MARSDEN S. BLOIS

CONCEPTUAL ISSUES IN COMPUTER-AIDED DIAGNOSIS AND THE HIERARCHICAL NATURE OF MEDICAL KNOWLEDGE

ABSTRACT. Attempts to formalize the diagnostic process are by no means a recent undertaking; what is new is the availability of an engine to process these formalizations. The digital computer has therefore been increasingly turned to in the expectation of developing systems which will assist or replace the physician in diagnosis.

Such efforts involve a number of assumptions regarding the nature of the diagnostic process: e.g. where it begins, and where it ends. 'Diagnosis' appears to include a number of quite different cognitive processes, some of which seem more subject to formalization than others. Underlying the difficulties inherent in these efforts at formalization, is the hierarchical structure of medical knowledge including that of diseases. (By 'hierarchical structure', I mean no more than that our descriptions of clinical objects necessarily refer to different levels of organization; atomic, molecular, cellular, physiological system, entire patient.)

Since diagnosis is in part a form of classification, measures of similarity between the clinical findings of a particular patient and a potential disease, are needed. This is complicated by the fact that some diseases are described using the vocabularies of physics and chemistry, others with the vocabulary of psychology, and most, with terms taken from a number of different hierarchical levels. The differences between the descriptions made using these vocabularies will be seen to exert systematic (if not uniform) influences upon the diagnostic process.

The degree to which computers may be useful in the aid or replacement of physicians in diagnosis, continues to be a topic of lively interest. This interest springs from a variety of views of the subject, with each contributing its ethical, and epistemological flavors. My purpose here is to explore some of the activities to which the term 'diagnosis' is applied, to consider how these processes may be

Marsden S. Blois, Section on Medical Information Science, University of California at San Francisco.

The Journal of Medicine and Philosophy 8 (1983), 29–50. 0360–5310/83/0081–0029 $02.20.

affected by the structure of medical knowledge, and to ask where computers may fit into these activities. Recent reports, for example, confirm that in certain "diagnostic" applications (Meehl, 1954, 1973; Warner, 1964; Crooks, 1959; Fitzgerald, 1966; Leaper, 1966), the use of Bayesian methods permits the accurate allocation of patients amongst a small number of mutually exclusive disease categories. The question is whether this is all, or the principal part, of what is meant by 'diagnosis'. And although such computations have well-defined starting and ending points, we will ask where the diagnostic process itself begins and ends (Blois, 1980).

The diagnostic programs cited above, begin with "well structured problems", in which the carefully organized data supplied to a computer program have already had their *relevance* certified. Physicians, in making a diagnosis, begin with no such things; they have only patients with complaints. For these Bayesian programs to function at all, then, some human must earlier have decided that in the case of *this particular patient* with pain *here* (e.g., the patient gesturing in the direction of the xiphoid), one should run a particular computer program (e.g., the chest pain program rather than the one for abdominal pain). Making this choice is possible only because someone has decided that certain of the patient's complaints and findings are relevant, and the rest merely coincidental and contingent. It is somewhere near here, in this early sorting out of affairs, that "diagnosis" would appear to begin.

Central to all of this is the matter of what physicians mean when they speak of "diagnosis". This appears to include several rather different kinds of activities:

(a) The application of a particular label to a patient (i.e., the assignment of a patient to a conventional disease category). For example, a patient will be said to have "acute depression" or "vitiligo" because one simply recognizes the disorder. With these diseases, the abnormality observed *is* the disease. Making the proper diagnosis in such cases is essential for choosing treatment, but it may explain little about the disease process.

(b) The detection of a particular malfunctioning or missing part (i.e., a physiological system, an organ, or a type of molecule), e.g. "renal failure", "mitral insufficiency", or "phenylketonuria". With these and similar diseases, the diagnosis (in conjunction with a knowledge of the disease) may provide a causal account for a number of the observed abnormalities.

(c) The identification of what is believed to be a proximate *exogenous* cause of an illness, e.g. "traumatic fracture of the neck of the femur" or "schistosomiasis".

(d) The identification of a process which is incomplete or temporarily suspended, before any of the above goals has been attained. A patient may be said to have an "infectious disease", a "surgical abdomen", or an "auto-immune disorder", with a more specific diagnosis awaiting the receipt of further information. What counts as "incomplete" depends on the state of medical knowledge which changes over time. The diagnosis "pneumonia", a perfectly acceptable one a century ago, would be regarded as incomplete today. We would insist on distinguishing between pneumonias due to viruses, bacteria, or to chemical irritants, since the appropriate treatment will differ. If a particular case of pneumonia were found to be bacterial in origin, we would expect to go on and identify the responsible organism and test its antibiotic sensitivity, so that an appropriate drug could be chosen.

(e) Because our knowledge of diseases and their appropriate treatment is uncertain at the edges, there is no formal end-point of the diagnostic process. There are two other factors operating here as well. One is the well known fact that if two different diagnoses are equally supported by the evidence, and one is a curable disease while the other is not, the physician will opt for the former. The other relates specifically to computer diagnosis and the fact that the physician will always have access to patient information which is not available to the program, e.g. that one of the diseases in the differential is at the moment epidemic in the patient's neighborhood. A physician may judge this to be relevant. Since only the physician can accept responsibility for the diagnosis, he may allow himself to be influenced by such information.

These meanings of 'diagnosis' are not disjunctive. They are vague, and they are not exhaustive, but they suggest the kinds of things physicians commonly have in mind when they use the term.

