Causal Representation of Patient Illness for Electrolyte and Acid-Base Diagnosis

by

Ramesh S. Patil

October 1981

This research was supported (in part) by the National Institutes of Health Grant No. 1 P01 LM 03374-0x from the National Library of Medicine.

Laboratory for Computer Science

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Carnbridge Massachusetts 02139

Causal Representation of Patient Illness for Electrolyte and Acid-Base Diagnosis

by

Ramesh S. Patil

Submitted to the Department of Electrical Engineering and Computer Science on October 24, 1981 in partial fulfillment of the requirements for the Degree of Doctor of Philosophy

Abstract

Much of the medical knowledge in the first generation *AI in Medicine* programs is phenomenological; that is, it describes the associations among phenomena without knowledge of the underlying causal mechanisms. Although these AIM programs provide a good first approximation to the way clinicians reason, they fail to reproduce clinicians' reasoning based on a deeper understanding of the phenomena. More specifically, they do not deal with the knowledge of disease at different levels of detail, nor do they utilize causal relations to organize and explain the clinical facts and disease hypotheses. They also cannot deal with illnesses resulting from multiple diseases, especially when one disease alters the presentation of the others. Finally, they are unable to capture the notions of adequacy and parsimony that play such a large role in diagnosis. To explore these issues and rectify these deficiencies, we have undertaken the task of providing expert consultation for electrolyte and acid-base disturbances.

This thesis reports the implementation of ABEL, the diagnostic component of the consultation program. In it, we explore the problems of modeling the causal understanding of a patient's illness. We develop techniques for dealing with illness resulting from multiple interacting diseases. We describe a multi-level representation of causal knowledge, and explore issues of the aggregation of available case specific knowledge into concise summaries of the patient's illness. We discuss structural criteria for evaluating parsimony, coherence and adequacy of diagnostic explanations. We also explore some of the issues involved in information gathering and propose expectation-driven diagnostic planning as a means of improving it. Finally, we discuss the issues of explanation and justification of the program's understanding and argue that these facilities are crucial for acceptability of a consultation program.

Thesis supervisor: Peter Szolovits
Associate Professor of Electrical Engineering and Computer Science

Keywords: Causal Representation, Medical Diagnosis

CONTENTS

FIGURES

Acknowledgments

I would like to express my thanks to all of the people who made this thesis possible:

Peter Szolovits, my thesis supervisor, for providing constant attention, guidance and for helping me formalize ideas when I could not see through the confusion;

William B. Schwartz for suggesting this thesis topic, teaching me about acid-base and electrolyte disturbances, and for articulating and refining many of the ideas presented in this thesis;

Randall Davis, for his helpful comments and suggestions on drafts of this document;

William Martin for introducing me to the area of knowledge based application systems and providing the intellectual environment in which these ideas germinated;

Lowell Hawkinson for developing XLMS and providing help in using it;

Bill Long for his encouragement and for being a very insightful sounding-board for ideas;

Glenn Burke for proofreading this document and for providing much needed system support;

Ken Church for many spirited arguments and discussions on this and other topics;

Bill Swartout for timely development of explanation methodology and for sharing his experience in using XLMS;

Stephen Pauker, Brian Smith, Ben Kuipers, Byron Davies, Howard Sherman, Harold Goldberger, Brij Masand, Gretchen Brown and other past and present members of Clinical Decision-Making group and Knowledge-Base System groups for providing the fertile environment and comradery needed to carry me through the thesis;

and finally, my wife Aruna and my family for bearing with me and providing constant encouragement without which this would not have been possible.

1. Introduction

In a 1970 article reviewing the role of emerging computer technology in medicine, Dr. William B. Schwartz notes

> *"If conventional remedies will not meet the demands imposed by society's broad commitment to extensions of health care, it is clear that new, even heretical, strategies must be devised. One such strategy will almost certainly involve exploitation of the computer as an "intellectual," "deductive" instrument — a consultant that is built into the very structure of the medical-care system and that augments or replaces many traditional activities of the physician. Already, several interesting steps have been taken in an attempt to extend the computer's role into this realm ... Indeed, it seems probable that in the not too distant future the physician and the computer will engage in frequent dialogue, the computer continuously taking note of history, physical findings, laboratory data, and the like, alerting the physician to the most probable diagnoses and suggesting the appropriate, safest course of action. One may hope that the computer, well equipped to store large volumes of information and ingeniously programmed to assist in decision making, will help free the physician to concentrate on the tasks that are uniquely human such as the application of bedside skills, the management of the emotional aspects of diseases, and the exercise of good judgment in the nonquantifiable areas of clinical care."*
>
> — Medicine and the Computer [Schwartz70, page 3]

The decade following these predictions saw a rapid growth in the field of *Artificial Intelligence in Medicine* (AIM) culminating in many promising programs, among which are Internist-I [Pople77], the Present Illness Program (PIP) [Pauker76], CASNET/Glaucoma [Weiss74], MYCIN [Shortliffe76] and Digitalis Therapy Advisor [Pauker76]. These programs represent the first efforts in the use of AI techniques in medical decision making, and can be characterized as the "first generation AIM programs". They have clearly demonstrated the feasibility and usefulness of AI techniques. Most of these programs have, in some trial, been judged to match expert physicians in their competence — this is indeed an outstanding achievement.

It is natural to question then: "What are the limits of their expertise? Why aren't we implementing these programs in many more areas of medicine and distributing them for clinical use?" To answer these questions we must take a deeper look at the programs and their performance. For example, although they are (on average) outstanding on their core set of anticipated applications, their performance can also be non-uniform; it tends to degrade rather ungracefully just outside their domain of expertise. Furthermore, these programs may be misled on difficult cases involving complex interactions or multiple disorders, even if these cases fall well within their domain of expertise. This leads to the ineyitable conclusion that although the models of representation and deduction used in these programs are capable of providing moderate coverage over the area of application, they are nonetheless inadequate.

These observations have led to a re-evaluation of the techniques used in the first generation of AIM programs. The following insights have been gained by this evaluation. Firstly, the notion of causality is inadequately exploited in the first generation AIM programs [Smith79, Patil79, Pople81]. They do not utilize the structure provided by causal relations to organize the patient facts and disease hypotheses. They fail to capture the human notion that explanation should rest on a chain of cause-effect deduction. Secondly, they cannot deal with the effects of more than one disease present in a patient simultaneously, especially when one of the diseases alters the presentation of the others. Thirdly, they do not deal with the knowledge of a disease phenomenon at different levels of detail that a physician clearly has. Finally, the numeric belief measures as used by the first generation AIM programs do not provide adequate criteria for diagnostic reasoning. They are unable to capture notions such as adequacy and parsimony of a diagnostic possibility.

Much of the medical knowledge contained in the first generation AIM programs can be characterized as being phenomenological; that is, it describes the associations among phenomena without the mechanisms underlying the observed associations. Such phenomenological descriptions provide a good first approximation to the way physicians reason, but they fail to capture the physicians' reasoning in recognizing and dealing with the inherent discrepancies in their knowledge and with deduction based on deeper understanding of the phenomena. Contrasting the behavior of the first generation AIM programs and human experts, Szolovits notes:

> *"Consider what happens when two "rules of thumb" (as we may identify a bit of phenomenological knowledge in medicine) conflict. Every AIM program written so far evaluates that conflict by reducing it to a numerical judgment of likelihood (or certainty, belief, etc.) in the hypotheses it holds: Mycin computes a revised certainty factor, CASNET computes new weights, Internist computes new scores, and the digitalis program often computes a weighted sum of its observations to evaluate their joint effect. Thus, conflict, just as agreement, is reduced to a manipulation of strength of belief. Yet, by contrast, we believe that human experts make a much more powerful use of occasions where they detect conflict. They are not satisfied by a simple revision of their degree of belief in the hypotheses which they have previously held; they seek a deeper, more detailed understanding of the causes of the conflict they have detected. For it is just at such times of conflicting information that interesting new facets of the problem are visible. Conflicts provide the occasion for contemplating a needed re-interpretation of previously-accepted data, the addition of possible new disorders to the set of hypotheses under consideration, and the reformulation of hypotheses thus far loosely held into a more satisfying, cohesive whole. Much of human experts' ability to do these things depends on their knowledge of the domain in greater depth than what is typically needed to interpret simple cases not involving conflict."*
>
> —Artificial Intelligence and Medicine [Szolovits81a, pages 16-17]

To move beyond the sometimes fragile nature of today's programs, we believe that future AIM

programs must contain medical knowledge similar in depth of detail to that used by expert physicians. They must have anatomical, physiological and pathophysiological knowledge sufficiently inclusive in both breadth and detail to allow the expression of any knowledge or hypothesis that usefully arises in medical reasoning.

One of the important areas of medical diagnosis not adequately addressed by the first generation of AIM programs is the evaluation of the effect of more than one disease present in the patient simultaneously, especially when one of the diseases alters the presentation of the others. For example, let us consider a patient with diarrhea and vomiting leading to severe hypokalemia. Let us also suppose that we know about the diarrhea, but we are not aware of the vomiting. The observed hypokalemia is too severe to be properly accounted for by the diarrhea alone and therefore diarrhea cannot be considered as complete explanation for the observed hypokalemia. Given this fact, the diarrhea is either not responsible for hypokalemia or is only partly responsible. If the diarrhea is not responsible, then further reasoning is relatively easy: the problem simplifies to finding the actual cause. However, if diarrhea is partly responsible, a correct partitioning of the total observed hypokalemia between its two suspected causes is required, with a judgment of how well the two separate causes combined in the estimated proportions account for the patient's condition.[1] Notice how inadequate the simple assignment of a probability linking diarrhea and hypokalemia (as is commonly done in existing programs) is to capture the problem being described here.

The complexity and depth of medical knowledge is well recognized [Szolovits78]. Our understanding of medical expert reasoning suggests that an expert physician may have an understanding of a difficult case in terms of several levels of detail. At the shallowest level that understanding may be in terms of commonly occurring associations of syndromes and diseases, whereas at the deepest it may include a biochemical and pathophysiological interaction of abnormal findings. While it may be easier for a program to reason succinctly with medical knowledge artificially represented at a uniform level of detail,[2] a range of representations are needed to reason at a sophisticated level of competence [Patil81]. Unfortunately, very little attention has been paid to developing methods for coping with it. We take this as the central issue of this thesis.

---

1. All the previous programs allow the entire hypokalemia to be accounted for by diarrhea. In particular, Internist-I after allowing the hypokalemia to be accounted for by diarrhea will not allow hypokalemia to lend any support to the hypothesis of vomiting. PIP, on the other hand, will allow the entire hypokalemia to lend support to the hypothesis of vomiting as well as allowing it to be explained by diarrhea.
2. This does not pose serious difficulty in medical domains where the pathophysiology of diseases is not well developed, because in such a domain a physician relies primarily on his phenomenological knowledge. However, in a domain such as electrolyte and acid-base disturbances we are constantly faced with this problem because, on the one hand, the pathophysiology of the disturbances is well developed, and on the other, the pathophysiology of many of the diseases leading to these disturbances is relatively poorly understood.

