Chapter 1 Hypothesis discovery

1. Introduction

This thesis is concerned with forming hypotheses by generalisation. We hope to expose a more interesting structure than would be expected from contemplation of the traditional recipe for generalisation which is the simple replacement of constants by variables. Thus we are committed from the start to a study of generalisation of sentences in the predicate calculus. First-order predicate calculus will be quite sufficient although some other formalisms will be mentioned. We assume, therefore, that the reader is acquainted with some standard formulation of first-order predicate calculus, such as that of Schoenfield (1967).

Some examination of generalisation processes in the predicate calculus may be found in Meltzer (1970a). He constructed a program to find general laws of group theory from particular examples. Some other work in the domain of predicate calculus has been done by Popplestone (1970), although this was not concerned with generalisation alone, but contained other rules for generating hypotheses, a heuristic search through the hypothesis space and a mechanism for generating "test" experiments to decide between competing hypotheses.

The largest single body of work in A.I. research on hypothesis formation has been on guessing grammars. A method for guessing finite-state grammars was devised by Chomsky and Miller (1957) and generalised to context-free grammars by Solomonoff. The adequacy of Solomonoff's (1964) method has been challenged by Shamir and Bar-Hillel (Shamir, 1962).

The first heuristic program for guessing finite-state grammars seems to be that of Feldman (1967).

Theoretically, we find contributions by Gold (1967), Feldman (1970), Feldman, Gips, Horning and Reder (1969) and Horning (1969).

Gold showed what kind of behaviour could theoretically be expected from grammar-guessing machines. Feldman (1970) continued these studies. In Feldman, Gips and Horning we find both theoretical results which are less general but stronger than those in Gold, and descriptions of a practical program for inferring pivot grammars, which form a subclass of the context-free grammars and which properly contain the finite-state grammars. Horning, too, conducts both a theoretical and practical investigation, based on Bayesian ideas.

His program has the special distinction of being, in one sense, theoretically optimal. Seemingly, however, it has less practical ability than the other, heuristic, programs.

Closely related work on guessing finite-state machines can be found in Perryman (1970) and Feldman and Biermann (1970). Taking this work together with that of the grammar guessers, we may conclude that only a little need yet be done to obtain a theoretically and practically good algorithm for guessing a finite-state grammar from a set of examples and a set of non-examples. Matters remain unsatisfactory, however, in the case of context-free grammars.

Amarel (1962, 1971) is concerned with guessing programs of a rather simple type with no looping, from samples of input-output pairs. He seems not to have programmed his method. Hewitt (1968) tries to guess programs with, possibly, recursion, but using as data traces of the program he is trying to guess. This method seems not to have been programmed.

Given descriptions of both examples and non-examples of a class of pictures, Winston (1970) attempts to generate a general description of the entire class. One interesting feature of his program is that new examples or non-examples may be taken into account by alteration of the current general description. The program is unique of its kind and seems quite successful.

This leads into the field of pattern recognition where we may, for example, regard the perceptron convergence theorem (Nilsson 1965, Minsky and Papert, 1969) as demonstrating the successful operation of a hypothesis-guessing machine.

Standing on its own is the work of Buchanan, Sutherland and Feigenbaum (1969, 1970) and Feigenbaum, Buchanan and Lederberg (1971) on hypothesis-formation in organic chemistry. The programs they developed form the most impressive hypotheses of any, although there are correspondingly strong assumptions.

We should mention also some work in psychology, in, essentially, a very restricted portion of the first-order predicate calculus. This concerns concept-learning. Notable books are those of Bruner, Goodnow and Austin (1956) and Hunt, Marin and Stone (1966).

Work on the related subject of analogy has been done by Evans (1968), Kling (1971) and Becker (1970). Evans was in fact, largely concerned with generalisation, although his work concerned analogy questions in I.Q. tests. The solutions generated by his program were in almost total agreement with those of the proposers of the tests. Kling was interested in the use of analogy to help in proving theorems analogous to already proven ones. Kling has programmed his method. Becker proposed his notions of analogy as an essential component in a model of "intermediate level cognition". He has not programmed his method.