DIAGNOSIS AS EXPLANATION

Although a common hope in diagnosis is to be able to provide an account for all the abnormal clinical findings, the attempt may fall short. In actual cases, the diagnostic process may terminate with the application of the name of some recognizable state such as "acute depression", providing no clue as to cause or mechanism. With other diseases, a diagnosis may make contact with causal linkages threading through a considerable body of medical knowledge, as seen with "phenylketonuria". There are important systematic differences between these two examples, and it will clarify matters to explore one of the meanings of 'explanation' in medicine. We will do this by considering the "hierarchical" features of the things involved.

In his Gifford Lectures, Eddington speaks of his "two tables". One, he explains, is the wooden table before him, upon which his ink-well and writing materials rests (Eddington, 1964). The other is the swarm of atomic nuclei and electrons (and mostly empty space) of which the table is composed when reduced to its (then known) "ultimate" constituents. Eddington then goes on to ask which of these two tables is the "real" one, and he answered, the latter. Nagel would, of course, disagree with this view of "reality", pointing out that the word "table" is not included in the vocabulary of atomic physics (Nagel, 1961).

Eddington might have answered differently if he had avoided this discontinuous transition from an ordinary table to a collection of atomic particles. He could easily have done so by pointing out that a table may be constructed out of lumber, that wood exhibits much of the structure of a formerly living tree, that it consists of cell walls, cementing substances, and other structures. These in turn are composed of biopolymers such as cellulose and complex molecules like lignin; moreover, these molecules are built up of simple sugars and smaller organic chemical units, and so on down to atomic particles. Since these intervening levels of organization are more intimately related, we can find points of contact between them, and frequently account for some of the properties of an object at one level in terms of the properties of those next below. On occasion, we may reverse this process, providing higher level accounts of lower level phenomena.

Let us apply this model to the human organism, and employ the same decomposition procedure. At a high-level of the hierarchy, we have entire, ordinary, functioning individuals. These we attempt to describe in the natural language used in our daily activities. We may then construct a sequence of whole-part relationships embodying our knowledge of the organization levels involved: the individual, organ system, organs/tissues, cells, cell parts/organelles, subcellular structures (membranes), biopolymers, molecules, and atoms. If we wish, we may continue on beyond the atomic particles known to Eddington, to the elementary particles (quarks, etc.) of modern physics. These whole-part relationships underlying the organization of matter, give rise to a series of hierarchical levels at which we describe events and construct theories. We must ask, however, whether these levels represent "natural" categories, or whether they merely reflect a convenient but arbitrary set of abstractions. That is, are they in some meaningful sense "real"? Such organizational levels arise from the decomposability properties of natural objects according to Simon (1962) and Bronowski (1970) and this view is not lacking in empirical support. While we can produce (or occasionally find) objects (e.g., atomic and molecular fragments, cells without nuclei, etc.) which do not fit squarely into these categories, they appear to be short-lived. The metaphysical arguments for such a multi-level ontology (for the case of biology), have been given by Grene (Grene, 1967)[1]. Here, we will simply note those features of hierarchical organization which seem most likely to affect the process of diagnosis.

To simplify matters, let us consider the levels corresponding to humans (high-levels), and atoms and molecules (low-levels), while keeping in mind the existence of the many intervening levels which serve to connect them. It goes without saying that low-level objects are best described using the vocabularies of physics and chemistry. Using these vocabularies, objects such as carbon atoms may be described as radioactive or not, being in excited states or in their ground state, or residing in this or that location. They cannot be spoken of, however, as being in pain or doubt, or being good or evil. The former attributes are low-level ones, while the latter are properties of high-level things, and, in particular, of humans. Our scientific grammar does not permit the casual mixing of terms belonging to the vocabularies of different hierarchical levels. When we predict what will happen to an atom under certain conditions, we must consistently use the vocabulary of physics. As a consequence of this, solving a physics problem primarily yields a physics answer. The same is the case with problems in chemistry, endocrinology, or psychology.

As a consequence of this multi-level ontology, scientific knowledge exhibits a curiously laminated quality. This is not to say that scientific explanations cannot extend across hierarchical levels because they may, at least in certain instances. And it is just these level-connecting or bridging explanations which make reductionism work. For example, the chemical bond (e.g., the covalent bond) has a satisfactory physical explanation, and quantum mechanics accounts for a number of other chemical properties as well. Yet not all of chemistry has in the same sense been reduced to physics. There does not appear to be a complete quantum mechanical theory of organic reactions which would replace the intuition of the synthetic organic chemist, and the use he or she makes of "named reactions" and of a knowledge of catalysis. Most scientific theories seem to have the property of accounting for things at one or two adjoining levels, and of being largely "horizontal" affairs. In a similar vein, biochemistry and physiology have their own theories operating at their respective levels, and new discoveries in these fields are made most readily along these levels. Systematic "vertical" explanations of significant range are uncommon.[2]

It is in the science of biology and the enterprise of medicine where we encounter a particular need for vertical explanations and where we experience an urgency to account for a phenomenon at one level by referring to states of affairs lying at lower or higher ones. In the history of science, the most striking achievements in explanation have been attained by carrying out a "downward" analysis and in detecting "upward causation" (i.e., explanation and causality flowing in opposite directions). Though our scientific understanding of the natural world has largely been achieved by an exploratory descent down these hierarchical levels, the reductionist (or "decompositionist") procedure would seem to be more of an historical fact than a deliberate research strategy. The opposite program of "upward" explanation (and "downward" causation) in biology can also muster empirical support in the case of stress-induced diseases and, more generally, of psychosomatic disorders. However, this approach runs counter to the prevailing paradigm, and has not achieved the power of the reductionist one.