Finally, we believe that the numerical (probabilistic or pseudo-probabilistic) belief measures as used by the first generation AIM programs for confirming diagnoses and guiding the diagnostic search do not provide adequate criteria for diagnostic reasoning. We believe that the evaluation methods for confirming a disease hypothesis should be different from the methods used for choosing the most promising disease hypothesis for diagnostic pursuit. A single criterion is almost certain to be inadequate for both these tasks. Furthermore, we believe that the probabilistic model by itself is inherently inadequate. For example, it fails to take into account the causal nature of the disease mechanisms, it fails to capture the notions of parsimony, coherence and adequacy of diagnostic explanation. In a study of problem solving activity of clinicians, Kassirer and Gorry note that

> *"In parallel with the processes by which the physicians built a case toward a final diagnosis, they assessed each diagnosis for coherence and adequacy. ...* (A diagnosis was considered *coherent* if all the symptoms and diseases contained in it were causally related to each other. A diagnosis was considered *adequate* when it accounted for all all known facts.) *... The physicians strove to attain parsimonious explanations for the findings and to accept a single explanation rather than make two or more diagnoses unless they were forced to do so."*
>
> — Clinical Problem Solving [Kassirer78, pages 249-250]

It is one of the central themes of this thesis that these problems cannot be avoided by relying solely on the numerical scoring mechanism; the programs must be provided with structural criteria to evaluate the disease hypotheses.

It is our belief that modeling the program's understanding of the patient's illness is crucial to capturing the expertise of clinicians. In this thesis, we will explore some of the issues involved in representing diagnosis. We will develop techniques for reconciling physiological reasoning with phenomenological reasoning and explore issues of aggregating all the available knowledge into concise summaries of the patient's illness. We will discuss structural criteria for evaluating parsimony, coherence and adequacy of diagnostic explanations. We will also explore some of the issues involved in information gathering and propose expectation-driven diagnostic planning as a means of improving it. Finally, we will discuss the issues relating to explanation and justification of the program's understanding.

To study these issues, we have chosen the task of providing expert consultation in cases of electrolyte and acid-base disturbances. The research presented in this thesis, the development of a program called ABEL (Acid-Base and ELectrolyte program), is a part of this overall effort. We describe a novel mechanism for representing ABEL's understanding of a patient's illness. This understanding is represented using a collection of data-structures called the *patient-specific models*(PSMs). Each PSM contains a hypothesis structure containing all known data about the patient, all currently held possible interpretations of these data, the causal interconnections among the known data and tenable hypotheses, and some indication of alternative interpretations

and their relevant evaluations. We describe the representation of medical knowledge and the processing strategies needed to enable ABEL to construct a PSM from the initial data presented to the program. The same representations and procedures are also used in revising the PSM during the process of diagnosis. Each PSM can be viewed as a partial explanation of the patient's illness.

Diagnostic problems are formulated by identifying the weaknesses and conflicts in the PSMs and by computing a *diagnostic closure* (DC) for each PSM. A DC associated with a PSM represents a collection of alternative completions of the partial explanation provided by the PSM. It brings together all the dependencies and expectations necessary for diagnostic inquiry, for evaluating real and apparent discrepancies in the incoming information, and for explaining the diagnostic alternatives under consideration. A plan for diagnostic inquiry is generated by decomposing a top level diagnostic problem into simple problems which can be directly solved by a question to the user. Finally, when an inquiry is completed, the new information gathered is assimilated into the PSMs and the diagnostic process is repeated.

1.1 Scope of Project

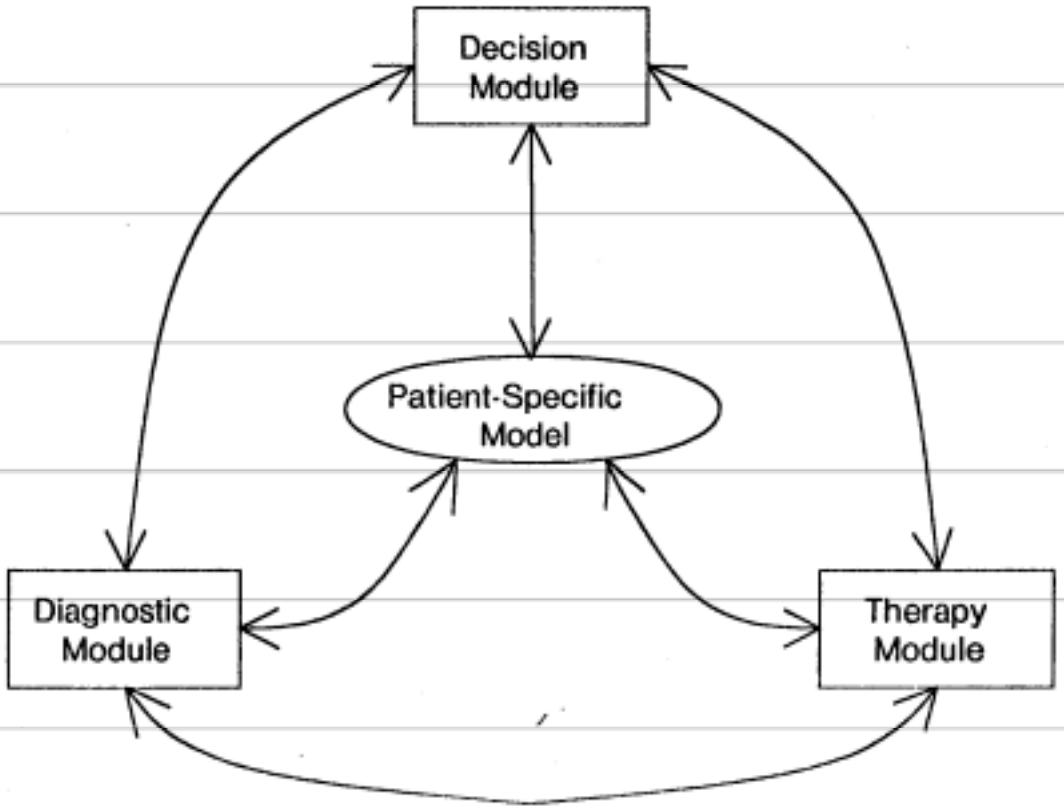
This thesis has three main objectives. The first is to develop a representation of causal medical knowledge. The second is to develop a case-specific "understanding" of illness. This understanding should be capable of describing subtle interactions between diseased and normal physiological mechanisms, and therapeutic interventions. The third is to develop a set of reasoning procedures to combine the aggregate phenomenological knowledge of disease associations with the detailed pathophysiological knowledge of disease processes. The first of these, the phenomenological knowledge, is necessary for efficient diagnostic exploration; the second, the pathophysiological knowledge, is necessary for proper understanding of a difficult case. The research reported in this thesis is conducted in the larger context of an Expert Consultant for Electrolyte and Acid-Base Disturbances [Patil79]. This section briefly reviews the organization of the overall system.

The objective of an expert medical consultant is to advise in the *proper management* of a patient. Proper management consists of collecting the relevant information about the patient, identifying the disease process(es) responsible for the patient's illness, and prescribing a proper course of action to correct the patient's condition. One of the complexities of this task is due to the fact that these subtasks do not have well defined boundaries. The patient may be presented to a clinician at different stages of a disease's evolution and treatment. During the course of management new information about the past history may become necessary as the diagnostic hypotheses evolve. The current diagnosis may depend on information that is presently unavailable. The disease itself may evolve through time, providing additional clues to its identity, or the response to certain therapeutic interventions may provide valuable diagnostic information. Finally, the patient's condition may require therapeutic intervention even before the diagnostic

issues can be reasonably resolved. Therefore, the next course of action must be chosen from a large range of alternatives. These alternatives may be broadly classified as *gathering information* (much of which may turn out to be irrelevant in the evolving clinical context), *ordering tests* (possibly involving expensive time delays and/or clinical costs), *waiting* for further development, *prescribing therapy* or some combination of the above. At every stage of consultation, the program must be able to choose between the alternative sets of actions with the patient's best interest in view. This can be achieved only by developing a program capable of forming a diagnosis, suggesting a therapy and making decisions. With this perspective we have embarked on the design of the Electrolyte and Acid-Base Consultant system. We have tried to separate and modularize different components of a physician's knowledge and expertise so as to be able to evaluate our understanding about each component and their interactions. This modularization should also allow us to further experiment with any component of the system without having to reimplement the entire program. A top level schematic for the overall system is shown in figure 1.

The Electrolyte and Acid-Base Consultant system consists of four major components: (1) the Global Decision Making component, (2) the Diagnosis component, (3) the Therapy component and (4) the Patient Specific Model. The patient specific model describes the physician's

Fig. 1. A schematic for the overall system



understanding of the state of the patient at any point during diagnosis and management; it is intended to be the central data structure which other components of the system may reason with. The global decision making component is the top level program which has the responsibility of calling the other programs with specific tasks. In general, the global decision program will call the diagnostic program with a task such as taking the initial history and elaborating some specific diagnosis. The diagnostic component then performs the specified task and reports the results to the main program. It also modifies the patient specific model to reflect the revised state of the patient. Similarly, if the global decision making program calls the therapy selection program, it attempts to formulate a set of alternate therapies for the patient along with a check list of items that must be tested before any specific therapy can be recommended. It also identifies information that will help discriminate between alternate therapy recommendations. Note that at every step the global decision maker can evaluate each of the possible sets of actions and choose the most desirable one. The decision making component will allow the program to make explicit the decision making that goes on in a physician's reasoning: is further diagnosis necessary, what treatment should be selected, should he wait before prescribing further treatment, can he choose some therapeutic action that would also provide diagnostic information making further diagnosis at this point unnecessary?

This thesis deals primarily with the development of the patient specific model which describes the program's understanding about the patient's illness. We have focused here because we believe that the level of expertise achievable by the program is inherently dependent upon the expressive capabilities of the patient specific model. The program can reason about subtle interactions between diseases in a given patient only if it can describe these interactions in the context of the patient. In addition a preliminary implementation of the diagnostic component to demonstrate the use of this patient-specific model is also discussed.

1.2 Choice of Domain

Careful selection of a domain is crucial for developing an application program. The domain chosen must be small enough to allow one to build a knowledge-base in a reasonable amount of time, and yet large enough to allow for realistic testing of the new ideas being implemented. Furthermore, the domain should be well defined and should lead to useful application, so that the program can be field-tested under realistic conditions. We have chosen the domain of electrolyte and acid-base disturbances as the test-bed for our theories of medical diagnosis.

The domain of electrolyte and acid-base disturbances is a well defined and relatively narrow area of medicine. It is an ideal domain for testing our theories about interactions between causal (physiological) reasoning and phenomenological (syndromic) reasoning, as on one hand the basic pathophysiology of the acid-base disturbances is well developed, and on the other, the pathophysiology of the diseases leading to these disturbances is relatively poorly understood. Thus constantly forcing us to develop reasoning mechanisms that can deal simultaneously with

well understood causal knowledge and poorly understood phenomenological knowledge. In addition, the feed-back nature of the electrolyte and acid-base homeostatic mechanism provides us, in a microcosm, with a variety of issues relating to "dynamic" systems that must be addressed in the management of a patient's illness.

Electrolyte and acid-base disturbances are a common complication of a large number of serious illnesses and medical interventions. In spite of their prevalence, this remains an area that most practicing physicians find somewhat difficult to deal with. This makes the field of acid-base disturbances an attractive domain for introducing expert computer consultant programs. One of the earliest programs for medical consultation [Bleich72] was in fact introduced in this very area.

Our primary concern, however is not with electrolyte and acid-base disturbances per se. Our basic purpose is to use this domain as a vehicle for evaluation of the existing techniques and development of new techniques for diagnosis and management of a patient's illness. In particular, in this thesis we will develop techniques for providing a coherent account of a patient's illness which incorporates the pathophysiological understanding of acid-base disturbances with the aggregate phenomenological understanding of the diseases causing these disturbances.