Of course one could go on for ever quoting work on hypothesis discovery. However the above gives a good indication of what has been done in the way of actually proposing and implementing algorithms.

While we will both propose and implement discovery algorithms, we will be mainly concerned with a theoretical analysis of the relation of generalisation and its use. It is hoped that this will provide some useful notation and techniques for further development of algorithms. The analysis sets up criteria derived from the philosophy of science. The possibility of doing this was explicitly stated by Buchanan (1966), to whom we owe some considerable debt. The hypothesis formation problem is seen as a stage in a continuing process of theory formation and criticism, data gathering, prediction and so on. Criteria that a hypothesis formation method should satisfy can then be set up. This forms the main part of this chapter. The criteria depend on what notion of explanation of the given data is employed. In chapter 2 we develop a definition of generalisation and so specialise the criteria that they result in a tractable formal problem of finding a machine which will produce hypotheses satisfying the criteria. In fact, a "nicest" such hypothesis will be required.

In chapter 3 the abstract theory of the generalisation relationship is developed and employed to examine the formal problem in chapter 4 and to provide several illustrations and other applications in chapter 5.

Finally, in order to partially correct the distorted emphasis on a single, idealised stage of hypothesis formation, we give in chapter 6 a generalisation of Feldman's theory of hypothesis identification in the limit, (Feldman, 1970).

The work started with a suggestion by R.J. Popplestone (private communication) that, just as the unification algorithm was fundamental to deduction, so might a converse be of use in induction. Unification is a basic idea in the theory of resolution given by Robinson (1965). We refer to his work for a complete description of the notation used in that theory.

For our immediate purposes, it is only important to note that a literal is an atomic formula or the negation of one and that a clause is a set of literals and it abbreviates the disjunction of its members (taken in some standard order). The letters L, M and N are used to stand for literals. The letters C, D and E are used to stand for clauses.

A literal is a <u>unification</u> of two literals iff it is an instance of each. A literal is a <u>most general</u> unification of two literals iff any other unification of them is an instance of it.

Similarly, a literal is a <u>generalisation</u> of two literals iff each of them is an instance of it. A literal is a <u>least general</u> generalisation of two literals iff it is an instance of any other generalisation of them.

For example, a most general unification of P(x,x) and P(f(y),f(g(z))) is P(f(g(z)),f(g(z))). A least general generalisation of them is P(x,y).

The existence of least general generalisations was soon shown. In fact they are easier to obtain than unifications. A necessary and sufficient condition that two literals have a least general generalisation is just that they have the same predicate letter and sign. However this is not enough even to generate simple universal laws of the form: $\forall x(P(x) \rightarrow Q(x))$. To do this it is necessary to consider generalisations of clauses. Let us say, again for the moment, that a clause C is more general than a clause D if C subsumes D - that is if there is a substitution σ such that $C\sigma \subseteq D$. (In resolution theory, substitutions are functions which operate on expressions; they are denoted by Greek letters.) One can then define the least general generalisation of two clauses. The key theorem is that any two clauses have a least general generalisation.

It is interesting to note that the similarity between deduction and induction breaks down here. What is useful is not a concept of unification of two clauses, but the deduction principle called resolution.

We then envisaged crude algorithms which would build up from clauses abbreviating implications between ground literals the set of least general generalisations and pick some subset as the proposed hypothesis. (A ground literal is one that contains no occurrences of variables.) Such a subset would have to be consistent with the data, but this does not rule out enough combinations and it is necessary to impose a condition that only the "nicest" (according, say, to some measure of simplicity) combination be picked.

Encountering Buchanan's work (1966) we realised that a hypothesis formation (or suggestion or discovery) method should be placed within a philosophical framework so that, for example, the suggested method could be criticised as not meeting various criteria found in the literature. Thus we arrive in almost the reverse order at the beginning of the development of this thesis. The main technical addition to the concept of a least general generalisation is that of generalisation relative to a body of knowledge. This enables, for example, a robot to form generalisations about its sensory experience, given in terms of light patterns on a retinal grid, in the more abstract language of objects such as bananas, faces, spectacles and so on.

