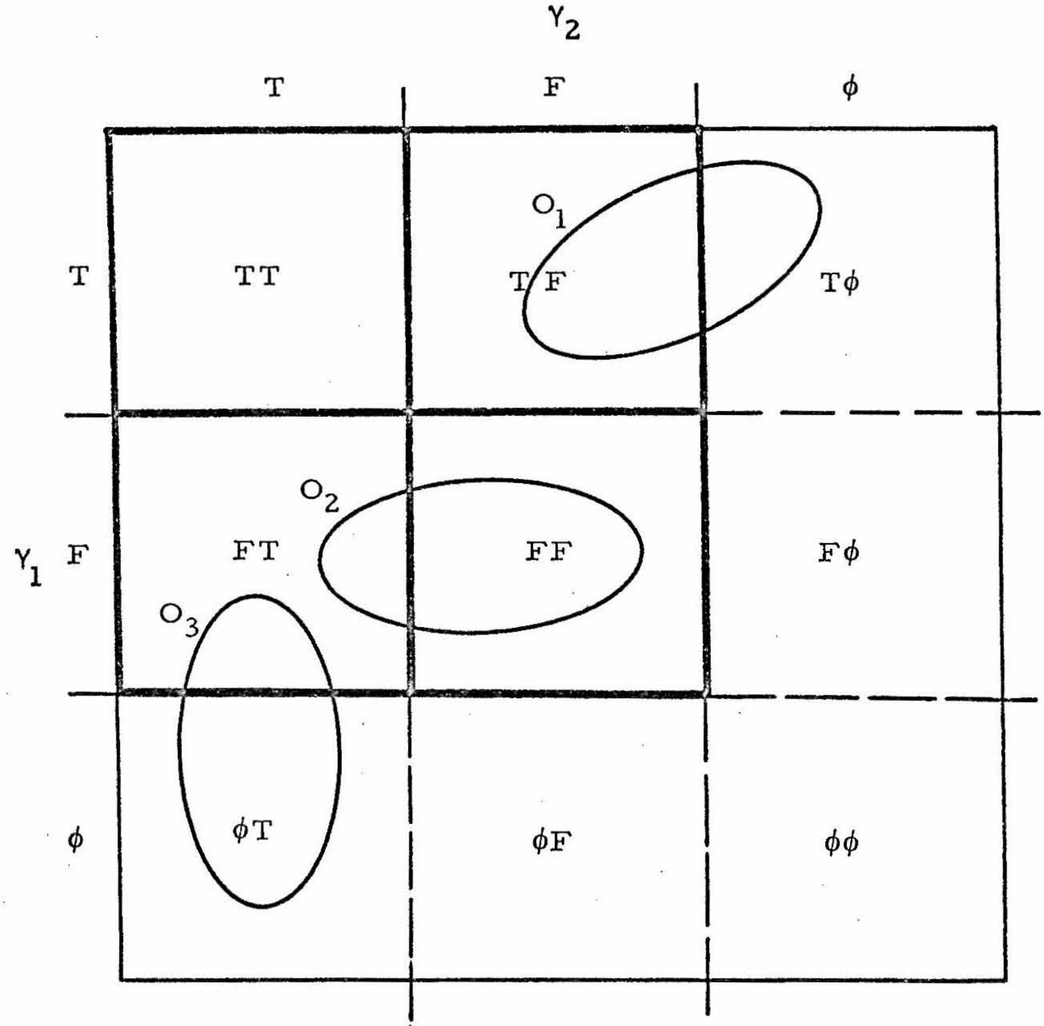


Figure 9. State diagram for language $\mathcal{L}_1$ and three possible observations, O_1, O_2, and O_3. The regions with heavy borders are basis elements.

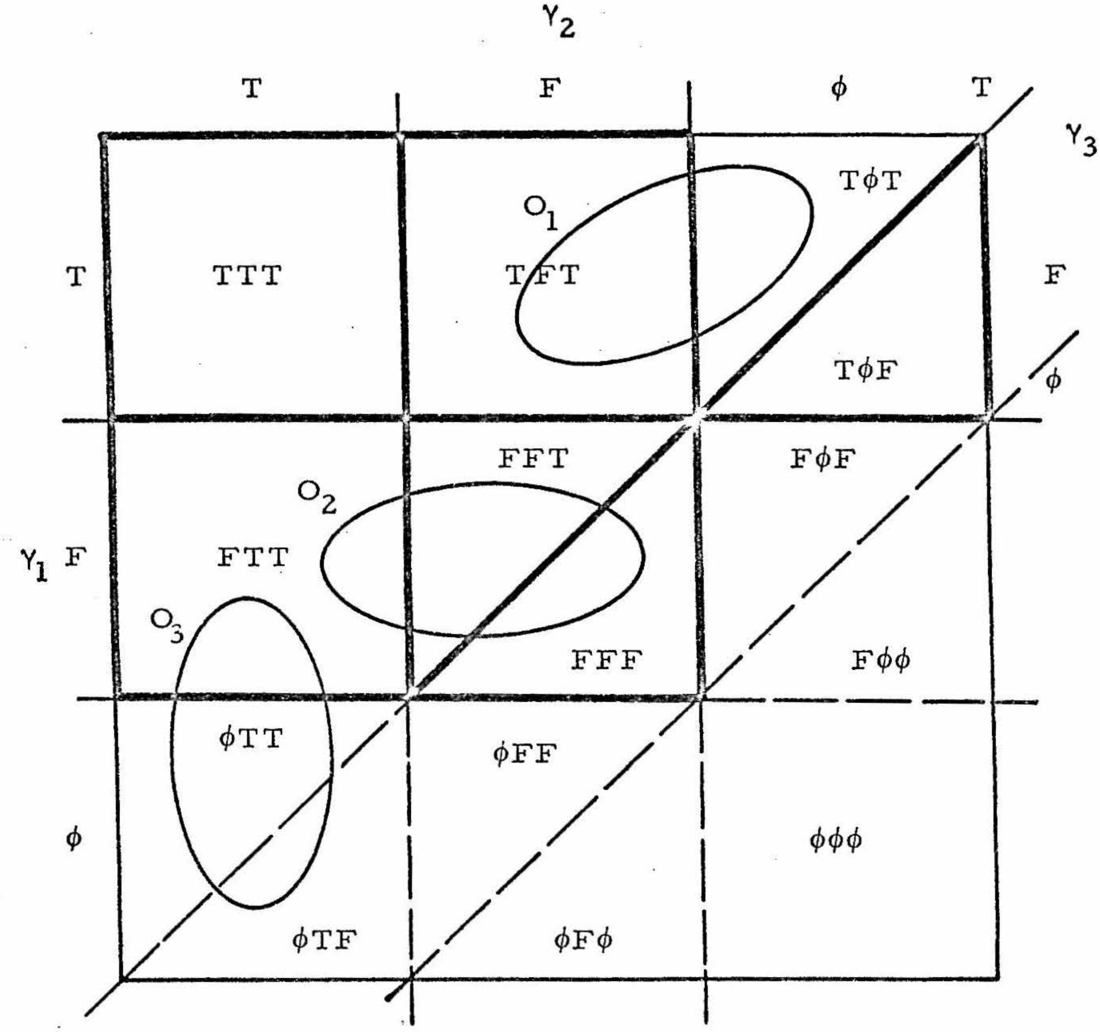


Figure 10. State diagram for language $\mathcal{L}_2$ showing the same observations as in Figure 9. The regions with heavy borders are basis elements.

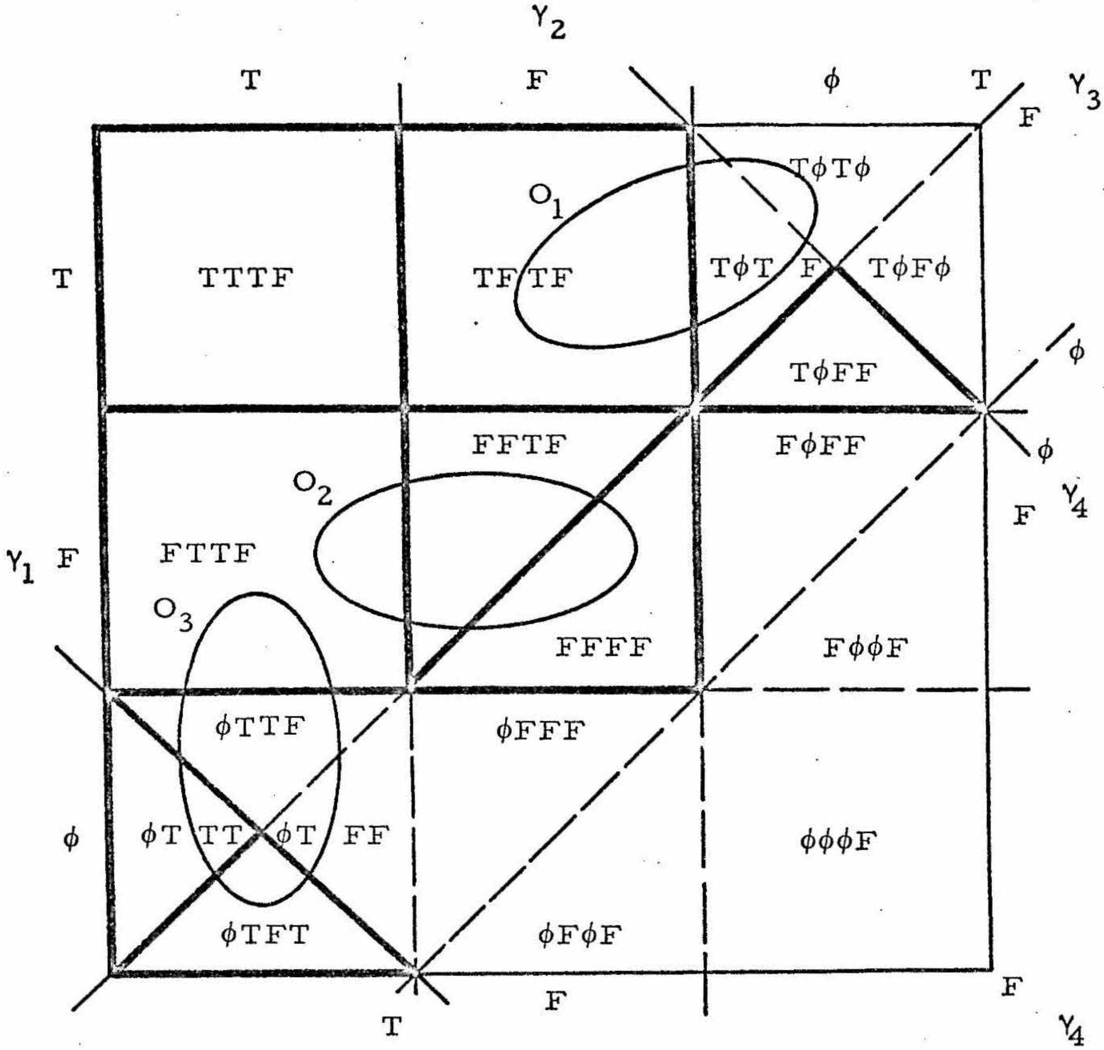


Figure 11. State diagram for language $\mathcal{L}_3$ showing the same observations as in Figures 9 and 10. The regions with heavy borders are basis elements.