1.3 Brief review of Electrolyte and Acid-Base Disorders

In this section we briefly describe the electrolyte and acid-base disturbances. This section is not intended as a full review of the subject matter, but is presented here to provide the readers with a framework for understanding the medical examples used in this document. Each example used in the document is accompanied by an explanation of the relevant medical knowledge.

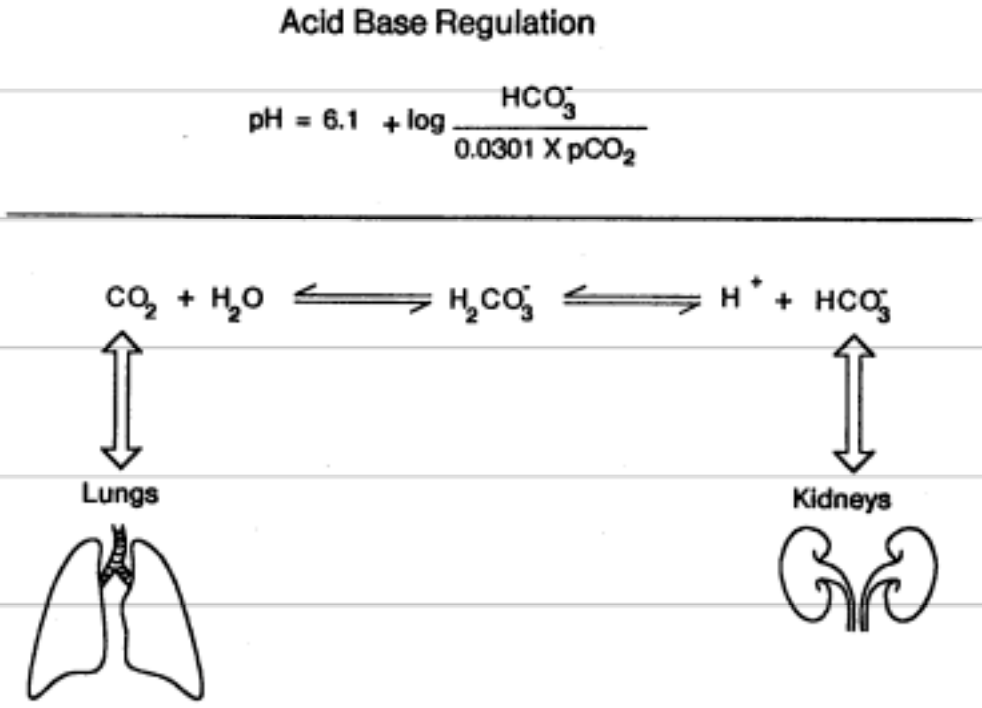
Fluid and electrolyte disturbances usually occur as complications of an underlying illness, therefore these disorders must be viewed not as isolated entities but in the context of the specific clinical settings in which they appear. As general background to the following discussion, it should be remembered that approximately 50 to 60 per cent of the body (by weight) consists of water distributed between the intracellular (within cells) and extracellular (outside cells) compartments. Water moves freely across cell boundaries, maintaining osmotic equilibrium between the different compartments. By contrast, owing to differences in their permeablity and active ionic pumps, the electrolytes are distributed in an asymmetric pattern, most of the ions in extracellular fluid consisting of sodium, chloride and bicarbonate and those in intracellular fluid of potassium and organic anions. Regulation of the external environment of cells, that is, the electrolyte concentration and acidity (pH) of the body fluids, is of primary importance. Perturbations in the regulation of this environment is the subject of electrolyte and acid-base disturbances.

The pH of the body fluids is regulated by three mechanisms: (1) the body buffers, (2) pulmonary regulation of the concentration of CO_2 in the body, and (3) renal excretion of acids and alkali. They act in a complementary fashion, first to minimize transient changes and then to correct any disturbances in acid-base balance by appropriate retention or excretion of hydrogen ions. To understand the mechanism of acid-base disturbances, it is instructive to consider the way in which the body deals with the normal daily acid load in maintaining a steady-state of acid-base equilibrium.

As food is oxidized to provide metabolic energy, both carbon dioxide (carbonic acid) and acids such as sulfuric and phosphoric acids are added to the extracellular fluid. They are immediately buffered to minimize the change in pH and transferred to the lungs and kidneys for excretion. Carbon dioxide is excreted almost entirely by the lungs while the other acids are excreted solely by the kidney. Bicarbonate is regenerated by the kidney as it excretes the excess acid, replenishing the bicarbonate stores that previously were depleted by the buffering of the dietary acid. From all these considerations it is evident that derangements in either the pulmonary or renal function, or the imposition of stresses that overwhelm normal regulatory mechanisms (such as vomiting, diarrhea, burns, etc.) can be expected to produce disturbances of acid-base equilibrium.

The equilibrium equation of the major buffer system in the extracellular fluid, the *carbonic acid — bicarbonate buffer system*, is shown in figure 2. This equation allows ready visualization of the directional changes that can be anticipated in both metabolic and respiratory disturbances of the acid-base equilibrium. For example, a primary reduction in bicarbonate concentration (*metabolic acidosis*) will cause the reaction to shift to the right, thus increasing hydrogen ion concentration, whereas a primary elevation in bicarbonate concentration (*metabolic alkalosis*) will cause the reaction to shift to the left, thus decreasing hydrogen ion concentration. Similarly, a primary rise in pCO_2 increases the hydrogen ion concentration (*respiratory acidosis*), and a fall has the reverse effect (*respiratory alkalosis*). However, the presentation of these disturbances is somewhat more complicated owing to the fact that the body reacts to these changes and attempts to compensate (in part) for the effect of these changes. Furthermore, different compensating mechanisms respond at different rates. A disturbance which has been properly compensated is called *compensated*, otherwise it is called *uncompensated*. The actual changes in the bicarbonate — carbonic acid concentrations in these disturbances is shown in figure 3. The nomogram of acid-base disturbances [Schwartz65, Cohen66] shown in figure 3 summarizes the normal physiologic response to the changes in HCO_3 and pCO_2 for each of the acid-base disturbances described above. For example, the nomogram shows that for a patient with adequately compensated metabolic acidosis and with serum concentration of HCO_3 of 15 meq/L the pCO_2 will be approximately 30 mmHg. The use of this nomogram for initial evaluation of a patient's acid-base state will be discussed later.

Fig. 2. Carbonic acid - bicarbonate buffer equation



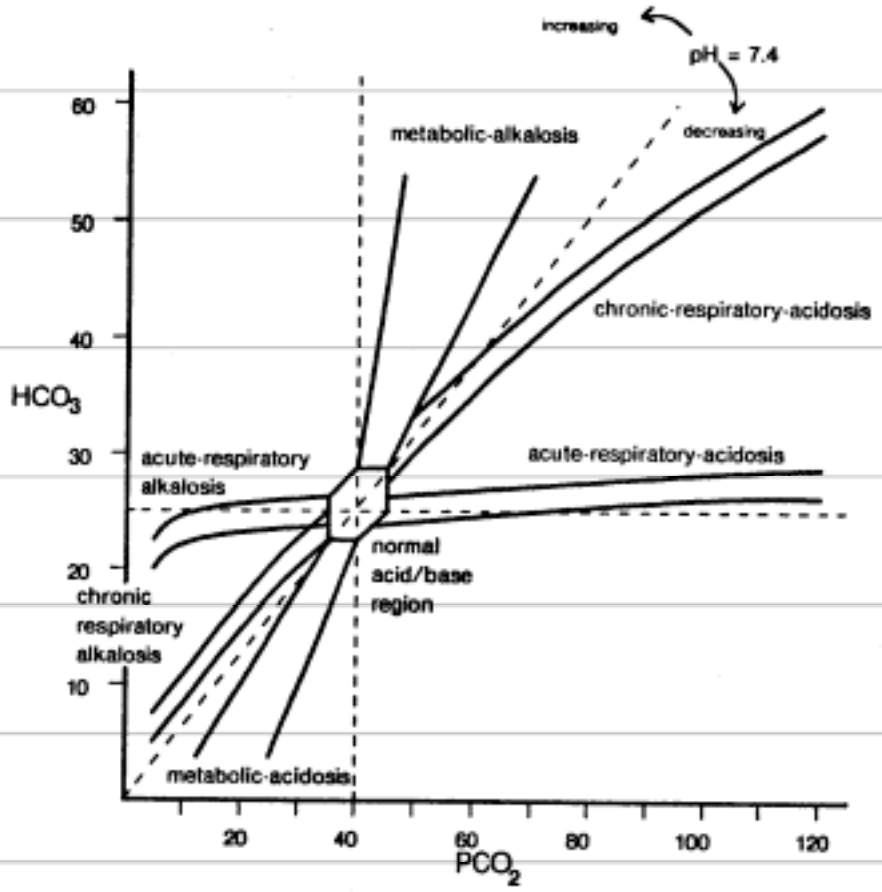
Regulation of Carbonic Acid / Bicarbonate buffer pair

Henderson - Hasselbalch Equation

The most frequently encountered clinical acid-base disorders occur as *single disorders* (also called simple disorders). The single disorders are: metabolic acidosis, metabolic alkalosis, respiratory acidosis, and respiratory alkalosis. There are, however, many clinical situations in which combinations of two or three disorders occur simultaneously, giving rise to *mixed disorders*. The recognition of mixed disorders is predicated upon a clear understanding of the pathophysiologic effects of simple disorders. To diagnose mixed disorders, one must know how each of the four simple disorders named above alter pH, pCO_2 and HCO_3 and the extent of renal or respiratory compensation that ought to occur for any given degree of primary disorder. However, since each of the disturbances can be caused by a variety of physiological states or diseases, the final differentiation between possible acid-base disorders must be made primarily on the basis of clinical information.

An important test in the diagnosis of electrolyte and acid-base disturbances is the laboratory analysis of a patient's blood sample. Also called the serum electrolytes, this test measures the concentrations of sodium (Na), potassium (K), chloride (Cl), and bicarbonate (HCO_3). Very often a test for concentration of creatinine is also made. This test does not, however, measure the

Fig. 3. Nomogram of acid-base disturbances



concentrations of anions such as phosphate, sulfate, proteins, and organic acids which are normally present in the blood in small amounts. The combined concentrations of these unmeasured anions is called the *anion gap*. The anion gap can be approximated by subtracting from the combined sodium and potassium concentrations the combined concentration of chloride and bicarbonate, an amount normally approximately 12 meq/L.

Determination of the anion gap is vital to the diagnosis and differentiation of metabolic acidosis. The anion gap differentiates metabolic acidosis into two categories: one with an increased anion gap and other with a normal anion gap. Metabolic acidosis with an increased anion gap is generally caused by increased production or impaired excretion of H^+ and unmeasured anions by the body. For example, diabetic ketoacidosis, in which the acidosis results from increased production of ketones. On the other hand normal anion gap acidosis is generally caused by loss of HCO_3. For example, diarrhea, in which HCO_3 rich gastrointestinal fluids are lost.