2. Criteria for hypothesis formation

We begin with some general arguments that induction may be useful in A.I. robot research and, if so, such criteria as may be provided by the philosophy of science should be accepted. Of course hypothesis formation, being one might say the intellectual activity par excellence requires no justification for its study for its own sake. As, however, we believe that integrated robot systems are of some importance in A.I., arguments relating such systems with inductive ones should be considered.

One can easily see in general terms how inductive abilities would be of help to a robot, i.e. an artificial rational man. Such an ideal entity should be a scientist - and so, a non-deductive reasoner of some kind, depending on one's philosophy of science. Again, a robot should have common sense and be able to talk a common-sensical language, such as English. It is a very defensible thesis that both English and common-sense involve a naive science. Both learning and using this naive science will therefore involve the robot in some naive non-deductive reasoning.

There is, even in present robots, an implicit form of inductive ability. For example, all present robots base their plans of action on rather brief glimpses of the world, since picture processing using available techniques is slow. This reflects an inductive expectation that the world will not change too much between looks. It is however quite unclear what inductive assumptions are definitely already built-in, let alone which ones ought to be.

A robot should have the explicit ability to learn predicates ostensively. Several presentations of a lamp standard before the robot's eye should result in his forming a general idea of lamp standards, (Barrow and Popplestone, 1971, Winston, 1970). He should also be able to perform inductions from activities or events and learn causal connections (Hayes, 1971).

Fancifully, by learning plausible beliefs about the effects of actions, but being prepared, if necessary to look or account for exceptions, it may be possible to ameliorate what McCarthy calls the "frame" problem (McCarthy and Hayes 1969, Hayes, 1971).

Most of the above examples show that some kind of non-demonstrative reasoning is required. It has not been shown that the most suitable form is that of hypothesis formation. Such a demonstration would require a much fuller specification of the robot's mental life than has been given. We do not have an integrated sketch of his ontology and epistemology and theory of perception and the structure of his knowledge and his methods of plan formation and his motivations and his physical being and so on.

Nonetheless, hypothesis formation is a leading candidate; analogical reasoning is the only other contender so far. It seems, therefore, worthwhile to develop hypothesis formation itself as much as possible.

We are concerned then with the automatic formation of hypotheses. At the most grandiose level, one would want a machine which could try to solve the problems of contemporary physics. At the lowest, one would wish to be able to (non-deductively) infer 'All crows are black' from 'That crow is black' and 'This crow is black'. In order to make quite clear the low level at which present general programs operate (without, however, any intention of denigrating the constructers of these systems) here are some examples:

a) Feldman, Gips, Horning and Reder (1969) have studied the induction of grammars (mainly context-free) from finite sets of samples of legal strings, and, perhaps, a set of illegal strings.

On being told that {AABB, AB, AAABBB} is a set of examples, a program of theirs, GRIN2, produced the following grammar:

X := AY

Y := XB/B.

On being told that {C, ACB, AACBB, AAACBBB} was a further set of examples another program GRIN2A, produced the following grammar:

X := AY/C

Y := XB/B.

b) Meltzer (1970) has looked at the problem of determining the axioms for a class of interpretations, given a finite number of facts about each of a finite class of structures.

He fed in a representation of the following facts about the cyclic groups of orders 2 and 4, whose domains are $\{e,a\}$ and $\{e,b,c,d\}$ respectively.

$e \circ e = a$; $a \circ e = a$; $(a \circ a) \circ e = a \circ (a \circ e)$; $(e \circ a)a = e$; $e \circ a \neq e$;
$a \circ a \neq a$; $b \circ c = c \circ b$; $(b \circ b) \circ b = c$; $(b \circ b) \circ c \neq c$; $(b \circ c) \circ c \neq b$.

The program generalised these representations, and obtained a representation of the following induced axioms.