Exploring the implications of this hierarchical model, we can summarize a few of the properties encountered with high and low level natural objects, and with their descriptions.

(a) Low-level objects give an impression of having fewer of what we will call "necessary" properties than do high-level ones. Put differently, the descriptions of such objects require fewer predicates. There are fewer relevant things that need to be said about them in order to account for their behavior. Under a "classical" atomic theory, we can see this as we proceed from the level of sub-atomic particles to that of atoms. The former (i.e., electrons, protons) are observed to have mass, electric charge, and magnetic moment. A satisfactory theory of these objects need only account for the observed relations of these attributes with general physical laws. A century of experimentation has disclosed no further properties. These attributes are what we may regard as *necessary* attributes, the specific values of which provide the means by which these objects are recognized or described under the theory. They also have innumerable *contingent* attributes such as specific location or velocity, the values of which are not permanent properties of the objects. These latter, while useful in describing them, play no role in recognizing them. Low-level objects such as these also have single names; they can be sharply defined and turn out to be monotonously uniform. For instance, the word 'electron' refers to one class of distinct objects, all of which are exactly alike (thus permitting no room for individuality). Neither is there a synonym for 'electron', nor the slightest need for one. Moreover, there is no ambiguity in the meaning of the term.

When we move to the next higher level – that of atoms – we find they have (actually or potentially) the attributes of the sub-atomic particles of which they are composed[3]. But additionally, we discover them to have still other properties which do not occur at the lower level. These kinds of properties are *emergent*[4] ones, and they play an important role in the present account. The emergent properties of atoms include such features as excitation, polarizability, and ionization, none of which are attributes of protons or electrons. Rather, these arise discontinuously at the next higher level (i.e., atomic) of organization. Emergent attributes frequently seem not to be predictable by theories which are thought sufficient to account for the behavior of objects at the lower level. When we better understand the higher levels, perhaps we will find examples of such theories. Do the psychological theories of the individual, which, while satisfactory at this level, fail to predict the properties which emerge in societies (e.g. mob behavior, political behavior)? Evidently, with each increase of hierarchical level, newly emergent properties appear, and they appear to do this more rapidly as we proceed to the higher levels.

(b) Low level objects not only appear to have fewer attributes, but our descriptions of them are sharp and clear ("hard edged"). No natural objects are known which are "sort of like" electrons (although there are many non-dogs which are apt to be confused with dogs). It is also appears possible to distinguish between the attributes which are necessary for the class-membership of low-level objects and those which are contingent. When, in contrast, we turn to higher level objects such as dogs, the number of attributes becomes much greater because of emergence. Instead of them all being alike, uniqueness appears, and it may become difficult to decide whether a particular attribute is necessary or contingent. (Is a dog still a dog if one of its legs were amputated? Two of its legs . . . ? If it were cut in two? Is a two headed "dog", a dog?)

(c) Vagueness and ambiguity in the use of the words which stand for objects and their attributes seems to increase when we consider higher level objects. There is no ambiguity with the word 'electron' or the expression 'electrical charge' (although these terms are just as much a part of natural language as 'fear' or 'love') because there is nothing we are likely to mistake them for. But we do have serious problems with 'dog' (what is the precise difference between dog and wolf?), with 'human' (what is the status of a patient with a flat EEG?), and with attributes like courage and honesty.

In contrast with low-level objects, high-level ones seem more complex, and the vocabularies we use in describing them often contain terms which seem vague, ambiguous, or "fuzzy" (Zadeh, 1976).

When we undertake to classify such objects, it is more difficult to distinguish between those attributes which are required for class membership, and those which may be merely contingent.[5] These changes appear to set in progressively as we move away from the objects of atomic physics or chemistry to the more complex ones of the biologist, the ordinary objects of everyday experience, and the still higher level phenomena of the sociologist and economist.

THE ABSTRACTION OF HIERARCHICAL DESCRIPTIONS

While natural language is our most powerful general tool for describing and analyzing the world, mathematics (including logic) is sometimes spoken of as the "language of science". Yet, while the term 'science' seems equally appropriate to the study of objects and processes at each of these hierarchical levels we have considered, mathematics seems most useful at the lower descriptive levels; it quickly loses its power as we encounter higher-level events.

We may analyze this behavior in the following way. Mathematics and classical logic seem applicable to everyday problems, only after these problems have been abstracted and formalized. Neither can be applied directly to the raw and unformalized situations which we encounter in the world. When we are faced with (immersed in) situations, we perform abstractions of them, simplifying them and formalizing them as we do so. Only then do we even have "problems". The result of this may be abstractions simple enough so that we can apply mathematics. The features of high-level objects entail long and complicated descriptions, and these are most in need of simplification. In performing these abstractions, we hope to identify and retain the necessary attributes and to discard the contingent ones. But as Whitehead reminded us, " . . . an abstraction is nothing else than the omission of part of the truth" (Whitehead, 1938). And the necessary and contingent attributes become difficult to distinguish from one another in the case of high-level objects such as humans, Eddington's tables, and Wittgenstein's games.

Nevertheless, by discarding much of the truth, we can frequently abstract situations to a point where mathematics can be applied. How useful this will prove, depends upon how much and what kinds of truth have been discarded during the abstraction. Since it seems that with the low-level objects of physics and chemistry things are simpler to begin with, we need not throw away so much. Our risk of discarding something important is far less. The power of mathematics, I would suggest, is greater with lower-level objects and processes, because the needed abstractions more closely resemble native situations than high-level ones.