1.4 Desiderata

In this section we discuss some of the characteristics required of the program if it is to be useful and effective as an expert consultant. They also serve as guiding principles for designing and evaluating the program. They are included here to communicate our aspirations. The goals described below have not been fully realized by the research reported here, nor can they all be fully realized by the current state of AIM technology. These characteristics are:

1.4.1 Making a Correct Diagnosis

The primary responsibility of the diagnostic program is to make a *correct diagnosis*. Without fulfilling this criterion, the program offers little possibility of being clinically useful. Although the issues involved in the evaluation of the efficacy of diagnosis by a program (or by a clinician) are difficult and controversial, it is clear that the diagnosis arrived at by the program must be a reasonable and thorough diagnosis in the light of the available information. Furthermore, a distinction must be made between a *working* diagnosis and the *correct* diagnosis. In practice, a correct diagnosis is often impossible owing to the high cost (medical and economic) of the information necessary to achieve it. A criterion for deciding when a working diagnosis has been achieved (for the purpose of management of a patient) should weigh the costs of gathering further information in terms of morbidity, time and money vs. the benefits of better diagnosis in terms of an improved management plan and a more reliable prognosis. For example, in situations in which the management plan for each of the diagnostic possibilities is the same, attempts to distinguish between diagnostic alternatives does not have any immediate utility. Hence, the working diagnosis should be considered sufficient. It should, however, re-evaluate the diagnosis as new information becomes available from the evolution of the disease or from the patient's response to therapy.

1.4.2 Continued Management of the Patient

Typically, a patient is examined by a physician more than once. The interaction between the patient and the physician can be divided into the initial interaction and the follow-ups. The follow-up sessions are used by physicians in evaluating the management plans and in refining the working diagnosis. In the majority of cases, follow-up sessions are essential for the proper practice of medicine. Furthermore, the ability to review the diagnostic decision during follow-up allows a program to revise its erroneous or incomplete conclusions.

1.4.3 Diagnostic Style

The diagnostic style used by a program is almost as important as reaching the correct diagnosis. Although good style is hard to characterize and even harder to embody in a program, certain aspects of diagnostic style are recognizable. For example, if the program pursues some low priority diagnostic problem in the face of more important issues, if it ignores a problem of life-

threatening character, or if the stream of questions seem pointless (i.e., if the program continues to ask questions when it should have been prescribing treatment), it is likely to be rejected by the user physician.

We wish to design a program which will exhibit focused, coherent and purposeful behavior in problem solving and will know when to call a halt to its question and make an interim diagnostic judgment. In a later section we will discuss how some of these requirements can be met using notions such as *hypothetical reasoning* and *planning*.

1.4.4 Mode of Interaction

A distinction is often made between two forms of data acquisition in diagnosis: *active* and *passive* [Gorry68]. A passive mode is one in which the program is provided with all the information at one point and must make a diagnosis based on this information. An active mode is one in which the program must ask a question in order to obtain each new piece of information. The active process suffers from the shortcoming that the physician may be aware of some facts potentially useful in the diagnosis, but may not be able to communicate them to the program because each new piece of information must be requested by the program. The passive approach avoids this problem but places the responsibility of identifying relevant information on the physician. This is an unacceptable demand on a physician who is not an expert in the medical domain of the program.

Therefore, we propose a compromise position involving *mixed initiative*. In this mode, as in the active mode, the primary responsibility of gathering information still rests with the program. However, at each point in the consultation the user physician is allowed to provide a suggestion. The program must analyze this suggestion,[3] even if it chooses to ignore the suggestion as being irrelevant.

1.4.5 Handling Discrepant Information

In virtually any diagnostic workup a large amount of discrepant information must be dealt with. Some of the discrepancies arise because patients are not always accurate observers of their symptoms and because laboratory tests and medical records are often in error. In other cases a seeming discrepancy may arise because of incomplete information, i.e. there may be a valid (but so far unknown) explanation for the apparent disagreement. Correct evaluation of each type of discrepancy is critical, if the program is to perform effectively. It is necessary for a diagnostic program to be able to identify the discrepant information as it is presented in order to

---

3. The program may not, as was the case with some previous programs, put these suggestions "on hold" without reasoning about them until it is ready to ask about them. If the program does not think that the suggestion is relevant, it must make that decision explicitly.

be able to evaluate a discrepancy and choose strategies for dealing with it before incorporating it in the patient model. We have observed that the expectations of the physician play an important role in identifying possible discrepancies in the incoming information. They allows the physician to locally evaluate these discrepancies (with respect to the available evidence, physiological possibilities and the current hypothesis) and act upon them before assimilating the new information into his patient descriptions. A similar mechanism in the program is desirable. Summarizing, the importance of good handling of discrepant information can not be overstated, especially when the system is expected to be used in a normal clinical setting as well as in experimental situations.

1.4.6 Explanation

To be acceptable in an application domain such as medicine, an AIM program must go beyond providing competent advice; it must be able to explain and justify its conclusions to the user physician — much as the human consultants do today — in a language that the physician is familiar with. After all, it is the physician who provides the medical care and is primarily responsible for the welfare of the patient. It is therefore natural (even desirable) for a physician to balk at accepting advice from a "black-box" program. This reluctance perhaps accounts for much of the reported antipathy of physicians even to the programs that on statistical analysis have been shown to be as good as the expert physicians [Yu79, Kulikowski81, Long80]

We believe that a program's acceptability depends crucially upon its ability to adequately explain its reasoning and justify its conclusions. It depends on the physician being able to challenge some part of the program's conclusions and having the program explore alternatives suggested by the physician. Consultation is a "two way street"; it can be effective only if the consultant (who is an expert in the subject matter) and the physician (who is familiar with the patient) cooperate. If any program is to be successful as an expert consultant it must allow for such an exchange.

The foregoing discussion may suggest that AIM programs be perfect, a requirement that can never be met in a real world of imperfect knowledge, where even the best of the expert physicians differ with one another. The thrust of our argument here is more limited. We are not demanding perfection from AIM programs, only that they be acceptable. Note that a program which is not as good as the best expert may nevertheless be fruitfully applied if it is acceptable and if its use improves the performance of the average clinician (who is not likely to be as good as the best expert in any given area of expertise).

In this thesis we are not extending the methodology of explanation generation. Our main thrust is in applying the available methodology to a much more complex domain than has been hitherto tried. However, since it has been demonstrated that generation of quality explanation can not be achieved by retrofitting a program with explanation capabilities, the program must be

designed with the explanation abilities in focus [Swartout80]. Our main interest is in designing explicit representation and reasoning mechanisms in the program which will provide us with the ability to justify the program's diagnoses as well as its reasoning in achieving those diagnoses.

1.5 Survey of AIM programs

Teaching of diagnostic medicine is often organized around diseases, with an emphasis on associations between the diseases and signs and symptoms typically associated with them. After all, the diagnostic task is to identify the disease hypothesis which represents the true state of the world by using all available data. Based on this observation we can conceive of a simple representation of diagnostic knowledge which draws associations between disease hypotheses and data. Given this "primitive" organization, we may already envision a diagnostic algorithm consisting of the following steps:

Diagnostic Reasoning:

(1) Whenever a new finding is reported, add it to the set of reported findings.
(2) Determine all the diseases linked to the new finding and add these diseases to the set of active hypotheses (which is initially empty).
(3) Score the active hypotheses by counting the number of expected findings observed for each disease hypothesis,
(4) rank-order the active hypotheses based on their scores and report the ranking.

Information Gathering:

(5) Select the highest-ranking disease, at least one of whose associated findings has not yet been either affirmed or denied, and ask about that finding.
(6) If step-5 fails to select a question, ask if the user is willing to volunteer a finding.
(7) If no findings are offered, report the rank ordered diagnoses and their supportive findings and stop. Otherwise, repeat steps 1 through 6.

The above algorithm, in spite of its simplicity, already captures the essential structure of a number of diagnostic programs. The association between diseases and findings forms its *static knowledge* about the domain. The set of observed findings and the rank-ordered set of active disease hypotheses are its *patient specific model* and its *understanding* of the patient's illness. The process of rank-ordering disease hypotheses is its *diagnostic evaluation*, and the selection of an appropriate finding for inquiry is its *information gathering strategy*.

The algorithm described above suffers from many inadequacies due to its oversimplification. Far more serious, however, are the problems fundamental to the model of the algorithm itself. For example, the above algorithm views diagnosis as the task of identifying that disease hypothesis which provides maximal coverage over the set of findings. Although this view of diagnosis suggests a relatively straightforward and intuitively appealing implementation, we believe this to be inadequate. Disease processes are causal; we believe that diagnosis involves providing an *adequate* explanation of the observed findings by reconstructing the possible sequence of causal events leading to the observed findings.

The program's information gathering strategy is limited to selecting one question at a time. At the end of this question, the program re-evaluates its diagnostic understanding, reformulates a new diagnostic problem (which may or may not be related to the previous problem) and selects the next question to ask. If after asking one question the diagnostic hypothesis being pursued is not confirmed, it must compete with all other active hypotheses for the attention of the diagnostic problem solver. In other words, the attention span of the program in solving any given problem is exactly one question. This results in diagnostic inefficiencies and incoherent question sequences. This problem is well recognized, and programs such as Internist-I and PIP have attempted to group diagnostic questions into meaningful packages, abating the problem somewhat. The work presented in this thesis is based on our belief that a substantial reformulation of the basic algorithm is needed before the problem can be adequately addressed [Szolovits81b].

In the remaining part of this section we will briefly review the four major AIM projects dealing with diagnosis, namely Internist-I, the Present Illness Program, CASNET/Glaucoma and Mycin. A detailed description of these programs can be found in [Szolovits81]. A good review of computer-based decision aids in medicine, using both AI and conventional computer methodologies is to be found in [Shortliffe79]. [Szolovits78] offers suggestions on the issues of choice of methodology and validation for acceptance for AIM programs. [Schwartz70] contains a discussion of acceptability issues from the viewpoint of physicians.

1.5.1 Internist-I and Present Illness Program

The Internist-I program [Pople75a, Pople77] is based on a large data base and a relatively simple evaluation and problem-selection strategy. The Internist-I data base is constructed by linking diseases and their manifestations with two subjectively assessed scores; an *evocation strength* which describes how strongly the manifestation should suggest a disease, and a *frequency* which describes how commonly the particular manifestation is observed in a patient with a given disease. Both of these are supplied by objective assessment by physicians. All the diseases are arranged into a hierarchy organized around organ-systems. Each non-terminal in this hierarchy is linked to manifestations that are common to all its inferiors. During each cycle of the algorithm, all diseases with at least one reported manifestation are evoked[4] and scored. Next, these disease hypotheses are partitioned into competing and complementary sets. This partitioning scheme represents an important contribution of the Internist-I program. It is based on two concepts: the *shelf* — a list of important manifestations that are not explained either by this diagnosis or any diagnoses previously confirmed, and the *dominance relation* — a hypothesis A is

4. If a disease (A) and one of its inferiors (B) are evoked simultaneously, then (1) if there are no known findings that can differentiate between B and any of its sibling hypotheses, B is considered to be subsumed by A and deleted from the active set. Otherwise, (2) A is replaced by the set of its immediate inferior diseases that are evoked by the manifestation.

said to dominate hypothesis B if the shelf of A is a proper subset of the shelf of B. The *competing set* is then said to contain hypotheses that either dominate or are dominated by the highest-ranking hypothesis. All other hypotheses are considered complementary and are ignored. The competing set is further reduced by considering only those hypotheses whose scores are within a fixed range of the highest-scoring hypothesis. Based on the number and relative scores of the hypotheses under consideration a diagnostic strategy (differentiate, confirm or rule-out) is selected and the next question computed. Finally, this question is asked and the diagnostic cycle is repeated.