$x \circ e = x$;
$(x \circ x) \circ y = w$ implies $x \circ (x \circ y) = w$;
$(y \circ a) \circ a = y$;
$x \circ y = y \circ x$;
$b \circ (b \circ b) = y$.

At their best, these and similar general programs such as our own are only slightly more magnificent. There are much more specific programs which generate more impressive hypotheses, of which a prime example is the Heuristic Dendral program of Buchanan, Sutherland and Feigenbaum (1969, 1970). This program's ability is almost entirely due to the availability of a large amount of chemical knowledge (although this knowledge was by no means there for the taking).

We will not attempt any general theory of theory generation. What is possible, in general, is to set up an outline which any generation method must fill in. This outline is derived from the philosophy of science. Apart from any use to which we actually put this work, we are convinced that much can be gained from its study. It provides discussion of the justification and criticism of hypotheses; of problems of explanation and the nature of simplicity. One can find categorizations of different types of hypotheses and various qualities of hypotheses. There are accounts of the actual and ideal progress of science, and accounts of the dynamics of theory construction. There is the problem of the nature of scientific language and its relation to the world; it is important to know the 'behind the scenes' assumptions made automatically when one chooses to use a particular theoretical language.

However, philosophy does not deal directly with our main concern, which is the generation of hypotheses. Indeed the philosophers generally delegate this problem to the psychologists. We therefore avoid a great deal of philosophical discussion since, for the most part, we have only extricated a schema of philosophical questions, rather than answers. So we expect that most of the current philosophical approaches could be so adapted as to fit in with our schema. This is not to say that we regard the different approaches as being essentially the same. The fact is that our schema is largely incomplete and typically (especially as regards the problem of justification of hypotheses) leaves unanswered just those questions at which the bones of contention arise. Neither have we succeeded in altogether avoiding philosophical commitment. We have tended to raise questions and problems in a way which is rather more acceptable to the Carnapians than the Popperians. Sometimes we use terminology in a way that is unacceptable to either. For example we discuss the problem of justification. A Carnapian would prefer the problem of confirmation, saying that no absolute justification is possible, in general. A Popperian would also dislike the problem of justification. He would say that the problem is how to decide between competing hypotheses. There is a problem of criticism. A Popperian might say that there is no problem of justification but perhaps one of corroboration. Nonetheless the debate does hinge on the existence and the importance of the problem of justification and the other related ones. We will therefore use the word justification to introduce this whole nexus of problems.

To rephrase our concerns, we are looking for a language to uniformly describe the various attempts made in A.I. to construct programs for hypothesis formation. It is believed that this will lead to a general theory in time to come, and that in the immediate future, when one wishes to program a method of theory construction, one will have available an interesting set of non-trivial questions about one's method.

Hypothesis formation is regarded as a process containing stages of theory construction, theory criticism, data gathering and data analysis. There will also be some supervisory mechanism to decide the order of these stages. As mentioned above we do not know and will not consider, how this fits into the general mental character of an ideal rational man. In fact, for the most part, we will only consider a simplified snapshot of the process, consisting of a stage of theory construction followed by one of theory criticism. The stage of theory criticism will be dealt with, on the whole, by reference to various possible philosophical positions.

There is no virtue in our omissions; they are blanks which should be filled. It will surely happen that when the process is considered as a whole many missing parameters will be discovered. The analysis should therefore be considered entirely provisional.

Suppose then that our robot is about to formulate and then criticise a theory after some stages of data gathering and analysis. The robot will possess knowledge and beliefs and pragmatic attitudes to these elements of knowledge and belief. Let us call all this k. He will have before him some body of phenomena, f, together with the relevant circumstances of their occurrence, e, for which it is required to find some kind of explanatory hypothesis, h. Generally f will be a set of phenomena, $\{f_i \| i=1,n\}$ and e will be a set of relevant attendant circumstances $\{e_i | i=1,n\}$ given by a function **Ev**, that is $e_i = \mathbf{Ev}(f_i)$. How the e_i are selected depends on an implicit stage of data gathering. The more naive this stage is, the more irrelevant information e_i will contain. The lack of any theory of good data gathering is certainly one of the most important omissions.