specific way. Thus, we will assume that there are three observers whose views of some domain are characterized by $\mathcal{L}_1$, $\mathcal{L}_2$, and $\mathcal{L}_3$, and we shall determine the *-probability of the observations O_1, O_2, and O_3 for each language. The result of this will indicate the relative informativeness of these observations with respect to the observers. Recall that the *-probability of an observation O is given by:

$$P^*(\mathfrak{m}_O) = \frac{\overline{P}\left(\bigcup_{\substack{B \in \mathfrak{B} \\ B /\!/ \mathfrak{m}_O}} B\right)}{\overline{P}\left(\bigcup_{B \in \mathfrak{B}} B\right)} .$$

We first calculate the a priori probability P (in these examples P and $\overline{P}$ are identical) of the union of all basis elements, which we denote as X^j for language $\mathcal{L}_j$. This is done simply by inspection of the three diagrams.

(1) For $\mathcal{L}_1$, $P(X^1) = 4/9$.

(2) For $\mathcal{L}_2$, $P(X^2) = 9/18$.

(3) For $\mathcal{L}_3$, $P(X^3) = 19/36$.

Now, let Y_i^j be the union of all basis elements of language $\mathcal{L}_j$ which are indistinguishable from observation O_i. We may then compute $P_j^*(\mathfrak{m}_{O_i})$ by the following formula:

$$P_j^*(\mathfrak{m}_{O_i}) = \frac{P(Y_i^j)}{P(X^j)} .$$

As one instance of this computation consider O_3 and $\mathcal{L}_2$ in Figure 10. In this case Y_i^j is the following:

$$Y_3^2 = \{TTT \cup FTT \cup T\emptyset F\} ,$$

and therefore

$$P_2^*(m_{O_3}) = \frac{P(Y_3^2)}{P(X^2)} = \frac{5/18}{9/18} = 5/9 .$$

The following table summarizes the value of P^* for each observation and each language:

	$\mathcal{L}_1$		$\mathcal{L}_2$		$\mathcal{L}_3$
O_1	1/2	>	4/9	<	9/19
O_2	1/2	>	4/9	>	8/19
O_3	1/2	<	5/9	<	11/19

Each of the observations thus exhibits a different behavior as the language changes from $\mathcal{L}_1$ to $\mathcal{L}_2$ to $\mathcal{L}_3$. Since the informativeness of an observation relative to a language is defined in terms of the logarithm of the *-probability, the informativeness becomes relatively higher as the *-probability decreases toward zero. Thus, the inequalities shown in the table among the probabilities are reversed when we consider the associated informativeness. Observation O_2 behaves in the manner we might at first expect. That is, as the language changes to become more expressive the expectation, relative to the language, of observing O_2 decreases; and hence O_2 is most informative for language $\mathcal{L}_3$. On the other hand, both O_1 and O_3 demonstrate a rather surprising phenomenon, namely that it is possible for the information obtained from an observation to decrease as one moves to a more expressive language. O_1 is particularly interesting because the information is greatest relative to language $\mathcal{L}_2$. What this suggests is that there may be formal languages which are optimum, in the sense of informativeness, for the description of certain classes of observations, and further that these are not necessarily the most expressive languages which could be employed. More specifically, for observation O_1, the language $\mathcal{L}_2$ is such that if we employ either the more expressive language $\mathcal{L}_3$ or the less expressive language $\mathcal{L}_1$, then the information we obtain on observing O_1 decreases.

This example again demonstrates the significant role of language in the characterization of an individual's view of his universe. If his language provides a rather loose or vague description of the phenomena he observes, there will be many detailed observations he can make which will yield little information; and conversely, if his language is overly expressive, the additional structural complexity of the language may diminish the informativeness of certain other observations. The reasons for this are directly related to the notion of meaninglessness; the fact that a formal language always makes various implicit or unstatable assumptions about the structural nature of its domain of discourse—as embodied in the semantic categories, for instance—causes sentences of the language to become meaningless on some models. This, in turn, means that indistinguishability is not an equivalence relation, as we have shown, and consequently very expressive languages may not be particularly informative for certain classes of observations.

Languages $\mathcal{L}_2$ and $\mathcal{L}_3$, shown in Figures 10 and 11, do not possess unique bases, as we mentioned previously. In addition to the bases outlined in the diagram, $\mathcal{L}_2$ has another basis, namely

$$\mathfrak{B} = \{TTT, TFT, FTT, FFT, FFF, \phi TF\} ,$$

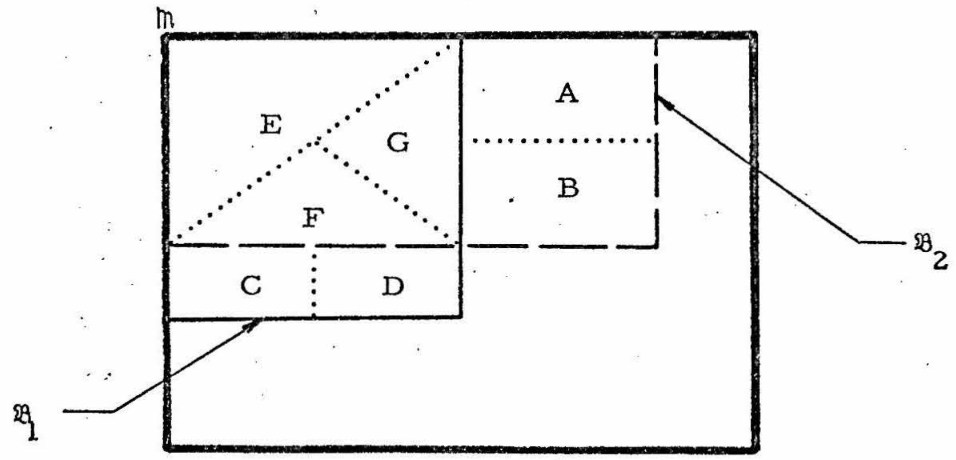
and similarly $\mathcal{L}_3$ also has the basis

$$\mathfrak{B} = \{TTTF, TFTF, FTTF, FFTF, FFFF, \phi TTT, \phi TFT, \phi TFF\} .$$

These can easily be shown to be the only other bases for $\mathcal{L}_2$ and $\mathcal{L}_3$. The question is, what are the implications of a non-unique basis for the probability and hence the informativeness of some observation? In the particular example we have chosen, if we associate the two bases above with $\mathcal{L}_2$ and $\mathcal{L}_3$, in lieu of those shown in Figures 10 and 11, then it can be verified that the probabilities assigned to the three observations O_1, O_2, and O_3 are unchanged for both languages. Consequently, the informational relationships among the languages also remain the same for these observations. In general, however, there will be other observations for which the choice of basis does affect the probability. This is particularly significant since it means that the syntax and semantics of a formal language may not be sufficient to uniquely determine the *-probability of all observations. More specifically, knowledge of the truth values of all sentences on all models may not characterize a unique basis. As a result, there may be situations where a particular formal language has more than one basis, and thus different *-probabilities could conceivably be associated with a single observation; and yet it would normally be assumed that a rational individual acts as though a given observation has only a single probability. This implies that an appropriate basis for the language must be chosen to satisfy criteria which we have not explicitly treated. The resolution of this question seems to lie in considerations involving the dynamics of language change. That is, as an individual's language evolves or develops to reflect his growing knowledge and changing perspective on his world, the mechanisms which determine exactly how his language will change presumably take into account aspects of his past experience, thereby causing one basis of his new language to become preferable to others. As we have previously indicated, an investigation into the nature of the processes involved in language change is beyond the scope of this thesis.

It is interesting to note that if a formal language does not possess a unique basis, there is nevertheless a more expressive formal language which does have a unique basis. The more expressive language simply includes sentences which distinguish the previously indistinguishable minimal states of the less expressive language. Of course, the converse is also true, as the previous example shows; that is, for a language with a unique basis there is a more expressive language with a non-unique basis. The languages $\mathcal{L}_2$ and $\mathcal{L}_1$ in Figures 9 and 10 are related in this way. We point this out merely to illustrate that for more realistic languages than those used in our examples, it is not necessarily the case that they would especially tend to have either a unique basis or multiple bases.

The example of Figures 9, 10, and 11 has shown that the information content of an observation does not generally strictly increase or decrease with the expressiveness of the language describing it. However, supposing that the languages we had used each possessed a unique basis, can we then demonstrate that the information would either increase or decrease monotonically with expressiveness? Again the answer is no, but rather than exhibiting further examples we will give a detailed analysis of the conditions under which information increases or decreases.



Suppose that we are given two formal languages $\mathfrak{L}_1$ and $\mathfrak{L}_2$, where $\mathfrak{L}_2 \geq \mathfrak{L}_1$, and suppose they have bases $\mathfrak{B}_1$ and $\mathfrak{B}_2$. The preceding diagram shows both bases superimposed on the set of models $\mathfrak{m}$.