The Present Illness Program (PIP) [Pauker76] is a frame based [Minsky75] program for taking the present illness in the domain of renal diseases. The PIP data base is implemented using disease frames, each containing the relation of the given disease to its expected findings and to other diseases, and a scoring criterion for evaluating the disease hypothesis. Some of the findings associated with a disease are specially designated as *triggers*. The complementary relation between diseases is described using *causal*, *complicational* and *associational* links; the competing relation is expressed using *differential* links. Each disease frame also contains two types of scoring functions; the *logical decision criteria* and the *numerical likelihood estimator* where the first is used for categorical evaluation and the second for probabilistic evaluation of the likelihood of the disease hypothesis under consideration [Szolovits78]. The diagnostic algorithm of PIP is similar to the basic algorithm discussed before. We should note that PIP does not use the disease-hierarchy or multiple diagnostic strategies used by the Internist-I program. On the other hand, PIP uses a substantially richer representation mechanism for describing findings and diseases as compared to Internist-I. For example, PIP allows one to describe the finding of edema observed in a given patient to be "severe", "worse in evening" and "pedal" (around legs). Finally, it uses categorical as well as probabilistic criteria for confirming diseases.

Internist-I and PIP represent medical knowledge as well as patient specific facts in phenomenological terms. The lack of physiological knowledge results in their weakness in dealing with patient illnesses with multiple interacting etiologies. The lack of physiological knowledge also results in activation of all phenomenologically possible hypotheses, including those that, based on the case-specific knowledge, are physiologically improbable. Thus, increasing the efforts needed in scoring and ruling out these hypotheses explicitly. Furthermore, the diagnostic algorithms in Internist-I and PIP alternate between obtaining a fact and evaluating the hypothesis list, resulting in a lack of focused diagnostic inquiry as discussed before.

The patient-specific model in Internist-I and PIP consists of a collection of patient facts and the list of active hypotheses; it does not relate different findings and hypotheses into causal explanations. As a result these programs have only a fragmentary understanding about the patient's condition and they often change their description of the patient's illness radically without substantial indications to that effect.

1.5.2 CASNET/Glaucoma

The Glaucoma program deals with the diagnosis and treatment of eye diseases. It is implemented using the CASNET [Weiss74] theory of representation of causal knowledge. The medical knowledge in Glaucoma is represented as a network of physiological states. These states are linked together by subjectively assessed transition probabilities, and by support values indicating how strongly certain test results support the presence of a particular condition (state). The transitional probabilities are used primarily as a means of selecting the most appropriate next state to investigate and the support values are used to evaluate the score (fuzzy likelihood [Gaines76, Zadeh65]) of a state, which is used to confirm or deny a state. Finally, the patterns of confirmed and denied states in the network are interpreted using a number of programs which compare the progress of the diseases in the given patient with the diseases known to the individual program.

The use of physiological knowledge gives the glaucoma program a better understanding of the mechanisms of disease evolution and interaction than the other programs discussed above. However, its use of causal knowledge is restricted to the local propagation of likelihood weights to determine the most appropriate next state for investigation. The program cannot use hypothesized diagnoses to guide its diagnostic inquiry: it separates the process of information gathering from that of diagnosis. The information gathering is directed solely towards confirming (or ruling out) states in the causal net.[5] Moreover, the program works in a domain where the disease physiology is uniformly well understood and each state can be confirmed directly using some test. Therefore, the techniques developed in this program are not easily extendable to programs working in other domains of medical expertise.

1.5.3 Mycin

Mycin is a rule-based program [Shortliffe76, Davis77] for diagnosis and treatment of infectious diseases — in particular, bacterial infections in the blood (and recently extended to other infectious diseases). It represents medical knowledge in terms of production rules [Davis77] and uses a collection of associative triples to represent the patient specific knowledge [Shortliffe75, Shortliffe76]. A novel mathematical model of confirmation [Shortliffe76] selects a set of organisms suspected of causing the illness. Diagnosis is carried out using a simple goal-directed control structure with backward chaining. The highest-level goal of Mycin is to determine if the patient is suffering from a significant infection which should be treated, and if he is, to select the appropriate therapy. It retrieves all the rules applicable to this goal and applies them sequentially as follows. It attempts to ascertain whether the "conclusion" of a rule is valid by evaluating each of its premises. If this information is already available in the data base, the

5. During this phase the program does not attempt to identify diseases responsible for the presence of these states. The diagnosis is attempted separately after the information gathering phase is completed.

program retrieves it. If not, determination of this premise becomes the new goal, and the program recurs. If after trying all the relevant rules, the answer still has not been discovered, the program asks the user for the relevant clinical information which will permit it to establish the validity of the premise clause. Thus, the rules "unwind" to produce a succession of goals, and it is this attempt to achieve each goal that drives the consultation.

The rules in Mycin are used to represent the domain knowledge as well as to encode the flow of control of the program. This takes away some of the advantages of modularity of knowledge because one must take into account the possible interactions between rules during problem solving. The goal structure of Mycin allows efficient problem solving and can be used for explaining the problem solving behavior of the program, but the program cannot explain the medical significance of its behavior as this information is compiled out while writing the rules.

The rule-based Mycin methodology is applicable in fields where the domain specific knowledge can be described using judgmental rules. It appears to require a field which has attained a certain level of formalization with a generally recognized set of primitives and a minimal understanding of basic processes and which does not have a high level of interaction between conceptual primitives [Davis77]. Finally, the rule-based methodology developed by Mycin and its derivative programs can be used effectively in encoding knowledge needed in handling specific well defined situations such as special heuristics for differentiation between two similar diseases which are difficult to differentiate using global differentiation heuristics.

The programs described above can be classified as the "first generation AIM programs". These programs have contributed immensely by demonstrating the feasibility of using computers (and AI techniques) in medical diagnosis. Some of the significant developments in this regard are summarized here.

The active hypothesis set introduced in PIP and the hierarchic organization of diseases introduced in Internist-I provide useful techniques for organizing programs for efficiency. A heuristic to partition the hypothesis set into competing and complementary sets was introduced in Internist-I. In spite of its shortcomings, the partitioning heuristic is intuitively appealing and empirically effective [Pople75a]. An improved technique for identifying complementary and competing hypotheses, especially for illnesses caused by multiple diseases, is one of the topics of interest in this thesis.

Recognizing that pathognomonic and important evocative findings help to focus the diagnostician's attention sharply, mechanisms to flag such findings and their use in focusing the programs attention were developed in Internist-I and PIP. Heuristics to help confirm or eliminate hypotheses categorically (without resorting to revised probabilities and thresholds) and explicit differential diagnosis links to indicate well-known points of diagnostic confusion were also added in PIP.

Causality as a major mechanism for tying together independent hypothesized disorders was identified as a fundamental mechanism in the CASNET/Glaucoma program, Internist-I and PIP. The Glaucoma program went a step beyond the others in the use of causality by defining disease as a progression of causally connected states. However, in all three programs, the use of causality is limited to propagating probability-like estimates of likelihood which remain the primary criterion for their clinical decisions.

The need for explanation and justification capabilities in an AIM programs was first recognized by and implemented in MYCIN. In this chapter we have argued that these capabilities are essential for the success of any consulting program. In this thesis we take this capability to be an essential component of the design of ABEL program.

1.6 Outline of the Thesis

This thesis contains seven chapters and two appendices. Chapter 2 previews the capabilities of the program with the help of two simple examples. Chapter 3 describes the representation of ABEL's medical knowledge. The medical knowledge consists of a hierarchic representation of anatomical, physiological, etiological and temporal knowledge. This forms the groundwork for an efficient representation of diseases and their pathophysiology in the domain of electrolyte and acid-base disturbances. The diseases are defined in terms of their loci along these four dimensions, providing a natural hierarchic organization to the disease definitions. This framework of basic medical knowledge provides us with a vocabulary for expressing phenomenological and pathophysiological knowledge.

An expert physician may have an understanding of a difficult case in terms of several levels of detail. As noted earlier, at the shallowest level that understanding may be in terms of commonly occurring associations of syndromes and diseases, whereas at the deepest it may include the biochemical and pathophysiological interaction of abnormal findings. Chapter 3 describes a multi-level description of pathophysiology, where each level of description can be viewed as a semantic net of relations between diseases and findings. Each node in the net represents a normal or abnormal state and each link represents a relation (causal, associational, etc.) between these states. Each node is associated with a set of attributes describing the temporal characteristics, severity or value, and other relevant attributes. Each link describes a causal relation between a cause node and an effect node by specifying a multivariate relation between attributes of the cause and the effect. Additional information to support mapping knowledge at one level to an adjacent level is also described.

In Chapter 4, we propose the use of a *coherent hypothesis* as the logical unit of hypothesis representation. This captures our notion, expressed above, that the reasoner's hypothesis structure must account for the total state of mind of the reasoner including its current uncertainties. In the program, each coherent hypothesis is represented using a *patient specific*

model (PSM). Each PSM represents a causal explanation of all the observed findings and their interrelationships at various levels of detail. Note that within each PSM all the diseases, findings, etc., are mutually complementary, while the alternate PSM's are mutually exclusive and competing.

The PSM is created by instantiating portions of ABEL's general medical knowledge and filling in its details from the specific case being considered. The instantiation of the PSM is very strongly guided by initially given data, because the PSM includes only those disorders and connections that are needed to explain the current case. Instantiation is accomplished by five major operators. *Initial formulation* creates an initial patient description from the presenting complaints and laboratory results. *Aggregation* and *elaboration* make connections between the levels of detail in the PSM by filling in the structure above and below a selected part of the network, respectively. In a domain such as ABEL's, multiple disorders in a single patient and the presence of homeostatic mechanisms require the program to reason about the joint effects of several mechanisms which collectively influence a single quantity or state. *Component decomposition* and *summation* relate disorders at the same level of detail by mutually constraining a total phenomenon and its components; the net change in any quantity must be consistent with the sum of individual changes in its parts. The final operator, *projection*, forges the causal links within a single level of detail in the search for causal explanations. The operators all interact because the complete PSM must be self-consistent both within each level and across all its levels. Therefore, each operation typically requires the invocation of others to complete or verify the creation of related parts of the PSM. Furthermore, PSM's are organized in a context tree allowing different PSM's to share structures common to them. The root of the PSM-tree also contains all the observed findings and diseases which have been concluded to be true so that they may be shared by all PSM's.

Locality is a desirable property for the reasoning and description schemes. It imposes modularity in the organization of knowledge, making acquisition and representation of knowledge tractable. Furthermore, it makes possible efficient reasoning schemes whose resource requirements do not grow with increasing size of the data-base.[6] To exploit the locality constraint in reasoning with causal networks, a program should be able to reason based only on the information locally available from the neighborhood of the mechanism under consideration. Although it is always possible to choose a level of abstraction at which the interaction between a given pair of states can be described locally, for a given level of detail it is not possible to impose the locality constraint on every interaction. The multiple-level causal model and the abstraction/elaboration process presented in this thesis allow us to overcome this problem. For example, if at some level of detail two distant states interact, we can aggregate the description of

---

6. Locality has been exploited in a large number of diverse problems, such as common-sense reasoning [Minsky73, Kuipers77, deKleer79] and natural language processing [Marcus79, Church80, Martin81]. For example, the constraint of "context freeness" in natural language is a specific instance of locality constraint.

intervening causal network to a level where the two states are adjacent to one another. The interaction between the two can now be computed locally.