In general, there will be many explanatory hypotheses and so it will be necessary to choose the nicest. We expect that there will be some measure, $\precsim$, of niceness. By $h \precsim h'$, we mean that h is at least as nice as h'; $\precsim$ will be transitive and reflexive. This niceness ordering will be defined relative to f and e and, perhaps, k.

Perhaps the most famous example is provided by Newton's theory of gravitation. Here f and e were both large and varied. Among the phenomena was the fact that an apple fell on Newton's head. The relevant circumstances were Newton's situation relative to the apple at the start of its flight and the apple's position in the earth's gravitational field.

The hypothesis h consisted of Newton's theory of gravitation. The knowledge, k, consisted of the axioms of Euclidean geometry and some axioms for time together with the interpretation of both of these via some theory of measurement. Of course, Newton himself did not formulate matters in these terms.

The coincidence of Newton's head with the apple is explained in two stages. First h and k together with the statement regarding the apple's position in the Earth's gravitational field imply that the apple will fall. This and the fact that Newton's head was directly below the apple imply that the apple will strike him on the head.

Putting the two stages together, we see that h and k together with the relevant circumstances imply that the apple will hit Newton's head. Notice that h and k are necessary in this implication. The relevant circumstances alone do not imply the coincidence of apple and head. Notice also that h and k and the relevant circumstances and the coincidence are, taken together, logically consistent.

The great beauty of Newton's theory is its success in explaining, with simple means, practically every mechanical and gravitational phenomenon known at that time, together with its successful prediction of many more. Perhaps the most spectacular prediction was the existence of Uranus. Indeed the theory possesses nearly every virtue described in the list of virtues given at the end of this chapter. It is, therefore, not too easy to describe $\mathcal{Z}$ formally.

In order to be able to formulate and criticise a theory, our robot should therefore possess answers to the following four questions:

<u>H1</u> Is h justified given f, e and k?

<u>H2</u> Is there a means of telling when h is justified, given f, e and k?

<u>H3</u> Does h provide a good explanation in that it is very nice (perhaps maximally) with respect to $\mathcal{Z}$ amongst those hypotheses which explain f, given e and k?

<u>H4</u> Is there a means of finding such a maximally nice h?

Answers to <u>H1</u>, <u>H2</u> and <u>H3</u> will enable the stage of criticism to be performed. An answer to <u>H4</u> is a hypothesis construction method. Before further analysis, it is helpful to consider four analogous questions which arise in mathematics.

Suppose we are looking for a theorem Th in some axiomatic theory T. We want a Th which provides a best answer to some question, Q, about T. Answers are needed for the following questions:

<u>D1</u> Does Th follow from T?

<u>D2</u> Is there a means of telling when Th follows from T?

D3 Does Th provide a best answer to Q?

D4 Is there a way to find a Th that is a best answer to Q?

Answers to D1, D2 and D3 represent the criticism stage and D4, the discovery stage in some process of mathematical activity. It may be the case that both hypothesis formation and mathematical discovery can be made to form part of some more general process of knowledge generation. However it should not be imagined that questions H1 and D1 will merge; or questions H2 and D2 and so on. This is because, for example, answering D4 may involve formulating hypotheses or proceeding by analogy with previous results. Similarly it may be necessary to prove theorems to show that h explains f, given e and k.

The distinction between questions D1 and D3 reflects a distinction between truth and interesting appropriateness. When, as here, we do suppose that both Th and T are formulated in the first order predicate calculus, then the Tarskian semantic notion of logical consequence is usually, and justifiably (Kreisel, 1967) taken as a proper formal analysis of the informal notion of logical consequence. Such a situation does not obtain for higher-order logics or modal logics or, certainly, natural language.

Any complete and consistent system of proof for first-order logic gives a correct, but only semi-effective, answer to D2. Note that, in general any answer to D2 can be consistent and/or complete with respect to D1. Similarly any answer to D4 can be consistent and/or complete with respect to D3. In all cases, consistency and completeness are necessary conditions for the coherence of the entire system.