The representation of the two bases is strictly schematic; the various lettered regions indicate sets of models contained in the basis and grouped together according to their indistinguishability or distinguishability relative to some set of models $\mathfrak{h} \subseteq \mathfrak{m}$ which is assumed to be associated with an observation. Basis $\mathfrak{B}_1$ consists of the sets C, D, E, F, G, and basis $\mathfrak{B}_2$ of the sets A, B, E, F, G. These sets are characterized by the following relationships with respect to the set of models $\mathfrak{h}$.

A: $\mathfrak{h} /_2 A$ B: $\mathfrak{h} //_2 B$

C: $\mathfrak{h} /_1 C$ D: $\mathfrak{h} //_1 D$

E: $\mathfrak{h} /_{1,2} E$ F: $\mathfrak{h} //_{1,2} F$

G: $\mathfrak{h} //_1 G$ and $\mathfrak{h} /_2 G$

The subscripts indicate for which of the two languages the relation holds. Thus, for example, $\mathfrak{h} //_1 D$ indicates that the set of models D is indistinguishable from $\mathfrak{h}$ in language $\mathfrak{L}_1$. Also, the models in D are not included in basis $\mathfrak{B}_2$. Assuming each of these sets is measurable, we can express the *-probabilities, $P_1^*(\mathfrak{h})$ and $P_2^*(\mathfrak{h})$, as follows:

$$P_1^*(h) = \frac{P(F) + P(G) + P(D)}{P(F) + P(G) + P(D) + P(C) + P(E)}$$

$$P_2^*(h) = \frac{P(F) + P(B)}{P(F) + P(G) + P(B) + P(A) + P(E)}$$

$P_1^*(h)$ is the *-probability of h for language $\mathcal{L}_1$, and $P_2^*(h)$ is the *-probability of h for language $\mathcal{L}_2$. The numerator of each fraction is the a priori probability of those models in the basis which are indistinguishable from h, and the denominator is the a priori probability of the union of all basis elements.

Since we wish to examine the difference in information content of the observation corresponding to h as the language changes from $\mathcal{L}_1$ to the more expressive language $\mathcal{L}_2$, we can use the ratio $P_1^*(h)/P_2^*(h)$ to accomplish this. That is, if for some specific h the ratio equals one, then the information is unchanged; and if the ratio is greater than one, $\mathcal{L}_2$ is more informative than $\mathcal{L}_1$. Conversely, if the ratio is less than one, $\mathcal{L}_2$ is less informative than $\mathcal{L}_1$, with respect to h. Notice that the sets of models E and F are defined so that they each have the same relationship to h in both languages; the others, however, namely A, B, C, D, and G, differ in their properties of distinguishability or indistinguishability as the language changes. An increase in the measure of any one of these sets, for example in P(A), affects the ratio as follows:

Increase in...	Change in ratio $P_1^*(h)/P_2^*(h)$
P(A)	increase
P(B)	decrease
P(C)	decrease
P(D)	increase
P(G)	increase

In each of the cases, all other quantities are assumed to remain fixed. The table shows that an increase in either P(A), P(D), or P(G) causes an increase in the ratio and hence an increase in the informativeness of the observation corresponding to h. Similarly, increasing P(B) or P(C) results in a decrease in informativeness. Note that h is assumed to be the same set of models in both languages. Also, changing from some h to a different h', will generally redistribute the sets of models among the various categories. However, the sets corresponding to $A \cup B$, $C \cup D$, and $E \cup F \cup G$ are fixed by the two languages $\mathcal{L}_1$ and $\mathcal{L}_2$ and are independent of h.

Before we interpret the intuitive meanings of the various possible changes in A, B, C, D, and G, we would like to establish that these are not interdependent. That is, in order to make this analysis meaningful we must show that each of the sets of models corresponding to A, B, C, D, and G may either be empty or non-empty in any combination. The following theorem shows that this is in general possible.

Theorem: There exist formal languages $\mathcal{L}_1$ and $\mathcal{L}_2 \geq \mathcal{L}_1$, such that there are sets of models η (corresponding to possible observations) for which the sets A, B, C, D, and G may be empty or non-empty in any combination, i.e. A, B, C, D, and G are independent of one another.

Proof: The proof is by specific examples, and because of its length it is contained in the appendix.

The preceding theorem implies that for two formal languages— one more expressive than the other—there will generally be some observations which are less informative relative to the stronger language; and further that the change in informativeness from one language to another is effected by variations in the measures of the sets A, B, C, D, and G, each of which could be large or small independent of the others. For example, consider the sets A and B. How can they be interpreted? Their union, $A \cup B$ contains all basis elements of the more expressive language $\mathcal{L}_2$ which are disjoint from every basis element of $\mathcal{L}_1$. These additional minimal states therefore represent new theories which are alternatives to those describable in $\mathcal{L}_1$. Here, we are identifying minimal states with complete theories within the language. That is, as we have previously discussed, the sentences defining a minimal state characterize some configuration of the universe as precisely as the language allows. The observation in question, namely the one associated with η, either partially confirms or denies the theories constituting A , and hence the information gained tends to increase in this more expressive language. On the other hand, the minimal states which B comprises are indistinguishable from $\mathfrak{h}$ and therefore, as theories, are generally neither refuted nor verified by the observation. Consequently, the information tends to decrease since the observation does not aid in resolving among the additional theories represented by B .

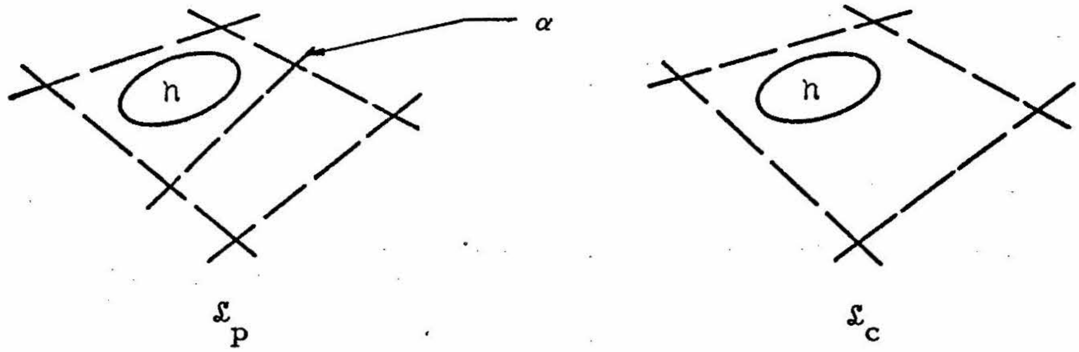
Now consider the sets C and D ; $C \cup D$ is composed of models which were members of basis $\mathfrak{B}_1$ but are not members of basis $\mathfrak{B}_2$. These excluded models are therefore indistinguishable from some basis elements of $\mathfrak{L}_2$. The members of the set C were previously distinguishable from the observation and thus represented distinct alternatives within the less expressive language. Their elimination from the basis $\mathfrak{B}_2$ means that they are no longer considered viable alternatives, and since $\mathfrak{h}$ is now distinguishable from fewer models in the basis, it becomes a less informing observation relative to $\mathfrak{L}_2$. The converse is true of the set D ; the models in D were not distinguishable from $\mathfrak{h}$ in $\mathfrak{L}_1$ and consequently, eliminating them from $\mathfrak{B}_2$ has the effect of increasing the informativeness of the observation.

Finally, the set G consists of models which are members of both bases, $\mathfrak{B}_1$ and $\mathfrak{B}_2$. In language $\mathfrak{L}_1$ these models were indistinguishable from the set $\mathfrak{h}$ and were thus considered to be equivalent to the associated observation. In language $\mathcal{L}_2$, however, these same models are distinguishable from $\mathfrak{h}$, and therefore $\mathcal{L}_2$ has sentences which make this distinction. This means that $\mathcal{L}_2$ can more precisely delimit the set of models $\mathfrak{h}$ than $\mathcal{L}_1$ can. Consequently, $\mathcal{L}_2$ is more informative with respect to this observation than $\mathcal{L}_1$ since the *-probability of $\mathfrak{h}$ will be less for $\mathcal{L}_2$ by an amount proportional to the measure of G, P(G) .

The result that increasing expressiveness may reduce information is at first contrary to intuition. This appears to stem from the fact that one considers that the conceptual apparatus of any reasonable language should be adequate to make all relevant distinctions among configurations of the universe. If this were the case, the only change in basis that could take place between one language and a second more expressive language would be of type G , and thus information would increase monotonically with expressiveness. This is not so. The classic result establishing this fact is Tarski's theorem on truth [39, p. 247], which states that the notion of truth for a formal language of sufficient expressive power can only be stated within another formal language that is necessarily more expressive. Once this essential inadequacy of the descriptive power of a language is seen, the potential for other changes in basis becomes apparent and hence the possibility for the reduction of information on passage to a more expressive language.

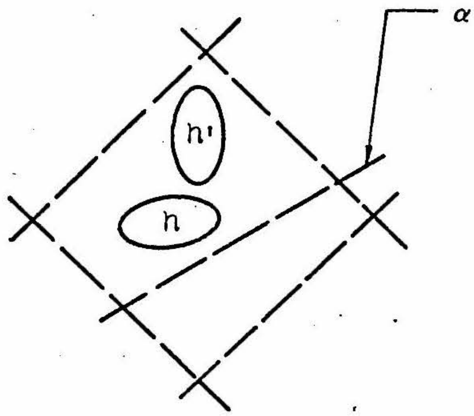
Having established some general properties of the change in information with language, we will now re-examine the problem of ramification of meaning which was presented by an example in the introduction. Recall that the example we used concerned two individuals, each of whom had a clearly different concept of the meaning of the word "transistor". In view of the apparatus we have developed we can now see that these two individuals can be considered to utilize two distinct, although related, formal languages. Each language thus embodies one individual's conceptualization of the relevant properties of transistors. We expect that each man has certain observational experience pertinent to transistors, and further that these two sets of observations are related, exactly in the sense defined in chapter III. As we have implied, we certainly can assume that their formal languages are in some way related since they are both concerned with describing and understanding many of the same basic phenomena. How might their languages be related? One plausible relationship would be that one language is more expressive than the other. That is, the solid-state physicist's language $\mathcal{L}_p$ may be strictly greater in expressive power than the language $\mathcal{L}_c$ of the circuit designer. In this relatively ideal case, the state diagram for the language $\mathcal{L}_p$ would properly partition the state diagram for $\mathcal{L}_c$. The following two figures illustrate corresponding portions of the total state diagrams for the languages.