Chapter 5 discusses the diagnostic problem solving activity. The diagnostic problems are formulated by identifying the weaknesses and conflicts in the PSM's. The task of the diagnostic problem solver is to resolve these conflicts and weaknesses by gathering new information. We note that the medical knowledge in the program consists of prototypes of the disease entities. However, this prototypical knowledge can be substantially constrained because the hypothesized disease entities must be consistent with the known facts and explanations. We introduce the notion of a *diagnostic closure* which extracts and tailors that part of medical knowledge that is directly relevant to the diagnostic task at hand. The diagnostic closure brings together all the dependencies and expectations necessary for planning a diagnostic inquiry, for evaluating real and apparent discrepancies in the incoming information, and provides a framework for explaining the alternatives under consideration and for justifying the selection of questions. Although we envision using recent advances in the planning paradigm [Fikes72, Sacerdoti75, Stefik81], the current implementation of the program generates a simple tree-structured plan for information gathering by decomposing the problem by successive applications of confirm, rule-out, differentiate, and group-and-differentiate strategies. Finally, when a sufficient amount of new information is available the program assimilates this information into the PSMs and the diagnostic process is repeated. The process terminates when an adequate explanation for the patient's illness is found or when all the information necessary for such an explanation is exhausted.

In chapter 6 we revisit the example described in chapter 2 in greater detail. Chapter 7 summarizes the experience gained and lessons learned in this enterprise and indicates pointers to future research. Finally, appendix 1 briefly summarizes the XLMS system (a knowledge representation system built on top of LISP) used by ABEL. Appendix 2 summarizes the techniques for translating the internal data structures of the program into English developed recently by Swartout [Swartout80] and discusses algorithms for organizing the concepts encoded in causal networks into a linear sequence of sentence level objects that can then be translated using the above-mentioned methodology.

2. Examples

This chapter presents the inner workings of ABEL with the help of annotated examples. In this chapter the reader is not expected to understand *how* the program accomplishes its task, but rather just *what* it does. The succeeding chapters will examine the structure of the program and the method by which each step is accomplished. We will consider two examples: (1) a patient suffering from moderately severe salmonellosis, and (2) a patient suffering from moderately severe salmonellosis and vomiting. The selection of the medical examples is motivated by our desire to make the medical contents of the examples as simple as possible. In chapter 6 we will revisit these examples and discuss how the program accomplishes each of its tasks.

2.1 Example 1: Salmonellosis

For the first example let us consider a 40 year old 70 Kg male patient who has been suffering from moderately severe salmonellosis and, as a result, has developed moderately severe metabolic acidosis and hypokalemia. To illustrate the program let us provide it initially with only the laboratory analysis of the patient's blood sample (serum analysis) without any clinical information.

```
Serum Analysis:
    Time:   0                   time of the session
     Sex:   male
      Na:   142 meq/l           normal
       K:   3   meq/l           moderately low
      Cl:   113 meq/l           normal
    HCO3:   15  meq/l           moderately low
    pCO2:   30  mmHg            moderately low
```

Based on these data, the program generates all possible acid-base disturbances that can account for the laboratory data. It then prunes and rank-orders these disturbances based on their complexity, likelihood and severity of each component. The rank-ordered list of likely disturbances is:

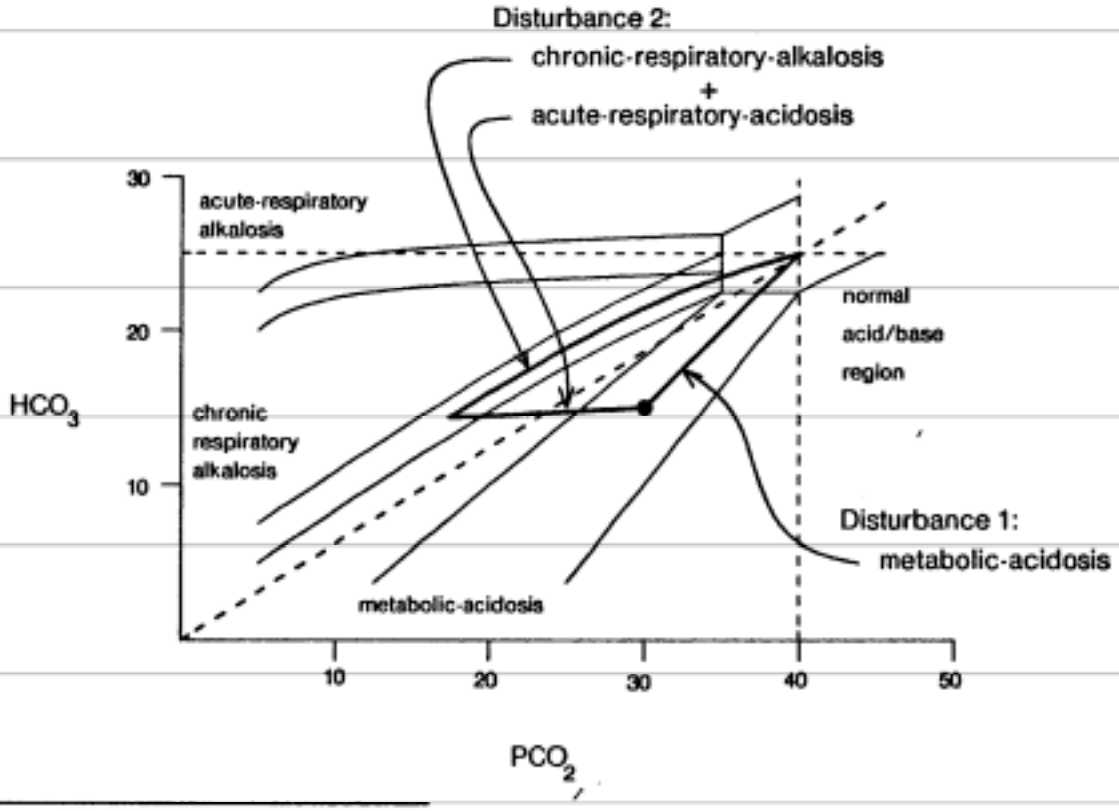
```
---- Patient Acid-Base Profile ----

1. metabolic-acidosis [severity: 0.4]                    very likely
2. chronic-respiratory-alkalosis [severity: 0.68]
   + acute-respiratory-acidosis [severity: 0.32]         unlikely
```

The computation of the acid-base profile is based on the Nomogram of Acid-Base Disturbances described in chapter 1. Figure 4 shows the relevant region of this nomogram with the loci of the two hypothesized disturbances. The estimation of the severity of a disturbance is

based on the length of the segment along the locus of that disturbance. Thus, we note that the severity of the single acid-base disturbance (metabolic acidosis) is only about 0.4 while an equivalent acid-base disturbance composed of chronic respiratory alkalosis and acute respiratory acidosis has severities of 0.68 and 0.32 respectively.[7]

Next, the program creates a PSM for each possible acid-base disturbance and interprets the laboratory data in the context defined by each acid-base disturbance. For example, with the assumption of fully compensated metabolic acidosis, the entire change in the PCO_2 may be considered chronic, therefore, the chronic component of the PCO_2 will be 30 meq/l, while with the assumption of chronic respiratory alkalosis and acute respiratory acidosis, the chronic component of PCO_2 is due only to the chronic component of this disturbance, therefore reading from the nomogram we find that the chronic value of PCO_2 in this case will be approximately 16 meq/l.

Fig. 4. Graphic depiction of the two Acid-Base hypotheses



7. The numbers corresponding to the acid-base disturbances computed above are the programs internal assessment of the severity of illness, they are not measurable.

The program then aggregates its patient-specific physiologic knowledge to formulate an interpretation of the laboratory data at the clinical level. The computer generated explanation of its interpretation of these data under the two major hypotheses is described in figures 5 and 6.

A quick look at the two clinical level explanations shows that the structure involving hypokalemia and acidemia is common to the two hypotheses. They differ in their accounting for acidemia. Note that the clinical level abstraction of the two hypotheses is fairly simple in structure and does not contain any feedback cycles. The cycles present at the intermediate level describing the interaction between the acidemia, hypobicarbonatemia and hypocapnia have been abstracted away. A closer look at these feedback cycles shows the principal difference between the two hypotheses. In the first case, the change in the acid-base state is a consequence of loss of HCO_3 from the body which causes hypobicarbonatemia, whereas in the second it enters as primary disturbance in ventilation which alters the PCO_2. Finally, we note that the first hypothesis has two unaccounted findings while the second hypothesis contains three unaccounted findings.

In the context of this initial analysis of the patient's condition, the program starts the diagnostic exploration. An annotated (in italics) transcript of the program's diagnostic behavior is shown next.

The program computes the diagnostic closures for the two hypotheses and decides to pursue the first hypothesis.

```
Differentiating between the causes of the leading
complete hypothesis.

     1    SALMONELLOSIS
     2    URETEROSIGMOIDOSTOMY
     3    VILLOUS-ADENOMA
     ----------------------------------
     4    DISTAL-RTA
     5    PROXIMAL-RTA
     6    ACUTE-RENAL-FAILURE
     7    CHRONIC-RENAL-FAILURE
continue? ==>
```

The list above contains all possible diseases that can explain some part of the first hypothesis. The list is divided into groups of diseases by the number of unaccounted findings that each disease can explain succinctly. Within each group the diseases are ordered by a secondary scoring criterion based on the quality of their match with the hypothesis and their potential to be ultimately confirmed.

```
Differentiating between
     SALMONELLOSIS URETEROSIGMOIDOSTOMY VILLOUS-ADENOMA
```

Fig. 5. Comparision of hypotheses 1 & 2 at clinical level

Hypothesis 1: Metabolic Acidosis

This is a 40 year old 70.0 kg male patient. His electrolytes are:

Na: 142.0	HCO3: 15.0	Anion Gap: 13.0
K: 3.0	pCO2: 30.0	
Cl: 113.0	pH: 7.32	

The patient has moderate metabolic acidosis and mild hypokalemia. The metabolic acidosis causes mild acidemia. The acidemia partly compensates the suspected moderate hypokalemia leading to the observed hypokalemia. The metabolic acidosis remains to be accounted for. The hypokalemia has only been partially accounted for.

Hypothesis 2: Chronic Resp. Alkalosis & Acute Resp. Acidosis

This is a 40 year old 70.0 kg male patient. His electrolytes are: ...

The patient has moderate chronic respiratory alkalosis, moderate acute respiratory acidosis and mild hypokalemia. The acute respiratory acidosis and chronic respiratory alkalosis cause mild acidemia. The acidemia partly compensates the suspected moderate hypokalemia leading to the observed hypokalemia. The chronic respiratory alkalosis and acute respiratory acidosis remain to be accounted for. The hypokalemia has only been partially accounted for.