A further condition of coherence is that the Th discovered as a best answer to D4 should, in fact, follow from T.

This condition seems, on the whole, rather strong. What would be better would be some process which, given a conjectured answer to Q, would either prove Th or suggest, perhaps using a constructed counter-example, some alteration to Q or Th.

Answers to D2 and D4 may be, or may not be, efficient. Not much work has been done on the efficiency of systems of proof. (Kreisel, 1970, Kowalski, 1969, 1970). It seems likely, however, that while there can be no most efficient system of proof, existing methods can be greatly improved. As regards D3 and D4, the amount of systematic work is practically zero, with the well-known exception of that of Pólya (1954, 1957, 1968).

An answer to H2 may be consistent, complete, and/or efficient with respect to H1. Similarly, an answer to H4 may be consistent, complete and/or efficient with respect to H3. A global coherence requirement is that maximally nice h's, which explain f, given e and k, should be justified given e and k. Although this seems rather strong, as does the analogous requirement in the deductive case, nonetheless we shall see in Chapter 5 that it will be satisfied in some simple cases, if e is sufficiently "large".

We will outline some of the points made by philosophers about H1-H4, before settling down to giving a more detailed account of H3.

The distinction made in H1 and H3 between justification and niceness reflects the importance of questions about the truth of hypotheses. At one time, it was held to be possible to discover in a fixed, finite number of steps, important general truths about the world. Mill was about the latest philosopher who halfway believed this. Once it was demonstrated that there was in general no way to determine the truth of general statements, opinion divided, roughly, in two. The Carnapians describe a logical confirmation function, c(h,e'), (Carnap, 1952). This quantity, c(h,e'), has been variously interpreted as the logical probability of h given e', or the betting odds that a rational man would accept on h's being true, given e'. The hypothesis, h, and the evidence for its being true, e', are both framed in first-order logic. Carnap, and this is a quite general opinion, holds that a science can be stated as a first-order axiomatic theory (Carnap, 1967). In our case we would take e' to be e and f and all the true observations contained in k. In any practical situation this would be a hopeless task and only the relevant parts of k would be considered. Probably this would include hypotheses not known to be true and to which only a degree of belief had been assigned. A predicate, Ac(h,e'), can be described, in terms of e', which specifies, in terms of a high, a posteriori confirmation, when h should be accepted given e', (Hintikka and Hilpinen, 1966). For a monadic language the confirmation function c is calculable, but for a general first-order language, there is not even a semi-effective way to calculate it (Hintikka, 1965, Kemeny, 1953, Putnam, 1956). If we require the strong global coherence condition to hold, it is unimportant that any efficient methods be sought. Nothing is known anyway about efficient methods in general. For the monadic case, there is a close relation with the problem of efficiently finding switching circuits (Quine, 1955).

Popper (1959) is less concerned with formal analysis. He notes that general theories can be falsified and makes this his central plank. For justification he would substitute the requirement that h survive in competition with other hypotheses an extended attempt to falsify it. Only strongly falsifiable and simple hypotheses should be chosen for consideration.

Carnap has little to say on H4, and Popper specifically excludes this from his consideration, as being in the psychological domain.

We turn now to looking at H3 in some detail. As this is not concerned with questions of truth, less has been said. An account of how general laws explain singular statements has been given by Hempel and Oppenheim (1948) and has been subject to criticism by, among others, Eberle, Kaplan and Montague (1961). Niceness is or is partly determined by simplicity. Popper has argued that the degree of falsifiability of a statement correlates strongly with its simplicity. Goodman (1961) rejects this. He also develops (1959) an account of the simplicity of the predicate basis of a theory.

We shall present a picture using a variant of Hempel and Oppenheims' explication of explanation and let the niceness be specified by a parameter. Values of this parameter corresponding, roughly, to different philosophical positions will be used and investigated at different places throughout the rest of the thesis.