The set of models $\mathfrak{h}$ corresponds to the observation which has been made by both individuals; we assume that this observation involves some transistor which is behaving in an interesting or novel

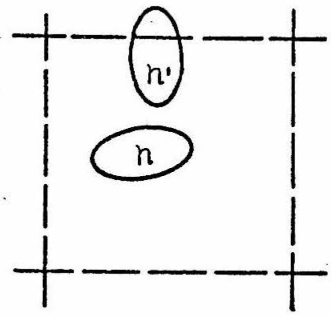


fashion. The dashed lines partition the model space in the usual way and are presumed to be associated with sentences about transistors. In this special case, we will further assume that the four sentences shown for $\mathcal{L}_c$ are identical, syntactically and semantically, with the corresponding four sentences of $\mathcal{L}_p$, i. e. those other than α. Consequently, if the physicist enunciates one of these statements, the circuit designer will agree that it is true of the observation, and vice-versa. But what of the statement α? α is a sentence of the physicist's more expressive language $\mathcal{L}_p$, but if it is also a sentence of $\mathcal{L}_c$ it does not have the same meaning and is therefore not useful in characterizing h. Thus, the physicist assigns a meaning to α in such a way that it is relevant to the observation. On the other hand, the circuit designer considers α either incomprehensible or irrelevant, i. e. meaningless on h or unnecessary to characterize h. Nevertheless, communication still takes place on a limited basis since there is some fundamental agreement about what has been observed. Even regarding α, there may be no disagreement; it is simply a non sequitur to the circuit designer. Notice that h is most precisely delimited by the language $\mathcal{L}_p$. This would indicate that the observation that was made is of greater importance to the physicist and presumably is more informative. As we have emphasized in this chapter, however, we can not conclude in general that h is more informing relative to $\mathcal{L}_p$ without knowing something about the other alternative theories and hypotheses within the language.

The preceding example is idealistic in that we have assumed complete agreement in meaning for at least some of the sentences pertinent to the observation. Thus, with the exception of the statement α, $\mathcal{L}_p$ and $\mathcal{L}_c$ are inter-translatable; at least this is true at the meta-level. It will generally not be true, however, that either individual "knows" that the meaning he attaches to some sentence is exactly the same as the meaning assigned to that sentence by the other individual. Indeed, it is possible that there will be an apparent agreement in concepts, but that this may break down on the basis of additional evidence. Suppose, for example, that the preceding diagrams were actually as shown in the following figures. Here we assume that the four sentences (excluding α) in $\mathcal{L}'_p$ are composed of the same words as the four sentences shown for $\mathcal{L}'_c$, although each individual clearly associates a somewhat different meaning



$\mathcal{L}'_p$



$\mathcal{L}'_c$

with them. h and h' are sets of models characterized by two distinct observations involving transistors. If both individuals make the observation corresponding to h, then the situation is almost exactly as before—namely there is agreement on all of the statements except possibly α. On the other hand, if the actual observation corresponded instead to h', then one of the statements which the physicist says is true of the observation, although understood by the circuit designer, is not strictly true of h' in his language $\mathcal{L}'_c$. The important point here is that as long as the observations shared by the two individuals have the characteristics of h, their languages will seemingly agree, in spite of the fact that the sentences do not have identical meanings. Only when some observation is associated with a set of models similar to h' will the actual conceptual differences be made explicit. In our example, h' might possibly be the set of models resulting from an observation of a peculiar type of semiconductor device which functions like a transistor in the physicist's view, but which the circuit designer does not consider to be a legitimate transistor because, for example, it only behaves properly at very high temperatures.

Notice that observations which make explicit the underlying differences in two languages can occur even when one language would be said to more precisely specify some concept, such as transistor, than the other. Consequently, the problem of communication between two individuals having more or less ramified views of some domain has two primary aspects. First, there may be statements like α which are relevant to one observer and not to the other. These are no fundamental barrier to communication since there may be apparent agreement on sufficiently many other statements so that ideas may still be adequately exchanged. Nevertheless, it is statements like α which tend to make certain observations more informative to one individual than the other, and hence to motivate or prompt communication. Therefore, the most relevant or precise statements which can be made are exactly those for which the communication breaks down, forcing one individual, or possibly both, to describe the observations in more general terms.

Secondly, this necessary abstraction from the most precise statements that could be made, even though it enables communication, may mask certain inconsistencies among the languages. These inconsistencies may be revealed by observations which have not as yet been made, but it is possible that such observations will never be made by the two individuals. In the latter case, there will always appear to be an agreement among the individuals, in spite of the fact that they each have a different conceptual view of the same domain.

VII. THE PROBLEM OF INDUCTION

Certainly one of the central problems in understanding science is to comprehend the inductive process—namely, how one arrives at general conclusions on the basis of specific observational evidence. The solution to this problem is one of the ultimate aims of this thesis, and although we have not been able to solve it directly, we have provided mechanisms that allow us to propose two new hypotheses. These hypotheses suggest directions for further research which may prove fruitful in resolving this difficult problem of the modern philosophy of science. As these hypotheses are firmly founded on our developed formalism, we consider them one of the central results of this thesis, and this chapter will be devoted primarily to a presentation of them and an investigation of their implications.

In chapter I we introduced a number of problems pertinent to an understanding of the scientific method, all but two of which have been discussed in terms of our formal apparatus. The remaining two concern the reinterpretation of the words of a language and the apparent disparity between the explicit and implicit scope of a scientific paper. We first show that both of these are instances of the general problem of induction. Thus consider the circumstances which cause a particular word to assume some new interpretation; this clearly comes about as a consequence of some relevant observational evidence. Suppose for example that we have observed five hundred white ducks and no black ducks. It is highly likely that our definition of "duck" includes the property of their being white. If we now see a black duck we have two obvious choices; we must either modify or reinterpret the meaning of the word "duck", or we can conclude that this new creature we have seen is not a "duck" at all. However, if we continue to observe black ducks, this latter course will become less and less natural until finally the lack of sensible differentiation, on some basis other than color, will cause the term "duck" to be redefined. Thus the accumulation of observational evidence refuting our belief that "ducks" are white will eventually induce a different meaning for the word "duck". The exact moment when this occurs can be characterized only if we fully understand the inductive process.

Now consider a scientific paper describing some experiment and the results obtained from it. Typically, the paper will detail the experimental apparatus, the experimenter's techniques, and the observed outcome of the experiment. Usually, if it is feasible, the experiment will have been repeated a number of times in order to reassure both the scientist and the reader that the specific outcome of the experiment was not a fluke or chance occurrence. In spite of this repetition, however, the actual circumstances under which the experiment was performed represent only an isolated instance out of many which might have been chosen. Indeed, the paper will generally make clear many of the specific restrictions of the experiment, and very often the explicitly stated conclusions will be confined to the most direct consequences of the observed results. However, as time goes by, and further consistent evidence is obtained, some subsequent paper may make the claim that "all entities of the type studied will behave in a manner consistent with the experimental results presented". Thus the inductive leap is made explicit. Long before this occurs though, the dissemination of previous results may lead others to inductively infer this same conclusion. Such implicit generalizations are yet another example of the inductive process.

In light of our previous discussions, it can easily be seen that induction may be validly characterized as a change in language. However, before commenting further on this, let us briefly examine the possibilities inherent in language change. First, we know that for many concepts, such as the integers or volume, there is no formal language which can constructively characterize them. Thus, any mechanism describing language change will necessarily fail to explain how such concepts can be sensibly interpreted, without some reference to extra-linguistic aspects of their meanings. Another salient point is that there is no most expressive language. This results directly from the Tarski theorem on the definability of truth for a language [39, p. 273], since if there were such a language, say $\mathcal{L}$, we would be unable to specify the meaning of "true in $\mathcal{L}$" for lack of a more expressive language in which to do this. This contradicts the fact that a meta-language for $\mathcal{L}$ can be formalized and truth for $\mathcal{L}$ defined within it. Consequently, continually moving from one language to another more expressive language is a non-terminating process. Furthermore, as we have seen, we may actually be losing information, thereby defeating our purpose. These basic limitations on language change suggest that the forces leading to such change are the product of other considerations.

One possibility is that the formal language that characterizes a person's view at a given instant is determined by his a priori probability and the sum total of his observational evidence. In situations where the need for tacit knowledge can be limited only by an infinite number of observations, any finite number of observations will leave room for language change compatible with these observations. Thus, one would expect an adjustment of tacit knowledge to take place as a function of the a priori probability, and within the latitude remaining over and above available observational evidence. What remains is to identify the forcing function or criterion involved. The literature considers, for example, such forcing functions as simplicity, though these remain ill-defined.

It is interesting to note that traditional treatments of induction have been based primarily on a fixed language which is presumed to be adequate to express all relevant concepts. By using a fixed language and defining induction within its rigid framework, several aspects we feel are basic to the inductive process have been overlooked. For instance, this precludes changes in the meanings of words as a factor in explicating induction. Also, if the language is static we can not make use of shifts in the level of abstraction or expressiveness relative to some concept. Furthermore, even though tacit knowledge—as embodied by the interpretation of referent words—has been generally accepted, its dynamic role in induction does not seem to have been recognized. As we mentioned, tacit understanding by its very nature allows for a certain flexibility in an observer's language. Thus linguistic changes can be made which do not conflict with the observer's previous observational experience, except in the limiting case of an infinite number of observations. Because we are convinced that the essence of the inductive process involves considerations of the temporal aspects of language, we seek an explanation which can fully utilize changes in the meanings of words, shifts in the generality or abstraction of language, and the freedom of expression resulting from the implicit nature of tacit knowledge.