THE HIERARCHICAL NATURE OF DISEASES

Medicine is one of the few human activities which is concerned with events at all of these hierarchical levels. The abnormalities found in sick patients may be an excess (or deficiency) of protons in the urine, too many or too few potassium ions in the serum, or too many (or too few) sugar molecules in the blood. In order to describe the diseases associated with these particular abnormalities, medicine requires the vocabularies (and tools and concepts) of physics and chemistry. These diseases will have other abnormalities describable only at higher levels, such as alterations in neural transmission or muscular response, and symptoms such as weakness or dizziness. To describe these, the vocabularies of physics and chemistry are insufficient, and we need those of physiology and psychology as well. We note that patients are not conscious of low-level abnormalities, that humans cannot experience a low blood-glucose level, although they may well experience the associated faintness. But neither the patient nor the physician (hearing a patient complain of faintness), can tell from this alone whether it is due to a low blood glucose, or say, a reduced cerebral blood flow. It is with respect to the hierarchical levels at which we describe, diagnose, and prescribe, that diseases differ so greatly in kind.

Some diseases, in their early stages, may have only low-level abnormalities and not yet produced symptoms, so that the individual in whom they are found is not "ill" (e.g. "chemical" diabetes, asymptomatic hypertension). Other diseases, in contrast, may have striking high-level abnormalities which cause the patient severe distress, and yet have no known lower-level dysfunctions (e.g. schizophrenia). Since low-level attributes can be detected only with the use of instruments or laboratory tests, they have come to be regarded as the "objective" evidence of disease, while the high-level ones (dizziness, pain, malaise), learned from a patient's account, are considered "subjective", and may be accorded a lower diagnostic worth.[6] Yet it is just the latter which makes people ill, and for which they seek relief.

A few diseases are understood in sufficient detail and over a number of different levels so that coherent causal accounts (in particular, those operating vertically and bridging levels) can be provided. Some diseases (e.g. inborn errors of metabolism) have features which can be more or less continuously traced from faulty gene expression to a missing enzyme, to the failure to carry out a particular chemical reaction, to structural abnormalities, and finally, to functional abnormalities and an awareness of being ill. When such a disease is diagnosed, a number of abnormalities may be explained at once and the diagnosis will have a high explanatory value.

Lest it be overlooked that "causality" works both up and down, we may consider the case of a patient who, after stressfull and unpleasant episodes at his work, has finally been fired. This situation is a high-level affair. And it may have resulted in the secretory cells of his stomach producing, in excess, hydrochloric acid and digestive enzymes. The stress to which he was exposed may well have produced other changes of which we remain ignorant, but the net result is the self-digestion of a small portion of his duodenal mucosa, and the patient now suffers from a peptic ulcer. There is a considerable body of evidence, describable at several hierarchical levels, which supports in broad outline the validity of such a causal account. There are thus "causal" connections between events in the world, and the behavior of cells and molecules in our bodies, as well as *vice versa*. And one of the meanings of 'diagnosis' is the attempt to identify and trace these connections, whether they lead us up or down the hierarchy.

THE COMPUTER PROCESSING OF LOW-LEVEL AND HIGH-LEVEL DESCRIPTIONS

With 'diagnosis' meaning so many different things and given the effect of hierarchical organization upon the descriptions of objects and processes, what does all this imply about computers and their use in diagnostic processes? We note first that computers can do no more than process the descriptions of things, that is, to manipulate symbols which stand for real or abstract things. The lowest-level attributes of diseases are the subject of physics and chemistry, and here we are least troubled with vagueness, ambiguity, and fuzziness. To such descriptions we can apply both mathematics and computers. It seems no accident that one of the earliest computer programs to be used in clinical medicine was Bleich's acid-base program (Bleich, 1969). Central to its function is the Henderson-Hasselbalch equation which provides the physico-chemical description of acid-base equilibria. As we examine phenomena occurring higher up in the hierarchy, we leave the "hard" sciences behind, and our descriptions of things become "softer". The fit between word and intended object becomes looser. When a patient says that something "hurts", we may ask whether it is a "pain", an "ache", or a "cramp" since all of these "hurt". We may go on and ask if a pain is "sharp" or "dull", "stabbing", "radiating" or "localized", "continuous" or "remitting". Describing a pain is a fuzzy undertaking, and we may count ourselves lucky if we can obtain a description corresponding to one of these conventional categories.

At the level of cells (e.g. a bacterium), individuality has already emerged; no two *E. coli* organisms are precisely alike, nor does a single one remain unchanged for long. At this level of hierarchical complexity, our abstractions begin to be seriously incomplete. If we leave in enough detail to make matters realistic, the abstractions remain too complex to express in mathematical terms. But if we simplify them sufficiently so as to be able to apply mathematics, they lose their biological relevance. When we do use mathematics in biology (at other than the very lowest levels), we succeed in capturing only selected aspects of phenomena and not an entire phenomenon as we do in physics. Explanation in biology and medicine cannot, it would appear, be satisfactorily managed with mathematics.[7]

Computers have long been known to be capable of carrying out functions beyond those which are commonly thought of as being mathematical (although they are no less restricted to processes which in a technical sense are "formal" and thus equivalent to mathematics). Computer programs operate best with the same kinds of abstraction as does mathematics. Each requires the use of vocabularies that are free of ambiguity. Neither mathematics nor computer programs could function if the meaning of some operator depended upon context. In mathematics, there is no provision for context; and with computers, the only thing they can "know" about contexts is that which has been explicitly represented in the programs. In order to represent in a computer program contexts rich enough to deal with ordinary human problems, a programmer would have to supply what we regard as commonsense knowledge. This task, as has been argued elsewhere (Dreyfus, 1979, Boden, 1977), raises serious epistemological questions about the limitations of computer programs.