The first important assumption is that h and k will be sets of first-order statements, framed in a theoretical and observational language. This over-simplifies the structure of the robot's belief-system k. In particular, all of k is now accepted as being true. This completely ignores the problem of how to deal with large collections of statements to which varying degrees of belief are assigned. Nor does Popper escape a similar dilemma. When k has been falsified, which part of it should be replaced?

The hypothesis, h, will in addition be restricted to be a member of $\mathcal{H}$, the hypothesis space.

The set of phenomena needing to be explained, f, is a set,

$f = \{f_i \mid i=1,n\}$ of first-order singular sentences.

$e = \{e_i \mid e=1,n\}$ is also a set of first-order singular sentences.

We will symbolise the relationship between each e_i and f_i by $\Rightarrow$. We might regard $\Rightarrow$ as a modal connective and say that the set of statements, $\{e_i \Rightarrow f_i \mid i=1,n\}$ in a certain modal logic represent the relation between the phenomena and their antecedents.

The interpretation of $e_i \Rightarrow f_i$ is, roughly, that e_i is a description of circumstances or preconditions which resulted in f_i's being true. While it may be a good idea to actually use a modal logic, we could not find any suitable one, and decided to deal with and use $\Rightarrow$ informally on the meta-level of our present discussion. Here is a list of some indications of possible interpretations of $e_i \Rightarrow f_i$.

1) Actions: f_i is the result of an action, or series of actions, described by e_i in circumstances also described by e_i.
2) Direct Cause: e_i is the description of a direct cause of f_i.
3) Experiment: e_i is the description of how an experiment was set up and f_i is the description of its result, for example f_i may be (a description of) a graph (such as one of temperature against pressure obtained from some experiment with gases).
4) Empirical Association: whenever e_i is true f_i occurs (with observed frequency such and such).
5) Temporal Succession: the event f_i followed very soon after the event e_i.
6) Essential Property: f_i is an essential property of objects with the description e_i. For example, having a date inscribed is an essential property of pennies, but being in Harry's pocket is not.
7) Definition: e_i is the appropriate case of the definition of the predicate occurring in f_i.

In the above, as in many places throughout this discussion, we are both using and mentioning the e_i and f_i. It is hoped that the reader will distinguish the different cases.

These possibilities by no means form an independent or complete list. We will adopt 3) as our standard interpretation. With this in mind, we restrict $\mathcal{H}$ to be some subset of the set of lawlike sentences, that is statements whose prenex form contains only universal quantifiers. This is a departure from the standard notion, which, in addition, requires that h contain no individual constants.

We do this because we wish at one extreme to regard a singular statement as being a (completely uninteresting) law. Further we want the niceness criterion, $\mathfrak{z}$, to be the factor that strives toward generality. Finally, we feel that the simple grammatical notion of absence of individual constants does not capture the ideal of generality exactly. Many general scientific laws do contain constants - for example the charge of an electron.

Next, we specify what it is for h to explain f given k and e. The belief system, k, is divided into two parts, Th and Irr. Thus $k = Th \wedge Irr$. The system Th consists of that part considered relevant to the set of phenomena at hand, and Irr is the rest, the irrelevant beliefs.

Then, h explains f given $k = Th \wedge Irr$ and e iff:

__E1__ For all i, $\vdash_{Th} h \wedge e_i \Rightarrow f_i$.

E2 For all i, it is not the case that $\vdash_{Th} e_i \Rightarrow f_i$.

E3 The sentence $k \wedge h \wedge \bigwedge_{i=1}^{n} (e_i \wedge f_i)$ is consistent.

E4 The hypothesis h is lawlike.

Strictly speaking, E3 should be written as 'The set of sentences $k \cup \{h, \bigwedge_{i=1}^{n} (e_i \wedge f_i)\}$ is consistent'. This, and similar, confusions will be perpetrated throughout the thesis.