A related problem of the philosophy of science is that of universals [30, chap. 6]. For example, consider the concept of being red. Clearly our knowledge of what is red is not directly related to knowledge of what specific objects are considered to be red; this we conclude because our concept of red obviously is not altered when, for one reason or another, an object we are viewing changes its color. Suppose we can account for tacit knowing in terms of inductive processes that determine how we exploit the leeway in the meaning of words above and beyond our observational evidence. If this can be done, universals can be readily explained as the "end product" of such processes.

Bearing in mind all of these factors, and as an outgrowth of our developed formalism, we now propose two hypotheses as possible avenues for explication of induction.

> (1) An individual seeks that language which maximizes the informativeness of his previous observational experience.

By this we mean that an individual attempts to adjust his language in such a way that it provides him with the most informative possible view of his observations. The information referred to here is to be interpreted exactly as defined in chapter IV. In chapter VI we demonstrated that there may be formal languages which maximize the information in any particular set of observations, and further that these languages are not necessarily highly expressive languages. This result lends credence to hypothesis (1) since it suggests the existence of formal languages which are optimum, in the sense of informativeness, for a given set of observations. In order to be able to state unequivocably that this is the case, however, we would need to develop other concepts which characterized the specific mechanisms bringing about language change. Thus, although we know that for some formal language and for particular observations there are other formal languages—both more expressive and less expressive—which are less informative, we do not know, nor do we suspect, that linguistic change is based purely on expressiveness. Consequently, further investigation is necessary to establish not only that maximally informative languages exist but also that, given their existence, they can be reached by the natural processes of language change.

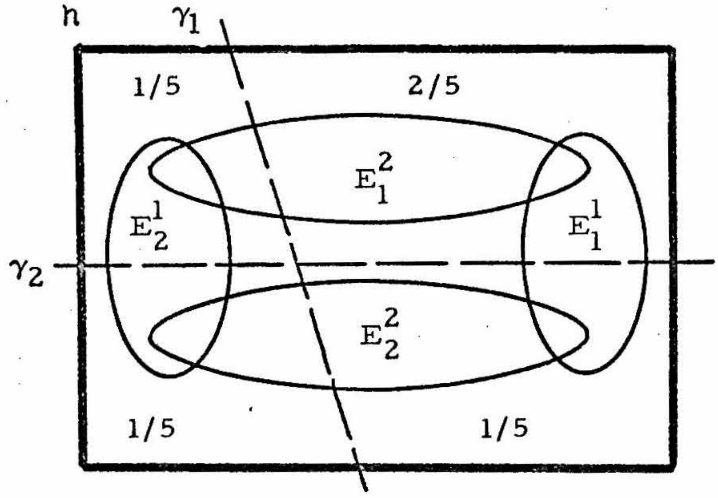
Many of the above points are also relevant to our second hypothesis, which we now state:

> (2) An individual seeks that language which maximizes his expected gain in information based on anticipated observational alternatives.

Although the first hypothesis was clear in terms of previous discussions, this second hypothesis needs further interpretation. Therefore, in order to gain insight into the implications of hypothesis (2), we will make use of the following rather trivial but suggestive example.

Suppose a scientist conceives of two experiments that he might perform to increase his understanding of some domain. The following diagram indicates how these hypothetical experiments, $\mathcal{E}_1$ and $\mathcal{E}_2$, might partition the set of models $\mathfrak{n} \subseteq \mathfrak{m}$. $\mathfrak{n}$ is presumed to be the entire set of models associated with the scientist's previous observational experience. Also shown are two hypotheses, γ_1 and γ_2, of the scientist's formal language.

Each of the two experiments has three possible outcomes, two of which are specifically labeled on the diagram. Thus, for



Case I

example, the outcomes of $\mathcal{E}_1$ are E_1^1, E_2^1, and E_3^1; E_1^1 and E_2^1 are delimited by ellipses on the diagram, whereas E_3^1 consists of all those models not in either of the other two. A similar situation holds for the second experiment $\mathcal{E}_2$. The two sentences γ_1 and γ_2 partition h into four regions, each presumably a state of the scientist's language. The numbers in these regions indicate the *-probabilities that the scientist associates with these states. We have previously discussed the expected informational gain for an experiment and developed a mathematical definition of it in terms of the *-probability. Suppose we now assume that the scientist plans to perform first experiment $\mathcal{E}_1$ and then experiment $\mathcal{E}_2$. The two experiments taken together have nine possible observational outcomes. For example, one of these is given by $E_1^1 \cap E_3^2$ and another by $E_2^1 \cap E_2^2$. In this simple case we can easily compute his anticipated informational gain by using the following formula:

$$\text{(i)}\quad I(\mathcal{E}_1 \text{ and } \mathcal{E}_2) = -\sum_{i=1}^{3}\sum_{j=1}^{3} P^*(E_i^1 \cap E_j^2)\log P^*(E_i^1 \cap E_j^2)\ .$$

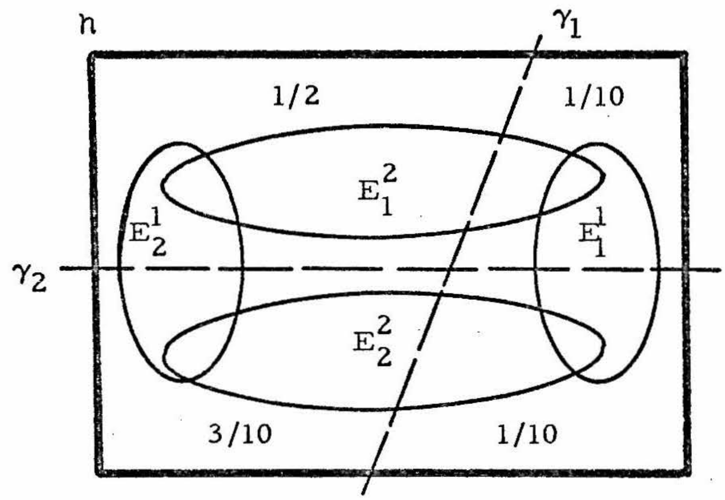
Arbitrarily choosing the natural logarithm we obtain the numerical answer, $I(\mathcal{E}_1 \text{ and } \mathcal{E}_2) = 2.681$. The absolute value of this answer has no particular significance, but we shall be interested in comparing it to several other cases.

Now consider the situation where the scientist has performed experiment $\mathcal{E}_1$ and has found that the outcome lies to the right of the line labeled γ_1, i.e. the observed outcome was E_1^1. He can now do experiment $\mathcal{E}_2$, knowing the result of $\mathcal{E}_1$, and determine which of the remaining admissible states of his language best characterizes the actual circumstances. What is the expected gain in information in this case? By taking the conditional *-probabilities we may compute the expected increase as before. The following formula,

$$\text{(ii)}\quad I(\mathcal{E}_2) \,|\, E_1^1) = -\sum_{i=1}^{3} P^*(E_i^2 \,|\, E_1^1)\log P^*(E_i^2 \,|\, E_1^1)$$

yields the numerical result $I(\mathcal{E}_2 | E_1^1) = 0.637$. As we would surmise the expected gain is less than for the previous case, since the scientist already has knowledge of the outcome of experiment $\mathcal{E}_1$.

Now we shall reproduce the preceding analysis with one change. Inherent in the previous diagram is the assumption that the scientist has certain tacit interpretations of the words of his language. Presumably his interpretations were inductively drawn from a finite number of observations. However, as discussed at length in chapter V, the meanings he has attached to these words may not be able to be fully delineated by any finite set of observations. Consequently, there is a degree of freedom remaining in the tacit meaning he associates with these words (as we have seen for the word "volume"). In the preceding diagram the scientist has exercised that freedom in a certain way. We draw the diagram again as it could appear had he exercised his freedom with regard to tacit meaning in a somewhat different way.



Case II

The salient difference between this diagram and the preceding concerns the hypothesis γ_1. This statement partitions the models in a different manner because of our presumed change in tacit meaning, and accordingly the *-probabilities of the four states are now altered. Using formula (i) we can again compute the expected informational gain prior to performing either experiment. The following result is obtained: $I(\mathcal{E}_1 \text{ and } \mathcal{E}_2) = 2.205$. In a similar fashion we use formula (ii) to calculate the increase in information for experiment $\mathcal{E}_2$, given that the outcome of experiment $\mathcal{E}_1$ is known. Our new result is: $I(\mathcal{E}_2 | E_1^1) = 0.693$.

Notice first that the expected gain in information prior to either experiment is higher in Case I than in Case II. Thus, at this point in time, by exercising his freedom in assigning meaning to the referent words, the scientist would presume that Case I represented the more meaningful position. However, the observed outcome of his first experiment eliminates previously admissible models in such a way as to render the initial (Case I) meanings of the tacit terms less useful informationally, as can be seen from the fact that the expected information for the second experiment is considerably higher in Case II than in Case I. Under these circumstances the scientist could be expected to conclude that a change in his tacitly accepted concepts is called for in light of experiment $\mathcal{E}_1$, which results in his modifying the meanings of the referent words. Let us see how this might work in practice.

Suppose an ornithologist is on a field trip, and observes some bird he has never seen before and says "That is a sparrow". Why does he not say instead that the bird he sees is an owl? It is tempting to say that it looks more like a sparrow to him and so he categorizes it as a sparrow, but that begs the question, it is not a satisfying explanation of the scientist's choice. For one thing, he is differentiating this creature he perceives not only from owls but also from a myriad of other possible entities. In addition, he presumably knows much more about the properties of sparrows than he has actually observed of this particular bird. Thus again we are faced with a situation involving inductive inference. However, our preceding example and the second hypothesis we proposed provide one possible explanation of the scientist's behavior. That is, he says that he sees a sparrow because he anticipates that this is an informing view relative to any subsequent experiments or tests he might carry out. More precisely, although he realizes that further examination might demonstrate that the creature is a stuffed bird or that it might hoot like an owl, the assumption at this point that it is an owl violates his expectations of the bird's behavior based upon prior experience (and a priori probability). The expected gain in informativeness will be maximized if he regards it as a sparrow. Thus, this situation is analogous to the hypothetical example of the two experiments, which was presented to aid in interpreting hypothesis (2). As we pointed out there, the expected gain in information for a given experiment may be different for distinct formal languages. In this particular case we would conclude that for the scientist to say that he sees a sparrow rather than an owl is more informative to him, based on his experience and the observational alternatives he can envision. Note that some other individual might well be best informed by calling the creature in question a bird, or a wounded sparrow, or perhaps simply an animal, depending on his orientation and anticipated activities.