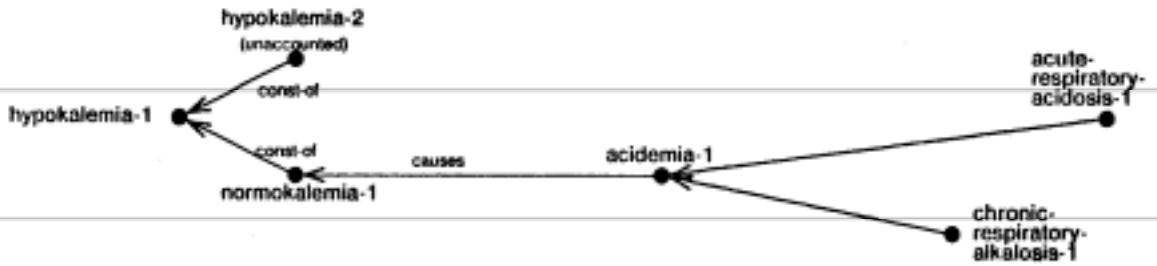
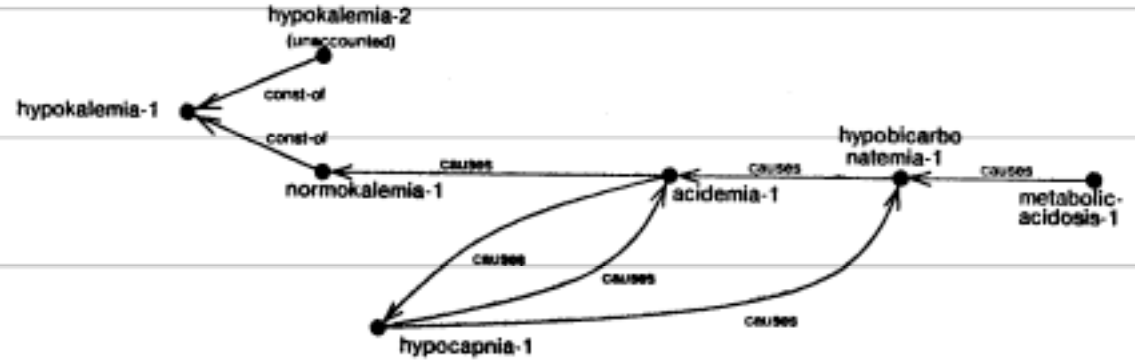


Fig. 6. Comparision of hypotheses 1 & 2 at intermediate level

Hypothesis 1: Metabolic Acidosis

This is a 40 year old 70.0 kg male patient. His electrolytes are: ....

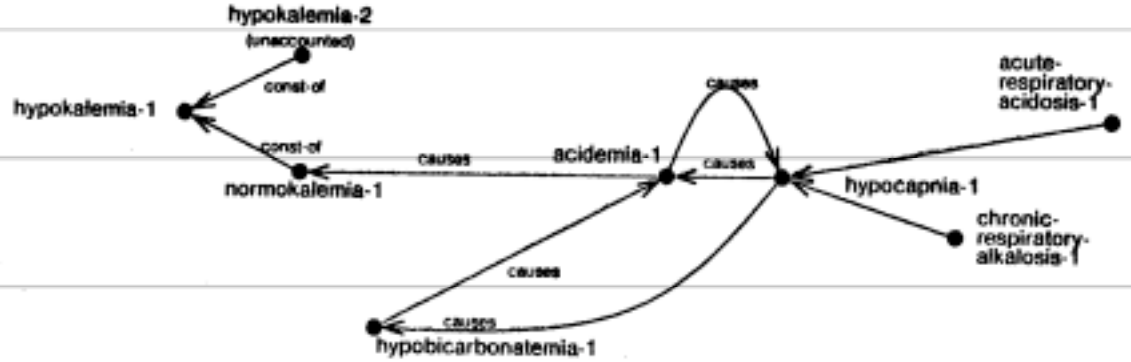
The patient has moderate metabolic acidosis, mild hypokalemia and moderate hypobicarbonatemia. The metabolic acidosis along with moderate hypocapnia causes hypobicarbonatemia. The hypobicarbonatemia along with hypocapnia causes mild acidemia. The acidemia partly compensates the suspected moderate hypokalemia leading to the observed hypokalemia. The metabolic acidosis remains to be accounted for. The hypokalemia has only been partially accounted for.



Hypothesis 2: Chronic Resp. Alkalosis & Acute Resp. Acidosis

This is a 40 year old 70.0 kg male patient. His electrolytes are: .....

The patient has moderate acute respiratory acidosis, moderate chronic respiratory alkalosis, mild hypokalemia and moderate hypobicarbonatemia. The chronic respiratory alkalosis and acute respiratory acidosis along with mild acidemia cause moderate hypocapnia, which causes hypobicarbonatemia. The hypobicarbonatemia and hypocapnia cause acidemia. The acidemia partly compensates the suspected moderate hypokalemia leading to the observed hypokalemia. The acute respiratory acidosis and chronic respiratory alkalosis remain to be accounted for. The hypokalemia has only been partially accounted for.



As a first step towards differentiation, the program asks if the user is already aware of any of the possible alternatives.

```
Does the patient have any of the following?
     1    SALMONELLOSIS
     2    URETEROSIGMOIDOSTOMY
     3    VILLOUS-ADENOMA

Please enter numbers between 1 and 3
Present: ==> E(xplain)
```

The user indicates that he would like an explanation. The program prompts by indicating in a menu the possible choices for explanation.

```
     1    DESCRIBE-HYPOTHESIS
     2    DESCRIBE-DIAGNOSTIC-REASONING
     3    DESCRIBE-FINDING
     4    DESCRIBE-CAUSE
     5    DONE
==> 2
The stack of diagnostic goals:

     3    DIFF-CAUSES
     2    CONFIRM-CH
     1    DIAGNOSE

    I am diagnosing the possible acid base disturbances.  I am
confirming the leading hypothesis.  I am differentiating between
its leading causes.

     1    DESCRIBE-HYPOTHESIS
     2    DESCRIBE-DIAGNOSTIC-REASONING
     3    DESCRIBE-FINDING
     4    DESCRIBE-CAUSE
     5    DONE
==> 4

Please enter a cause for viewing: ==> salmonellosis

    Salmonellosis may account for moderate metabolic acidosis and
moderate hypokalemia as follows.
    Moderate salmonellosis may cause metabolic acidosis.
    Moderate salmonellosis may cause hypokalemia

     1    DESCRIBE-HYPOTHESIS
     2    DESCRIBE-DIAGNOSTIC-REASONING
     3    DESCRIBE-FINDING
     4    DESCRIBE-CAUSE
     5    DONE
==> 5
```

```
Does the patient have any of the following?
     1    SALMONELLOSIS
     2    URETEROSIGMOIDOSTOMY
     3    VILLOUS-ADENOMA

Present: ==> none      Absent: ==> none      Unknown: 1 2 3

Differentiating between
     SALMONELLOSIS URETEROSIGMOIDOSTOMY VILLOUS-ADENOMA
```

Salmonellosis, ureterosigmoidostomy and villous-adenoma all cause a reduction in extracellular fluid, resulting in dehydration. However, the program notices that some of the diseases in the second set (e.g., renal failure) may have the exact opposite effect of causing edema. Therefore, while exploring the state of extracellular fluid the program includes edema in the question.

```
Does the patient have one of the following?
     1    DEHYDRATION
     2    EDEMA

Present: ==> none      Absent: ==> none      Unknown: 1 2
```

The program is expecting dehydration. Therefore, when we fail to confirm or deny the dehydration the program pursues the finding further.

```
I would like to ask about the effects of SALMONELLOSIS.

Is the value of SERUM-CREATININE known?      ==> E(xplain)


    High serum creatinine may be caused by moderate salmonellosis
as follows:

     Moderately high serum creatinine may be caused by moderate
dehydration, which may be caused by salmonellosis.
```

The user indicates that he would like a justification for this question. The program generates the explanation by tracing back the causal path from serum creatinine in the DC associated with the goal of the question.

The program's diagnostic reasoning at this point can be described as follows: The top level goal of the program is to do diagnosis. In order to do the diagnosis the program evaluates the two alternate hypotheses and selects the first hypothesis (metabolic-acidosis) for confirmation. To confirm this hypothesis it selects the set of leading causes for the first hypothesis, namely, salmonellosis, ureterosigmoidostomy and villous-adenoma. It then attempts to differentiate between these causes. It determines the findings predicted by each of these three causes, orders them according to their discriminatory power, and asks about them, the most discriminating finding first.

```
Is the value of SERUM-CREATININE known? ==> yes

Please enter the attributes of SERUM-CREATININE
What is the VALUE of SERUM-CREATININE ? ==> 3
What is the START-TIME of SERUM-CREATININE ? ==> 0

Is the value of MEAN-ARTERIAL-BLOOD-PRESSURE known? ==> yes

Please enter the attributes of MEAN-ARTERIAL-BLOOD-PRESSURE
What is the VALUE of MEAN-ARTERIAL-BLOOD-PRESSURE ? ==> 75
```

The program has now completed one full cycle of its planned diagnostic inquiry. It now incorporates this information into both hypotheses and starts the next cycle of diagnostic planning.

```
Starting next cycle of diagnosis

Differentiating between the causes of the leading complete hypothesis.

     1    SALMONELLOSIS
     ----------------------------------
     2    VILLOUS-ADENOMA
     3    URETEROSIGMOIDOSTOMY
     ----------------------------------
     ----------------------------------
     4    ADRENAL-INSUFFICIENCY
     5    DIABETES-INSIPIDUS
     6    ACUTE-RENAL-FAILURE
     7    CHRONIC-RENAL-FAILURE
     ----------------------------------
     8    DISTAL-RTA
     9    PROXIMAL-RTA
continue? ==>
```

The program has already gathered sufficient information to confirm salmonellosis. It is unable to do so because we have not implemented the criteria for confirming a disease yet. However, we note that the information gathered has provided substantial categorical separation between the gastrointestinal and the renal causes of the metabolic-acidosis with hypokalemia.

```
Pursuing SALMONELLOSIS: no findings available

Differentiating between VILLOUS-ADENOMA URETEROSIGMOIDOSTOMY:
     no finding available

Differentiating between ADRENAL-INSUFFICIENCY DIABETES-INSIPIDUS
ACUTE-RENAL-FAILURE CHRONIC-RENAL-FAILURE
```

As there are more than two items in the differentiation set, the program groups these items into renal and extra-renal sets. It then pursues the renal set first.

```
Does the patient have any of the following?
     1    CHRONIC-RENAL-FAILURE
     2    ACUTE-RENAL-FAILURE

Present: ==> none     Absent: ==> 1 2

Does the patient have any of the following?
     1    ADRENAL-INSUFFICIENCY
     2    DIABETES-INSIPIDUS

Present: ==> none     Absent: ==> 1     Unknown: 2

Differentiating between DISTAL-RTA PROXIMAL-RTA

Does the patient have any of the following?
     1    PROXIMAL-RTA
     2    DISTAL-RTA

Present: ==> none     Absent: ==> none     Unknown: 1 2
```

The program has now completed the second cycle of its planned diagnostic inquiry. No new finding was uncovered during this cycle of diagnosis and therefore, the program does not need to revise the two hypotheses. However, during the planning of the next cycle of diagnostic inquiry it discovers that all the questions relevant to the etiology of metabolic-acidosis and hypokalemia have already been exhausted. Therefore, in order to proceed it requests that some possible etiology be assumed.

```
Starting next cycle of diagnosis
No new finding added in the previous cycle.

    All possible etiologies that could explain the patient's
illness are unknown.  In order to proceed we must at least
hypothetically assume one of them.  Possible etiologies that could
explain the patient's illness listed in decreasing order are:

     1    SALMONELLOSIS
     ----------------------------------
     2    VILLOUS-ADENOMA
     3    URETEROSIGMOIDOSTOMY
     ----------------------------------
     ----------------------------------
     4    DIABETES-INSIPIDUS
     ----------------------------------
     5    DISTAL-RTA
     6    PROXIMAL-RTA

Would you like to assume SALMONELLOSIS ?     ==> yes

Assuming MODERATE ACUTE SALMONELLOSIS.
```