Providing explanations, such as we might expect from a human, may be asking too much of computers. A more useful question may be, "In what ways might computers be expected to be of help to physicians in diagnosis?" This turns the matter around, and invites us to make use of computers where they perform best. Where might this be?

It turns out that there are a number of such applications. We will consider two of them. The kinds of mistakes which physicians make in diagnosis have been carefully studied (Elstein, 1978). One common method of diagnosis is that of hypothesis-creation and deduction, the procedure which C. S. Peirce called "abduction". Here, the physician first creates a list of hypothesized diseases which could account for the clinical findings in a particular case. Utilizing the principle of parsimony, he or she will, if possible, choose diseases each of which could account for all the findings, and invoke multiple diseases only when necessary. This list can then be refined deductively, by obtaining further data which will confirm or reject the hypothesized diseases in turn.

The point at which the hypothetico-deductive process is most likely to fail is of course the first step. A physician cannot diagnose a disease without having first thought of it. The correct diagnosis may not have been included in the original list because it is uncommon, the physician's knowledge of it may be sketchy, he may have been hurried or fatigued, or the patient may have presented with an unusual combination of findings or with multiple diseases. All of these might be overcome to some degree with the use of computers, and this "prompting" use of computers has been widely advocated. Scadding (1967) has referred to this as the "library" function of the computer in diagnosis.

Lindberg and his associates (1968) wrote a program of this sort a decade or so ago called CONSIDER. This program matched patient attributes which had been entered (using direct lexical matching) against the descriptions of a group of over 3000 diseases, and listed all the diseases which were characterized by these attributes. We have described elsewhere a program called RECONSIDER, which operates much as did Lindberg's, with the additional features that disease-attributes are weighted according to the "selective power" of the attributes, so that the final list of diseases can be ordered (Blois *et al.*, 1981). A synonymy capability was also provided so that if a particular attribute could be described by more than one word (e.g., 'itch' and 'pruritus'), the appropriate matches could still be made. Programs such as these two are not intended to produce a single, correct diagnosis but to suggest that certain diseases be considered for inclusion in the physician's differential diagnosis. They are perhaps best described as "diagnosis prompting programs".

"KNOWLEDGE BASES" AND "EXPERT SYSTEMS"

Another class of computer programs, which have become known as "expert systems", has been developed and studied in recent years. The goal of their designers has been to attempt to capture the special knowledge of experts in a given subject matter field and to encode or represent this in a "knowledge base" which will be computer-processable. These attempts involve two separate kinds of activities: (1) the development of programs embodying the logic to be applied, so that the computer will function as an "inference engine", and (2) the employment of a suitable means for representing this special or expert knowledge. These activities now constitute a significant portion of the field referred to as "artificial intelligence" (AI), and they raise further conceptual issues.

The idea that there exists a certain kind of knowledge called "expert knowledge" suggests that there may be another sort which we might call "commonsense knowledge". The latter may be viewed as the kind of knowing which permits people to deal with the every-day circumstances of life. This is contrasted with the special knowledge exhibited in the performances of, say, experienced chemists or highly successful diagnosticians. This attempted demarcation raises difficulties, however, since the effectiveness of expert knowledge might in some fields presuppose access to commonsense knowledge as well. There is also the further question as to how much of an expert's performance is due to skill, and how much to propositional knowledge. This is the difference between "knowing how" (Polanyi's "personal knowledge"), and "knowing that" (Popper's "objective knowledge"), with the latter being the sort that is most readily articulated and incorporated into computer programs. If Polanyi is correct that portions of "knowing how" cannot be translated into explicit rules, and if a significant part of the expert diagnostician's performance derives from this kind of knowing, then attempts at formally describing it would seem to be in for trouble (Polanyi, 1958). The difficulty, it should be stressed, lies in having the expert articulate his relevant knowledge in propositional form. But rather than inquiring into these general issues, let us see what the experimental data reveal in actual cases.

Two large scale expert systems have been developed in recent years in the attempt to capture and apply the knowledge of chemists and diagnosticians to practical problems. One is the program called DENDRAL, which was developed for use in analytical chemistry (Buchanan and Feigenbaum, 1978). If the elementary analysis of an unknown compound permits a chemist to write the molecular formula say, $C_nH_mO_o$, the question remaining is which of the particular permissible structures corresponding to this formula is that of the unknown. If *n, m*, or *o* have values of more than perhaps 10, the numbers of potential chemical structures may lie in the many thousands. In order for a chemist to choose the correct one, further information is necessary. Two analytical instruments of wide applicability are available to provide these data: mass spectroscopy and nuclear magnetic resonance. Given the additional data obtained from these sources, DENDRAL could in principle then compute all the possible structures corresponding to the molecular formula (using the constraints imposed by valence, atomic size, bond angles, etc.). It could compare each of these many candidates with the fragmentation pattern and the n.m.r. spectrum. Such a brute-force method is computationally unattractive because of the combinatorial explosion which occurs with larger molecules. DENDRAL's designers have circumvented this by incorporating various chemical "rules" in the program. To obtain these rules, chemists were asked to articulate their knowledge in if-then form: "if the n.m.r. spectrum shows five equally-spaced absorption lines with intensities in the ratio . . . , then consider a methyl group . . . ". The use of sets of such rules excludes whole classes of potential candidates and greatly reduces the size of the search-space. The program thus more rapidly converges on the smaller number of structures which are compatible with both the molecular formula and the experimental data. Occasionally, the proper structure is uniquely identified at once. This program has become very useful in analytical chemistry, and it has been distributed to a number of different laboratories where it is in routine use. DENDRAL has come to be regarded by some as the most useful accomplishment in AI to date. It will be noted that commonsense knowledge of the world is not much called upon in this application; ordinary experience is not very useful in this low-level domain.