Hypothesis (2) states that the choice among these is determined by a combination of the a priori probability, the sets of models to which he is confined by prior observations, the formal language, and the experiments or tests which the observer considers to be relevant.

In the discussion of the second hypothesis we have primarily considered changes in language resulting from a change in the meaning of referent words. We do not mean to suggest that all legitimate changes are of this nature. Examples of this type were used for convenience only, and in reality we would expect changes in semantic categories, semantic transformations or other aspects of the formal language. Also, we emphasize that hypothesis (2) does not fully explicate language change, just as hypothesis (1) did not. Again, we have not accounted for the actual mechanisms of language change, and thus we do not know that maximally informing languages, in the sense of hypothesis (2), can always be arrived at in a natural manner.

We now wish to re-examine the two remaining examples in the introduction in light of the two hypotheses we have proposed. As the first of these, consider the re-definition of the term "isotope". Prior to the discovery of deuterium, isotopes of the same element were held to be chemically inseparable, but deuterium was chemically separable from light hydrogen. Clearly, once deuterium was discovered, it was either to be called an isotope or it was not. What factors aided in resolving this question? As we know, the term "isotope" was tacitly re-defined to include things like deuterium, and of course this change caused corresponding modifications of the associated formal languages. If we take the viewpoint of hypothesis (1), we would say that the change in meaning of the word "isotope" resulted from adopting a more informative position with respect to the previously observed properties of deuterium. That is, the scientists in question found that classification of deuterium as an isotope was more informing than considering it to be in a category of its own. Thus deuterium was observed to be sufficiently much like other isotopes so that a change in the concept of isotopy could be accommodated without conflicting with knowledge of their properties.

The outlook on this problem afforded by hypothesis (2) is slightly different. The second hypothesis suggests that to regard deuterium as an isotope, thereby modifying the notion of isotopy, was more informative with respect to its potentially testable—but as yet unverified—properties. We might say that the essential difference between the two hypotheses is that the first relies most heavily on previous experience while the second combines this with anticipatory factors. This "looking ahead" of the second hypothesis is based on the particular tests and experiments which seem plausible and relevant to the theories to be investigated. We should mention that Polanyi's [26, p. 111] analysis of this example states that "isotope" was re-defined to reflect its "truer meaning". It is difficult to know what basis one could have for such an absolute judgement. If one were to construe the word "true" in the sense of better fit with observation as in "the carpenter trues up the wall of the house", then Polanyi's comment is directly related to the view presented here. We would emphasize that our hypotheses do not depend upon an absolute truth. Rather, they rely on the individual's conception of what appears most informative to him in the light of previous experience and his metaphysical assumptions. This admits the possibility that his view is actually false of reality, i.e. some of his assumptions are false of the true model M_0.

With regard to the second example, which concerns the explicit and implicit scope of a scientific paper, we can suggest one possible resolution of the apparent disparity. The salient point here is that the scientist finds it most informative to employ one formal language in writing a paper for a journal and another in guiding his personal research. Thus, suppose that he is a biologist studying the functions of vision in the wolf spider. For purposes of his own research and as a means of determining appropriate experiments to perform, he is likely to employ a formal language which facilitates generalizations of the specific experimental outcomes he perceives. That is, he may ignore certain known idiosyncrasies of his experimental apparatus and the particular spiders he is using as subjects on the assumption that these distinctions will eventually be shown to be irrelevant. In the sense of our second hypothesis he will find it most informing—based on experiments he expects to perform—to make use of a formal language whose states coincide well with the scientist's long term expectations regarding alternative, and as yet unproven, theories about the general phenomena he is studying. This may, for example, result in his ignoring the peculiarities of the wolf spider since he anticipates that his conclusions will eventually prove valid for a much broader class of animals.

On the other hand, when the scientist is writing a paper about his results, different considerations may motivate his choice of formal language. For one thing, in contrast with his personal speculations, he is now very definitely concerned with making statements which are "true", or hypotheses which are almost certainly correct. But what he actually knows to be true falls far short of what he expects will later be proven true. Consequently, he will take great care to ensure that he fully describes the specific and detailed conditions of his experiment and that he does not over-generalize the results he has obtained. Thus, in this situation he is confronted with the problem of communicating his results to some community of scientists with similar backgrounds, without making the paper either unnecessarily detailed and trivial or overly vague and speculative. Since he generally presumes, although perhaps incorrectly, that his readers have a common understanding of many of the concepts involved in his paper, the scientist can attempt to employ some formal language which will be most informative to them. Thus, even though he does not explicitly state many of the generalizations of his results which he personally suspects to be true, he can tailor the language he uses to allow others to inductively infer what are to them the most informative consequences of his results. In this sense, hypothesis (1) would seem to provide the most reasonable explication of his behavior. That is, he seeks a language which maximizes the information relative to his specific observational experience. He then can assume—based on the presumed similarities of his readers' languages—that they will augment his statements with hypotheses and generalizations analogous to those he feels will ultimately be proven correct. In this respect, hypothesis (2) seems to be most applicable.

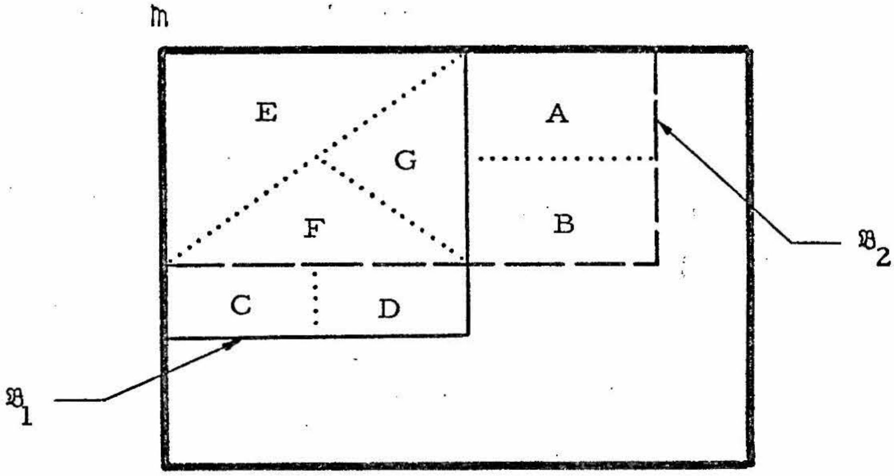
Neither of the examples presented in this chapter as characteristic of the inductive process could be considered to be solved by our hypotheses. Nevertheless, it does appear that they provide a new and useful insight into these difficult problems. As we have mentioned, the primary missing ingredient is our current inability to characterize the actual mechanisms of language change. We have, however, suggested informativeness as the force underlying such changes, and either of our two hypotheses, or perhaps some combination of them, provide possible guidelines for an investigation of the dynamics of formal languages.

Let us summarize what we have accomplished in this thesis. A scientist tries to find structure in the world around him. He does this by formulating and testing scientific theories. At one time it was thought that this could be done by the slow accretion of scientific knowledge, by a systematic procedure often referred to as the scientific method. We now know that this is not the way science develops, and that there are many difficult and unsolved problems in explaining science and the scientific method. In this thesis we have developed a formal method for attacking these problems. The concept of a formal language has been introduced to explicate the notion of structure. A clear distinction has been made between language, or conceptualization, on the one hand and observation on the other hand. We have made clear the necessity of tacit knowledge which could not be fully delimited by any finite amount of observational evidence. The existence of this tacit knowledge provides the leeway within which a scientist can shift his language to fit his ever increasing experimental evidence. By introducing the notion of probability we have developed a related measure of information as a function of formal language, as well as of observation. Finally we have suggested that either information or expected information, or some combination of these, can be used as a forcing function to define the concept of an optimal formal language for a given body of observational evidence. Thus we suggest that the process of science is not one of discovery of structure, so much as it is a fitting of structure to experimental evidence; and more importantly, we have provided a formal apparatus for making this suggestion precise.

We have attempted to study these problems inherent in the scientific method from the point of view of the meanings of sentences, without a correspondingly detailed analysis of the meanings of individual words. Such analysis we believe is a natural extension of what we have done and will certainly be necessary in conducting further investigations of induction and other related problems of the methodology of science. It has also become apparent that the temporal processes of language change, as well as the static informational aspects of language, are involved in the development of science—particularly, as we have seen, in the problem of induction.

APPENDIX

In this appendix, we prove the theorem stated in chapter VI which asserts the independence of the various quantities affecting the informativeness of an observation as the language changes from less expressive to more expressive. For convenience, we repeat the following diagram.



$\mathfrak{B}_1$ is the basis for language $\mathcal{L}_1$ and $\mathfrak{B}_2$ the basis for $\mathcal{L}_2$. Also, $\mathcal{L}_2$ is assumed to be at least as expressive as $\mathcal{L}_1$, i.e. $\mathcal{L}_2 \geq \mathcal{L}_1$. The sets A, B, C, D, E, F, and G may be characterized as follows:

A: $h/_2\, A$ B: $h/\!/_2\, B$

C: $h/_1\, C$ D: $h/\!/_1\, D$

E: $h/_{1,2}\, E$ F: $h/\!/_{1,2}\, F$

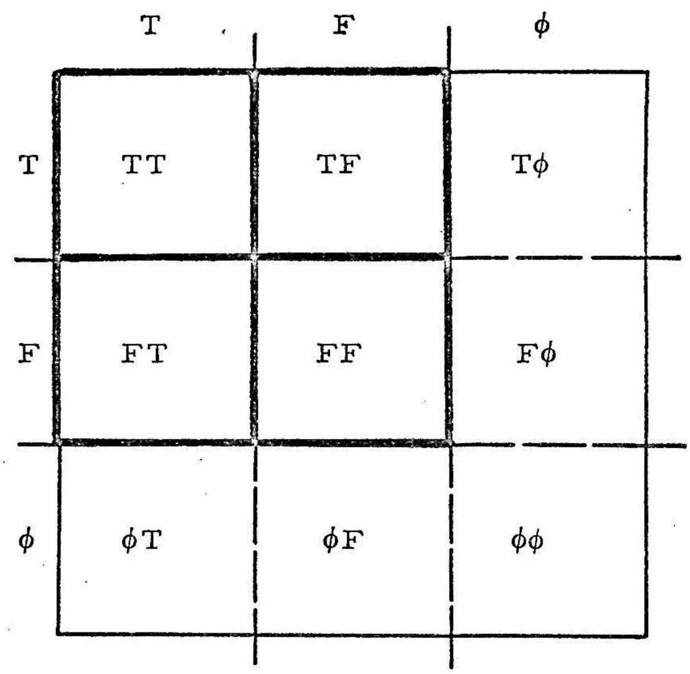
G: $h/\!/_1\, G$ and $h/_2\, G$

The set of models η is assumed to correspond to some arbitrary observation. Since E and F have the same properties with respect to η in both languages, they are not of interest. The sets A, B, C, D, and G are the ones we wish to prove independent of one another.