Requirement E1 is certainly the least that h can do if it is to explain each f_i, given e_i and Th. The second requirement, E2, ensures that the phenomena are not trivial, that is that there is something new to be explained. Requirement E3 is a minimal requirement if h is to be incorporated into the body of beliefs. Whether or not one's beliefs are true, they should certainly be consistent with one's observations. Requirement E4 is imposed as a result of the interpretation we adopted of =>. Other interpretations would require other conditions.

By themselves, E1-E4 would not capture some of the "feel" of explanation. This should perhaps be inserted in the description of the niceness relation, $\dashv$. Mario Bunge has compiled a very impressive list of possible "niceness" qualities of hypotheses (1961). We will give very brief accounts of each of them. This will give yet another indication of the sheer size of the problem.

Syntactical Requirements:

1) Well-formedness. For us, this is just the requirement that h actually be a wff of the first-order predicate calculus.

2) Connectedness. If a hypothesis is thought of as a conjunction of postulates, and each predicate symbol occurs in many of these postulates, it is well-connected.

Semantical Requirements:

3) Linguistic Exactness. The ambiguity, vagueness and obscurity of a hypothesis should be minimal. Such terms as 'hot' or 'historical necessity' are not welcome.

4) Empirical Interpretability. The hypothesis must make empirical predictions.

5) Representativeness. The theory should deal with actual events and processes. Thus theories of action at a distance are replaced by field theories, showing exactly how action at a distance actually works.

6) Semantical Simplicity. The world should be constructed simply from simple parts. The theory of quarks is an extreme example.

Epistemological Requirements:

7) External Consistency. The hypothesis should be consistent with the bulk of one's knowledge.

8) Explanatory Power. The hypothesis should explain many known empirical facts and generalisations.

9) Predictive Power. The hypothesis should entail many unknown facts.

10) Depth. The hypothesis should explain essentials and reach

deeply into the structure of reality.

11) <u>Extensibility</u>. The hypothesis should be extensible in order to cover new domains, not previously thought of as being relevant to the hypothesis.

12) <u>Fertility</u>. The hypothesis must have exploratory power.

13) <u>Originality</u>.

<u>Methodological Requirements</u>:

14) <u>Scrutability</u>. The predicates involved in the hypothesis must be open to scrutiny by the general public. That is, techniques, tests and evidence must be intersubjective. For example, events should not occur through God's will, nor should Mrs. Smith's female intuition count as a theoretical entity.

15) <u>Refutability</u>. It must be possible to imagine circumstances which could refute the hypothesis. Critical experiments can be set up.

16) <u>Confirmability</u>. The theory must have consequences which agree with observation.

17) <u>Methodological Simplicity</u>. It must be technically possible to subject the theory to empirical tests.

<u>Philosophical Requirements</u>:

18) <u>Level Parsimony</u>. The hypothesis must be parsimonious in its references to sections of reality other than those directly

involved.

19) Meta-Scientific Soundness. The hypothesis must be compatible with fertile metascientific principles such as the requirement that it be lawlike.

20) World-View Compatibility. The hypothesis should be in line with the general world-view of scientists. One should be led to reject crackpot theories, but accept, eventually, genuine scientific revolutions.

It will be seen that some of these requirements are, from our systematiser's point of view, no more than hopeful hand-waving. Some have been covered, some seem extremely hard to formalise and some impossible.

Let us recapitulate the parameters entering into H3.

There is a hypothesis space $\mathcal{H}$.

There is a set of phenomena and their circumstances $\{e_i \Rightarrow f_i \mid i=1,n\}$, with restrictions on the nature of the e_i and f_i and some interpretation of $\Rightarrow$.

There is the knowledge, $k = Th \wedge Irr$.

There is the type of explanation through which each e_i is explained by h(in $\mathcal{H}$) given k and $f_i(i=1,n)$.

There is some notion of niceness, $\mathcal{N}$.

Once these parameters have been specified one obtains a method of answering H3 given h, and a formal problem:

"Answer H4 so as to obtain a consistent, complete and efficient algorithm for finding an h which answers H3."

Such an algorithm should cover a large range of possible sets of phenomena, and perhaps some range of possible belief systems k.