A second expert system, one under development for several years, is INTERNIST (Pople, 1977). The goal of this project has been to formalize the clinical expertise of Dr. Jack Myers, a highly experienced professor of medicine, in computer processable representations, so that a program using this knowledge could produce a diagnosis when provided with the clinical findings of a case. The operation of the program has been discussed by McMullin, and need not be recounted here. It is reported that INTERNIST has successfully performed with a number of difficult cases, including clinical-pathological cases taken from the *New England Journal of Medicine*. A systematic review of its performance, including an analysis of the difficulties encountered, has not yet been published, and an improved version (INTERNIST II, recently renamed CADUCEUS) continues under development.[8]

What I would draw attention to as distinguishing between these two programs is that DENDRAL operates in a problem-area lying at a low hierarchical level of description. The necessary attributes of atoms and molecules are well known and invariant, and the laws of chemical structures have been formalized. One can state rules about chemical and physical behavior which admit no exceptions. In this sense, analytical chemistry is hard (low-level) science, and though the process which DENDRAL carries out may be viewed as a kind of diagnosis, the process itself is readily formalized.

INTERNIST, in contrast, deals with an object which is describable over a wide range of hierarchical levels. At the lowest levels, the expert knowledge it uses is similar to that used by DENDRAL: a decreased creatinine clearance means decreased kidney function. No matter what the cause of this may be, the conclusion follows. And the values of the attributes are expressible in numbers, a characteristic of lower-levels. With higher level (clinical) attributes ranging from alopecia to xerosis, their association is for the most part with clusters or kinds of diseases, rather than with single ones, and they take on values more apt to be expressed in terms of "linguistic variables" such as 'marked', 'slight', or 'perhaps'. Such attributes can in principle be handled with Bayesian methods, either by the method of abstracting very large numbers of medical records and calculating the appropriate conditional probabilities, or by obtaining an expert's estimates of them. With such a list in hand (giving the likelihood that a patient with actinomycosis will have anorexia, that one with agranulocytosis will do likewise, . . .), Bayes' Theorem may do the trick. Developing such a matrix, however, presents formidable problems of scale, validity, and selection.

More difficult to understand is the antecedent step during which a patient's clinical abnormalities are represented by the physician with the use of carefully chosen descriptors. The evidence for these abnormalities are appearances, sounds and smells, and the words and gestures used by the patient in describing the illness. These are the clinical primitives, but it is not these which are written in the medical record or typed into the diagnostic program. Interposed between these sense data and the diagnostic process itself, is the physician, acting both as filter and interpreter. By deciding that certain features are irrelevant and others germane, the course of the diagnostic process is affected (even, determined) at the outset. And in calling a set of phenomena "proximal muscular weakness" or "Kayser–Fleischer rings", the physician is doing far more than reporting raw clinical data. When such findings correspond to high level observables, their interpretation and representation seems to call for greater skill than interpreting a laboratory test result. The specificity of a particular finding for a particular disease, has nothing to do with this; pathognomonic findings may occur at all descriptive levels, though increasingly at lower ones. The problem here is the insightful labelling of higher level clinical findings, a pre-diagnostic process which is simpler at lower levels.

It is not yet possible to compare the respective performances of DENDRAL and INTERNIST. However, when the necessary evaluations have been completed, I would expect (on the basis of the analysis sketched here) that INTERNIST will perform significantly better with diseases which are primarily characterized by low-level attributes (e.g., the results given by laboratory tests) rather than by those expressed through a patient's description of aches, pains, and other discomforts. If this should turn out to be the case, the program will have performed more as a clinical pathologist than as an internist.

Given the practical ends of medicine, it makes little difference how a physician goes about making a diagnosis (subject to ethical and economic considerations), so long as an appropriate one is reached. And if INTERNIST (or CADUCEUS) should turn out to perform better when dealing with low-level findings than with high-level ones, this would not detract from its achievement. Physicians themselves are increasingly finding this to be the case in their own work, which is why they depend so heavily on the clinical laboratory and the X-ray department. If CADUCEUS should behave in a similar manner, then its goal of simulation would have indeed been even more closely realized.

WHERE DOES DIAGNOSIS BEGIN?

There is a further kind of diagnosis which deserves comment: the direct recognition of disease. In mentioning both the Bayesian and the abductive models of diagnosis, we have stated that the information given by the clinical findings is assumed to be available at the outset. This has occasionally led to the notion that diagnosis is a two step process: a physician first collects data, and only when this process is completed, reflects upon it (or turns it over to a computer) for the purpose of inferring a diagnosis. Bayesian programs usually work this way, but such a notion would distort what seems to go on in a physician's mind. From the time a physician first sees a patient or hears a patient's complaints, there is probably not a moment when he does not have one or more diagnoses in mind (Elstein, 1978). Moreover, it is just this entertaining of possibilities which guides the physician's questioning and examination. Observation is said to be "theory laden": we observe particular things in a situation because we have theories which impute relevance to some of them and not to others (Popper, 1972). A purely neutral or indifferent collection of neutral clinical facts would lead to little.