Theorem: There exist formal languages $\mathcal{L}_1$ and $\mathcal{L}_2$, $\mathcal{L}_2 \geq \mathcal{L}_1$ such that there are sets of models η (corresponding to possible observations) for which the sets A, B, C, D, and G may be empty or non-empty in any combination, i.e. A, B, C, D, and G are independent of one another.

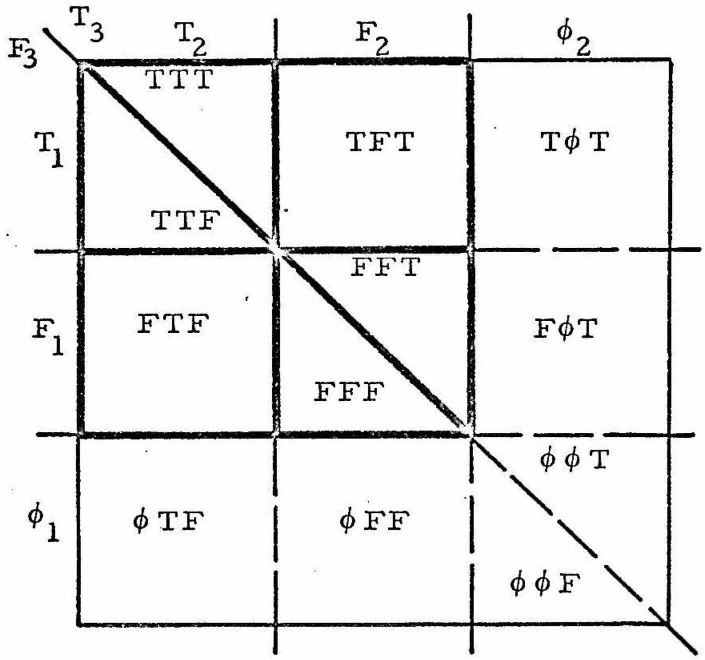
Proof: We shall exhibit a language $\mathcal{L}_1$ and four languages corresponding to $\mathcal{L}_2$ (this is the minimum possible number), and for each of the $2^5 = 32$ cases where some combination of A, B, C, D, and G could be non-empty, we will choose an η to show that this combination is actually realizable. Each of the languages we shall illustrate has a unique basis, as indicated by the elements with the heavy borders. The first figure shows the state diagram for language $\mathcal{L}_1$. The next four figures show the possible cases for language $\mathcal{L}_2$, and in each case $\mathcal{L}_2 \geq \mathcal{L}_1$. Following each of these figures is a table designating some set of models η and, for each η, the status of the sets A, B, C, D, and G. In these tables, a "1" indicates the corresponding set is non-empty and a "0" that it is empty.

If a table entry lists more than one element, where these are separated by commas, then h is the set of models obtained by taking the union of all elements listed. The characteristics of the set h with respect to the basis for $\mathcal{L}_1$ and for the particular case of $\mathcal{L}_2$ determine whether A, B, C, D, and G are empty or non-empty.



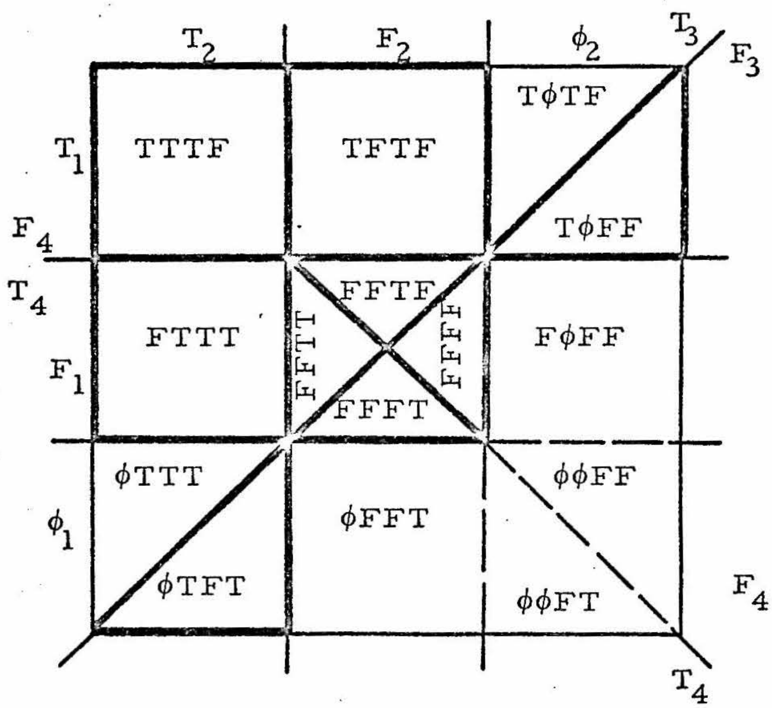
The above figure shows the state diagram for language $\mathcal{L}_1$. For this and each of the subsequent figures, the truth values are subscripted to indicate the sentence to which they correspond. Thus, for example, F_2 designates the region where sentence γ_2 is false.

Case I. A = B = C = D = 0. Suppose $\mathcal{L}_2$ is —



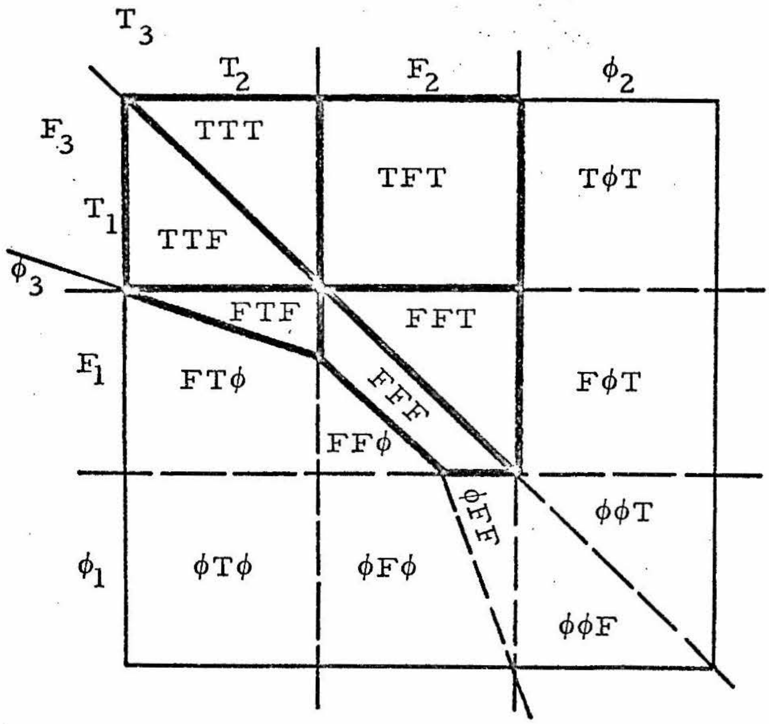
h	A	B	C	D	G
FTF	0	0	0	0	0
TTF	0	0	0	0	1

Case II. C = D = 0. Suppose $\mathcal{L}_2$ is —



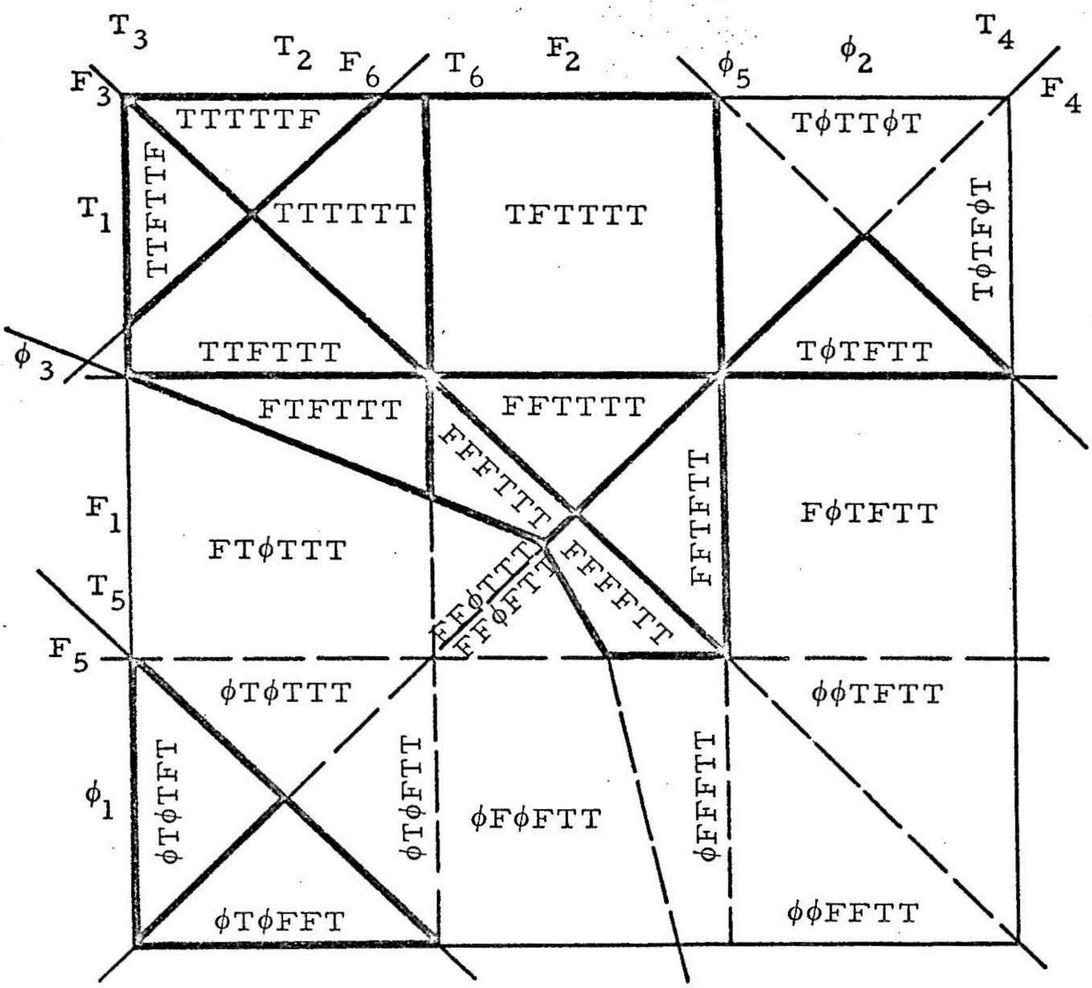
h	A	B	C	D	G
ØTTT, ØTFT, TØFF, TØTF	0	1	0	0	0
ØTFT, TØFF	0	1	0	0	1
FTTT	1	0	0	0	0
FFFT	1	0	0	0	1
TØFF, TØTF	1	1	0	0	0
TØFF	1	1	0	0	1

Case III. A = B = 0. Suppose $\mathcal{L}_2$ is —



h	A	B	C	D	G
FTφ, FFφ	0	0	0	1	0
FTφ, FFφ, FFF	0	0	0	1	1
TFT	0	0	1	0	0
TTT	0	0	1	0	1
FTφ	0	0	1	1	0
FFφ, FFF	0	0	1	1	1

Case IV. Suppose $\mathcal{L}_2$ is —



h	A	B	C	D	G
TØTFTT, ØTØFFT, ØTØTFT, TTFTTF, TTTTTF	0	1	0	1	0
TØTTØT, TØTFØT, FØTFTT	0	1	0	1	1
TØTTØT, TØTFØT, TTFTTF, TTTFTT, TTTTTF	0	1	1	0	0

Case IV (continued)

h	A	B	C	D	G
TØTTØT, TØTFØT	0	1	1	0	1
ØTØFFT, ØTØTFT, ØTØFTT, ØTØTTT, FTØTTT, FTFTTT, TTFTTT, TTTTTT, TTFTTF, TTTTTF	0	1	1	1	0
ØTØFFT, ØTØTFT, ØTØTTT, ØTØFTT, FTØTTT, FTFTTT, TTFTTT, TTTTTT	0	1	1	1	1
FTØTTT, FFØFTT, FFØTTT	1	0	0	1	0
FTFTTT, FFTTTT, FFTFTT	1	0	0	1	1
TFTTTT	1	0	1	0	0
TTTTTT	1	0	1	0	1
FFØFTT, FFTTTT	1	0	1	1	0
FFØFTT	1	0	1	1	1
ØTØFTT, FFØTTT, FFTTTT, TTFTTF, TTTTTF	1	1	0	1	0
ØTØFTT, FFØTTT	1	1	0	1	1
TØTTTT, TTFTTT TTTTTF, TTFTTF	1	1	1	0	0
TØTTTT	1	1	1	0	1
ØTØTTT, ØTØFTT, TTTTTF, TTFTTF	1	1	1	1	0
ØFØFTT	1	1	1	1	1

BIBLIOGRAPHY

1. Beach, F. A.: "The Snark Was a Boojum"; The American Psychologist, volume 5, 1950, pp. 115-124.

2. Bruner, J. S., Goodnow, J. J. and Austin, G. A.: A Study of Thinking. J. Wiley and Sons: New York; 1956.

3. Carnap, R.: Logical Foundations of Probability. University of Chicago Press: Chicago, Ill.; 1950.

4. Carnap, R.: Meaning and Necessity. University of Chicago Press: Chicago, Ill.; 1956.

5. Chomsky, N.: Syntactic Structures. Mouton and Company: The Hague; 1957.

6. Chomsky, N.: "Formal Properties of Grammars"; chapter 12 in Handbook of Mathematical Psychology, edited by Luce, Bush, and Galanter, volume 2. J. Wiley and Sons: New York; 1963, pp. 323-418.

7. Chomsky, N.: Current Issues in Linguistic Theory. Mouton and Company: The Hague; 1964.

8. Chomsky, N.: Aspects of the Theory of Syntax. M. I. T. Press: Cambridge, Mass.; 1965.

9. Chomsky, N. and Miller, G. A.: "Introduction to the Formal Analysis of Natural Languages"; chapter 11 in Handbook of Mathematical Psychology, edited by Luce, Bush, Galanter. J. Wiley and Sons: New York; 1963.

10. Church, A.: "An Unsolvable Problem of Elementary Number Theory"; in The Undecidable, edited by M. Davis. Raven Press: Hewlett, New York; 1965, pp. 88-107.

11. Cohen, P. J.: Set Theory and the Continuum Hypothesis. W. A. Benjamin, Inc.: New York; 1966.

12. Davis, M.: Computability and Unsolvability. McGraw-Hill: New York; 1958.

13. DuBridge, L. A.: The President's Report, 1967/1968. California Institute of Technology: Pasadena; November, 1968.

14. Frege, G.: "On Sense and Nominatum"; Readings in Philosophical Analysis, edited by Feigl and Sellars. New York; 1949.

15. Ginsburg, S.: The Mathematical Theory of Context-Free Languages. McGraw-Hill: New York; 1966.

16. Gödel, K.: The Consistency of the Continuum Hypothesis. Princeton University Press: Princeton, N. J.; 1940.

17. Gordon, C. K., Jr.: "The Third Great Crisis in Mathematics"; IEEE Spectrum, May, 1968, pp. 47-52.

18. Harris, Z. S.: "Transformational Theory"; Language, July-September 1965, volume 41 # 3, pp. 363-401.

19. Khinchin, A. I.: Mathematical Foundations of Information Theory, translated by R. A. Silverman and M. D. Friedman. Dover Publications: New York; 1957.

20. Kleene, S. C.: Introduction to Metamathematics. D. Van Nostrand Company: Princeton, N. J.; 1950.

21. Kreisel, G. and Krivine, J. L.: Elements of Mathematical Logic. North-Holland Publishing Co.: Amsterdam; 1967.

22. Kuhn, T. S.: The Structure of Scientific Revolutions. University of Chicago Press: Chicago, Ill.; 1962.

23. Lyndon, R. C.: Notes on Logic. Van Nostrand Co.: Princeton, N. J.; 1966.

24. McCann, G. D. and Dill, J. C.: "Fundamental Properties of Intensity, Form, and Motion Perception in the Visual Systems of Calliphora phaenicia and Musca domestica"; Journal of General Physiology, 1969, volume 53, pp. 385-413.

25. Nagel, E.: The Structure of Science. Harcourt, Brace, and World: New York; 1961.

26. Polanyi, M.: Personal Knowledge. University of Chicago Press: Chicago, Ill.; 1958.

27. Popper, K. R.: The Logic of Scientific Discovery. Harper and Row: New York; 1959.

28. Post, E.: "Recursive Unsolvability of a Problem of Thue"; The Undecidable, edited by Davis. Raven Press: Hewlett, New York; 1965, pp. 292-303.

29. Quine, W. V.: Mathematical Logic. Harvard University Press: Cambridge, Mass.; 1951.

30. Quine, W. V.: From a Logical Point of View. Harper and Row: New York; 1963.

31. Robinson, A.: Introduction to Model Theory and to the Metamathematics of Algebra. North-Holland Publishing Co.: Amsterdam; 1963.

32. Russell, B.: "On Denoting"; Mind, 1905, volume 14, pp. 479-493.

33. Russell, B.: Mysticism and Logic. W. W. Norton and Company: New York; 1929.

34. Russell, B.: Inquiry into Meaning and Truth. W. W. Norton and Company: New York; 1940.

35. Savage, L. J.: The Foundations of Statistics. J. Wiley and Sons: New York; 1954.

36. Schwartz, L. S.: Principles of Coding, Filtering, and Information Theory. Spartan Books: Baltimore; 1963.

37. Shannon, C. E. and Weaver, W.: The Mathematical Theory of Communication. University of Illinois Press: Urbana, Ill.; 1964.

38. Sinsheimer, R. L.: "The End of the Beginning"; Bulletin of the California Institute of Technology, March, 1967, volume 76 # 1, pp. 1-8.

39. Tarski, A.: Logic, Semantics, Metamathematics. Oxford University Press: London; 1956.

40. Tarski, A. and Banach, S.: "Sur la Décomposition des Ensembles de Points en Partie Respectivement Congruentes"; Fundamenta Mathematica, 1924, volume 6, pp. 244-277.

41. Thompson, F. B.: The Semantic Interface in Man-Machine Communication. Technical Report RM 63TMP-35, September, 1963. TEMPO, General Electric Co., Santa Barbara, Calif.

42. Thompson, F. B.: "English for the Computer"; AFIPS Conference Proceedings, 1966 Fall Joint Computer Conference, volume 29, pp. 349-356. Spartan Books: Washington, D.C.; 1966.

43. Turing, A. M.: "On Computable Numbers, with an Application to the Entscheidungsproblem"; in The Undecidable, edited by Davis. Raven Press: Hewlett, New York; 1965, pp. 115-154.

44. Whorf, B. L.: Language, Thought, and Reality. M. I. T. Press: Cambridge, Mass.; 1965.

45. Wirth, N. and Weber, H.: "EULER; A Generalization of ALGOL and Its Formal Definition---Part I"; Communications of the ACM, January, 1966, volume 9 # 1, pp. 13-23.

46. Woodger, J. H.: The Axiomatic Method in Biology. University Press: Cambridge, England; 1937.