Hypothesis-making and clinical observation appear to proceed together. These processes are also iterative. Before the desired data have all been obtained, many diseases will have been thought of and discarded, and some perhaps entertained again. As the clinical data are being collected, they are also expressed in medically significant linguistic forms. A physician doesn't report that he heard a "funny noise" in the chest, but that he heard a "diastolic murmur over the aortic valve". When he does this, he is doing far more than reporting a primitive sense-datum. When raw sensory impressions are refined and expressed in such specific terms, the diagnostic process is already well advanced. And when sets of such carefully expressed observations are then passed on to a diagnostic program, it can well be argued that much of the diagnostic process has been completed. It is for this reason that we should probably speak of "computer-aided" diagnosis rather than computer diagnosis.

Then too, physicians do not seem (consciously at least) to operate in particular diagnostic modes, although all present computer programs do. Physicians do not make a mental commitment that they will think about a particular case in, say, a "Bayesian" way, and then later consciously shift to a pattern matching mode. Instead they appear to dip into their bag of cognitive tricks, and use whatever strategy seems most appropriate in a particular situation at a given moment, whether this be Occam's Razor, "Sutton's Law", Bayes' Theorem or something entirely different. The rich and varied repertoire used in medical inference has not yet been decomposed into sets of formal procedures, nor has it been shown to be decomposable in such a manner.

None of this is meant to suggest that computer programs have not or may not be usefully employed in certain aspects of the diagnostic process. It is not certain, however, that the most useful applications have been identified, or that a suitable allocation of labor has been made between man and machine. Even more important is the proper assignment of responsibility, a question we will not consider here. Nor is it clear that a strategy based upon the attempt to simulate a physician's diagnostic methods is the best one. There is, however, considerable evidence that computer support in diagnosis would be well received by physicians, if this can be shown to result in more accurate diagnoses.[9] It seems possible that diagnostic programs having fairly limited aims, such as "prompting", but which are broadly based (i.e., having an extensive "knowledge base" such as programs which generate useful differential diagnoses when provided with clinical laboratory data), might be amongst the first to find acceptance. It also appears probable that the computer processing of low-level attributes will prove more practical than with high-level ones, since it is the former which are more readily formalized.

Nevertheless, it is the high-level attributes of diseases that bother patients. These are not only the most difficult to formalize, but in the case of psychiatric disorders, diagnosing a disease may in a sense require understanding what it is like to have it. It is these distinctly human features of diseases which raise the greatest difficulties for the formalist, and for which computer-aided diagnosis appears (at least for the present) to have the least to offer.

ACKNOWLEDGMENT

I am indebted to Azad Balour, Dana Ludwig, and Mark Tuttle for their comments on an earlier draft of this paper, to Ernan McMullin for discussions of several points, and to Chris Cherniak for a number of helpful suggestions. This research was supported in part by a grant from the National Library of Medicine (5T 15 LM 07000–06).

NOTES

1 The extension of this multilevel ontology downward to the molecular and atomic levels (and upward too, for that matter), would appear a reasonable move given a theory of decomposability, and in the overall interests of parsimony or "unity".

2 Evolutionary theories of biology and cosmological theories attempt to provide general "vertical" explanations extending across many levels. Schaffner has proposed that the Jacob-Monod theory of gene repression and derepression is such a vertical theory. (Schaffner, 1980).

3 Not all of these need be expressed as properties of the higher level object. Scalar properties may be (a person's mass is nothing else than the scalar sum of the atomic masses which comprise his body). But the momentum of a tangible object at rest is zero, because the vector sum of the atomic momenta comprising it is zero, and not because their atomic momenta are all individually zero. It is the same with charge, magnetic moment, and other properties which may be internally compensated, and thus at a higher level, latent. But these internal balances need not be permanent; the electrically neutral atoms and molecules of our bodies may become ionized, and must, if cell membranes are to function

and nerves to work. When a molecular radical displays a magnetic moment, it is not because this property has emerged at this level. Magnetic moment was latent there all the while and then becomes expressed. The actualization of these atomic properties results in local rather than global effects (patients do not become 'magnetized'), and the results are no more than would be predicted by the laws of physics and chemistry.

[4] Emergent properties are usually thought of as arising at high-levels of organization, such as in living organisms, and to be exemplified by such properties as "living" or "conscious". (Medawar, 1977) It appears, however, that they arise at all upward shifts of hierarchical levels, beginning with the very lowest. When we say that a particular whole is greater than the sum of its parts, this vague claim can mean no more than that the bringing of parts together to give rise to the whole calls into being emergent properties which were not displayed by any of the parts alone.

[5] Low-level objects are frequently classified according to a "monothetic" procedure, one in which a single property may suffice for the construction of a classification scheme. The chemical elements (and their isotopes) are so classified by using their atomic number. In the classification of organisms, a "polythetic" method has been claimed to be necessary (Sokal, 1974) where all their attributes must be used (and are usually weighted equally). This procedure may be resorted to because it does not require us to distinguish between necessity and contingency.

[6] Some of the confusion over "objective" and "subjective" clinical findings (and their "reality" and explanatory power) would seem to disappear when these terms are translated as "low-level" and "high-level" findings. An analysis along similar lines would seem helpful in examining the evolution of technology in medicine, by way of amplifying the strictly historical accounts, such as that of Reiser (Reiser, 1978).

[7] Hilary Putnam (Putnam, 1973), in taking issue with the reductionists, and arguing that we cannot explain high-level behavior on the basis of low-level descriptions, remarked that reductionism breeds " . . . physics worship coupled with neglect of the 'higher level' sciences".

[8] A report of an evaluation of INTERNIST (Miller *et al.*, 1982) appeared while this paper was in press.

[9] Shortliffe has suggested that physicians' acceptance of computer-proffered advice will be greater if the program is able to recount, in a comprehensible way, the nature of the inference involved. (Shortliffe, 1976). A human consultant's opinion will be accepted and acted upon, not because he is an expert, but because he has succeeded in justifying it. It is for this reason that the recommendations delivered by purely statistical programs may be viewed with skepticism. The connections between the input and output of the program are not explicit, and while a physician may be persuaded of the accuracy of a diagnostic program's *average performance*, he may well have reservations in a particular case.

REFERENCES

Bleich, H. L.: 1969, 'Computer evaluation of acid-base disorders', *Journal Clinical Investigation* 48, 1689–1696.

Blois, M. S.: 1980, 'Clinical judgment and computers', *New England Journal of Medicine* 303, 192–197.

Blois, M. S. *et al.*: 1981, 'RECONSIDER: A program for generating differential diagnoses', in *Proc. of the Fifth Annual Symposium on Computer Applications in Medical Care, Washington, D.C., Nov. 1981*, IEEE, New York.

Boden, M. A.: 1977, *Artificial Intelligence and Natural Man*, Basic Books, New York, N.Y.

Bronowski, J.: 1970, 'New concepts in the evolution of complexity', in J. Bronowski (ed.), *A Sense of the Future*, M.I.T. Press, Cambridge, Mass.

Buchanan, B. C. and E. A. Feigenbaum: 1978, 'Dendral and metadendral: their applications dimension', *Artificial Intelligence* 11, 5–24.

Crooks, J. *et al.*: 1959, 'Statistical methods applied to the clinical diagnosis of thyrotoxicosis', *Quarterly Journal of Medicine* 28, 211–234.

DeDombal, F. T. *et al.*: 1972, 'Computer-aided diagnosis of acute abdominal pain', *British Medical Journal* 2, 9–13.

Dreyfus, H. L.: 1979, *What Computers Can't Do: The Limits of Artificial Intelligence*, revised ed., Harper and Row, New York.

Eddington, A.: 1964, *The Nature of the Physical World*, Everyman's, London.

Elstein, A. S. *et al.*: 1978, *Medical Problem Solving: An Analysis of Medical Reasoning*, Harvard Univ. Press, Cambridge, Massachusetts.

Fitzgerald, L. T. *et al.*: 1966, 'A Computer program for diagnosis of thyroid disease', *American Journal of Roentgenology, Radium Therapy, and Nuclear Medicine* 97, 901–905.

Grene, M.: 1967, 'Biology and the levels of reality', *The New Scholasticism* 41, 94–123.

Lindberg, D. A. B. *et al.*: 1968, 'CONSIDER: A computer program for medical instruction', in *9th IBM* Medical Symposium.

McMullin, E.: 1983, 'Diagnosis by Computer', this issue, pp. 5–27.

Medawar, P.: 1977, 'Emergence', in *The Fontana Dictionary of Modern Thought*, Fontana/Collins, London.

Meehl, P. E.: 1954, *Clinical versus Statistical Prediction; a Theoretical Analysis and a Review of the Evidence*, University of Minnesota Press, Minneapolis.

Meehl, P. E.: 1973, *Psychodiagnosis*, University of Minnesota Press, Minneapolis.

Miller, R. A. *et al.*: 1982, 'INTERNIST-I, an experimental computer-based diagnostic consultant for general internal medicine', *New England Journal of Medicine* 307, 468–476.

Nagel, E.: 1961, *The Structure of Science, Problems in the Logic of Scientific Explanation*, Routledge and Kegan Paul, London.

Polanyi, M.: 1958, *Personal Knowledge*, Routledge and Kegan Paul, London.

Pople, H. E.: 1977, 'The formation of composite hypotheses in diagnostic problem solving: an exercise in synthetic reasoning', in *Proceedings of the 5th International Joint Conference on Artificial Intelligence*, Cambridge, Massachusetts.

Popper, K. R.:1972, *Objective Knowledge*, Clarendon Press, Oxford.

Putnam, H.: *Cognition* 2, 13–146.

Reiser, S. J.: 1978, *Medicine and the Reign of Technology*, Cambridge University Press, New York.

Scadding, D.: 1967, 'Diagnosis: The clinician and the computer', *Lancet* 876.

Schaffner, K. F.: 1980, 'Theory structure in the biomedical sciences', *Journal of Medicine and Philosophy* 5, 57–97.

Shortliffe, E. H.: 1976, *Computer-Based Medical Consultations: MYCIN*, Elsevier, New York.

Shortliffe, E. H. *et al.*: 1979, 'Knowledge engineering for medical decision making: A review of computer-based clinical decision aids', *Proceedings of the IEEE* 67, 1207–1224.

Simon, H. E.: 1962, 'The architecture of complexity', *Proceedings American Philosophical Society* 106, 467–482. (Reprinted in H. E. Simon: 1981, *The Sciences of the Artificial*, M.I. T. Press, Cambridge, Mass.)

Sokal, R. R.: 1974, 'Classification: Purposes, principles, progress, and prospects', *Science* 185, 1115–1123.

Warner, H. *et al.*: 1964, 'Experience with Bayes Theorem for computer diagnosis of congenital heart disease', *Annals New York Academy of Science* 115, 558–567.

Whitehead, A. N.: 1938, *Modes of Thought*, MacMillan, New York.

Zadeh, L. A.: 1976, 'The concept of a linguistic variable and its application to approximate reasoning', *Information Science* 9, 43–80.