The program adds salmonellosis to the patient models and re-evaluates two hypotheses before resuming the diagnosis again. However, it immediately recognizes that it has found the missing piece in the puzzle! Based on the assumption that the patient has salmonellosis, the program selects the first PSM (metabolic acidosis) as the explanation of the patient's acid-base disorder. The program's explanations of the successful and the alternate PSMs are shown next. Note, however, that these explanations are provided with the assumption that the patient has salmonellosis, an assumption that needs to be verified.

```
Diagnosis completed.  The successful diagnosis is:

    This is a 40 year old 70.0 kg male patient with moderate
salmonellosis.  His electrolytes are:

        Na: 142.0          HCO3: 15.0        Anion Gap: 13.0
         K: 3.0            pCO2: 30.0
        Cl: 113.0            pH: 7.32      Creatinine: 3.0

    The salmonellosis causes moderate metabolic acidosis and
moderate dehydration.  The dehydration causes moderate
hypotension and moderate high creatinine disturbance.  The
metabolic acidosis causes mild acidemia.  The salmonellosis and
acidemia cause mild hypokalemia.  All findings have been
accounted for.
```

As can be seen from the above explanation, the program has concluded that salmonellosis provides an adequate explanation for the patient's illness and that the first hypothesis is substantially superior to the alternate hypothesis. The alternate hypothesis with all the available information added is shown below.

```
   This is a 40 year old 70.0 kg male patient with salmonellosis.
His electrolytes are:

        Na: 142.0          HCO3: 15.0        Anion Gap: 13.0
         K: 3.0            pCO2: 30.0
        Cl: 113.0            pH: 7.32      Creatinine: 3.0

   The salmonellosis causes moderate metabolic acidosis and
moderate dehydration.  The dehydration causes moderate
hypotension and moderate high creatinine disturbance.
Moderate acute respiratory acidosis, moderate chronic
respiratory alkalosis and metabolic acidosis partly
compensate the suspected mild alkalemia leading to the
observed mild acidemia.  The salmonellosis and acidemia cause
mild hypokalemia.  The chronic respiratory alkalosis and
acute respiratory acidosis remain to be accounted for.  The
alkalemia has only been partially accounted for.
```

Notice the difference in the two explanations. The first explanation contains only one acid-base disturbance, while the second explanation contains a total of three acid-base disturbances. Furthermore, all the findings in the first hypothesis have been accounted for while the second hypothesis has three acid-base disturbances still to be accounted for.

2.2 Example 2: Vomiting and Salmonellosis

The next example illustrates the program's capabilities in dealing with multiple etiologies, namely salmonellosis and vomiting, which offset the effects of each other on the acid-base balance. We will focus on the program's understanding and its ability to reformulate this understanding when new information is provided.

Let us consider a patient who is suffering from moderately severe vomiting for the past two days who then develops salmonellosis. Note that the electrolyte and acid-base disturbances in vomiting result from the excessive loss of upper gastrointestinal fluid, while in salmonellosis they result from the loss of lower gastrointestinal fluid. The upper GI fluid is acidic while the lower GI fluid is alkaline, therefore the two tend to have offsetting effects on the acid-base balance. However, vomiting and salmonellosis both cause hypokalemia and dehydration, therefore they compound these effects of each other. For this example, let us consider a patient in which the presentation of vomiting and salmonellosis are such that each exactly cancels the acid-base effect of the other, leaving the patient with no acid-base disturbance. We will illustrate the program's handling of this case by describing the program's understanding of the case at three points during the diagnostic process: (1) just after the electrolyte values are entered in the program, (2) after the finding of vomiting has been presented, and (3) at the end of the diagnostic process.

The program's evaluation of the serum electrolytes and the English explanation of its initial hypothesis are:

```
Serum Analysis:
    Time:  0
     Sex:  male
      Na:  141  meq/l      normal
       K:  2    meq/l      low
      Cl:  108  meq/l      normal
    HCO3:  25   meq/l      normal
    pCO2:  39   mmHg       normal

    ---- Patient Acid-Base Profile ----
    1.  normal-acid-base-state

    This is a 40 year old 70.0 kg male patient with moderate
hypokalemia.  His electrolytes are: .....
```

The serum analysis reveals only one abnormal finding, hypokalemia. The program starts the diagnostic process by attempting to differentiate between the possible causes of hypokalemia which include vomiting and salmonellosis along with other etiologies such as laxative abuse, diuretic use, hyperaldosteronism etc. The summary of the program's hypothesis after the finding

of moderately severe vomiting has been presented is:

```
     This is a 40 year old 70.0 kg male patient with moderate
vomiting.  His electrolytes are:

        Na: 143.0          HCO3: 25.0          Anion Gap: 12.0
         K: 2.0            pCO2: 39.0
        Cl: 108.0            pH: 7.42

     The vomiting causes moderate metabolic alkalosis.  Moderate
hypokalemia is partly caused by vomiting leaving some
additional factor causing hypokalemia still unaccounted for. The
hypokalemia and moderate acidemia have only been partially
accounted for.
```

Notice that the vomiting partially accounts for the observed hypokalemia. However, in order to account for the hypokalemia the program must assume that there has been substantial upper GI fluid loss sufficient to also cause metabolic alkalosis. As this metabolic alkalosis is not consistent with the normal acid-base state the program must decompose the normal acid-base state into offsetting alkalemia and acidemia. The alkalemia which is accounted for by metabolic alkalosis, and acidemia which remains unaccounted for. The remaining unaccounted components now present a picture similar to that of the previous case (example 1) and the diagnosis proceeds similarly. The diagnosis is completed when the program is told about salmonellosis (the remaining disturbance). A summary of the programs' final diagnosis is described next.

```
     This is a 40 year old 70.0 kg male patient with moderate
vomiting and moderate salmonellosis.  His electrolytes are:

        Na: 143.0          HCO3: 25.0          Anion Gap: 12.0
         K: 2.0            pCO2: 39.0
        Cl: 108.0            pH: 7.42        Creatinine: 3.0

     The vomiting causes moderate metabolic alkalosis.  The
salmonellosis and vomiting cause moderate dehydration, which
causes moderate hypotension.  The dehydration also causes
moderate high creatinine disturbance.  The salmonellosis causes
moderate metabolic acidosis.  The metabolic acidosis and
metabolic alkalosis cause normal ph.  The salmonellosis, normal
ph and vomiting cause moderate acute hypokalemia.  All findings
have been accounted for.
```

The primary focus of this thesis is in developing a methodology for knowledge representation and manipulation that allows our program to exhibit the understanding of patient illness demonstrated above. In the next three chapters we will study in detail this methodology and its

implementation before revisiting the same examples again in greater detail.

3. Representation of Medical Knowledge

Illness can be described as a change in the normal state or function in a patient. To describe an illness, we need a formalism to represent the states, the state changes, the normal and the abnormal functions and their interactions in terms of the primitives known to the system. This knowledge is organized in the program with the help of (1) an anatomy component, which includes a *part-of* hierarchy for organ systems, *contained-in* and *position* relations for major anatomical features, and a *connected-to* relation which provides material flow information. (2) A physiology component, where our concentration has been only on the fluid and electrolytes, describes the fluid compartments of the body, the spaces of distribution of various solutes, and the relative distribution of losses and gains in the various compartments under different conditions. (3) A pathophysiology component, which contains some primitive knowledge about disease etiologies, a taxonomy of disease processes, and causal relations which describe how the changes in a given state influence other states.

It is also important to recognize commonly occurring constellations of abnormal states as special composite situations. Conceptualization of these composite situations in a diagnostic system is important because it provides us with the ability to reason at a high level of abstraction, and to organize a large number of seemingly unrelated facts into a coherent whole. We have argued that it is crucial for any diagnostic system to have the ability to reason simultaneously at a high level of abstraction consisting of phenomenological knowledge as well as at a physiological level. We accomplish this with the help of a multi-level model for representation of diseases and causal phenomena. This is motivated by the observations made by Lynch while studying the *conceptual maps* of metropolitan regions. He notes

> *"Rather than a single comprehensive image for the entire environment, there seemed to be sets of images, which more or less overlapped and interrelated. They were typically arranged in a series of levels, roughly by the scale of area involved, so that the observer moved as necessary from an image at street level to levels of a neighborhood, a city, or a metropolitan region."*
>
> — The image of the city [Lynch60, pages 85-86].

The structure of the cognitive map described above is a product of the necessity to cope with *large-scale maps*; maps that are too large to be perceived at once, too large to be stored in the short-term memory by their users at a single instance of time, and too complex to be computationally tractable in solving problems (such as finding an efficient path between two points on the map). An important observation in formulating cognitive maps is that they are organized around *landmarks*. The conceptualization can be achieved by expanding the denotation of a landmark to subsume the local topology surrounding the designated location. If this conceptualization is carried out carefully, so that the areas subsumed by these landmarks overlap and cover the entire detailed map, it is possible to maintain sufficient coherence

(mapping) to be able to move between different levels of description.

Based on these observations and similar observations of a physician's use of medical knowledge, we have developed a hierarchical multi-level representation scheme to describe medical knowledge. The lowest level of this description consists of pathophysiological knowledge about diseases, which is successively aggregated (summarized) into higher level concepts and relations, gradually shifting the content of the description from physiological to syndromic knowledge. The aggregate syndromic knowledge provides us with a concise global perspective and helps in the efficient exploration of the diagnostic alternatives. The physiological knowledge, on the other hand, provides us the capabilities of handling complex clinical situations arising in patients with multiple disturbances, evaluating the physiological validity of the diagnostic possibilities being explored, and organizing a number of fragmented and seemingly unrelated facts into a coherent causal description.

3.1 Anatomical Knowledge

The anatomical knowledge of the system includes (1) a *part-of* hierarchy for organ systems, (2) *connected-to* relations, which provide the material flow information, and (3) *contained-in* and *position* relations which provide gross anatomical relations between anatomical entities.

3.1.1 Anatomical Taxonomy

The part-of hierarchy defines the various anatomical parts of the body by defining each organ system in relation to the body, and each sub-organ in relation to the organ-system containing it. The part-of hierarchy provides us with the taxonomic hierarchy for anatomical parts. A small section of the part-of hierarchy[8] and its graphical representation is shown in figure 7.

3.1.2 Material Flow Pathways

Material flow (e.g. the flow of glomerular filtrate) is represented by the *connected-to* relation. For example, the path of the filtrate in the kidney can be described as shown in figure 8. As can be seen from the figure, the material flow relation is specified at various levels of detail. The rationale for this multiple level description is provided later on in this section.

The anatomical knowledge that follows in the remainder of this section has been included to provide a fuller description of ABEL's knowledge base. However, this knowledge is currently not used by the program in its diagnostic reasoning.

---

8. The data are expressed in XLMS [Hawkinson80], which is briefly described in appendix 1.

Fig. 7. The part-of hierarchy

```
[body
   [urinary-system = (anat-entity*s "urinary-system")↑u
      [kidney = (anat-entity*s "kidney")↑u
         [cortex = (anat-entity*s "cortex")↑u]
            [medulla = (anat-entity*s "medulla")↑u]
          [nephron = (anat-entity*s "nephron")↑u
             [tubule = (anat-entity*s "tubule")↑u
                 [proximal-tubule
                 = (anat-entity*s "proximal-tubule")↑u]
               [loop-of-henle
                  = (anat-entity*s "loop-of-henle")↑u]
               [distal-tubule
                  = (anat-entity*s "distal-tubule")↑u]]
            [glomerulus = (anat-entity*s "glomerulus")↑u]]
          [collecting-duct
           = (anat-entity*s "collecting-duct")↑u]]
      [ureter = (anat-entity*s "ureter")↑u]
      [bladder = (anat-entity*s "bladder")↑u]
      [urethra = (anat-entity*s "urethra")↑u]]]
```

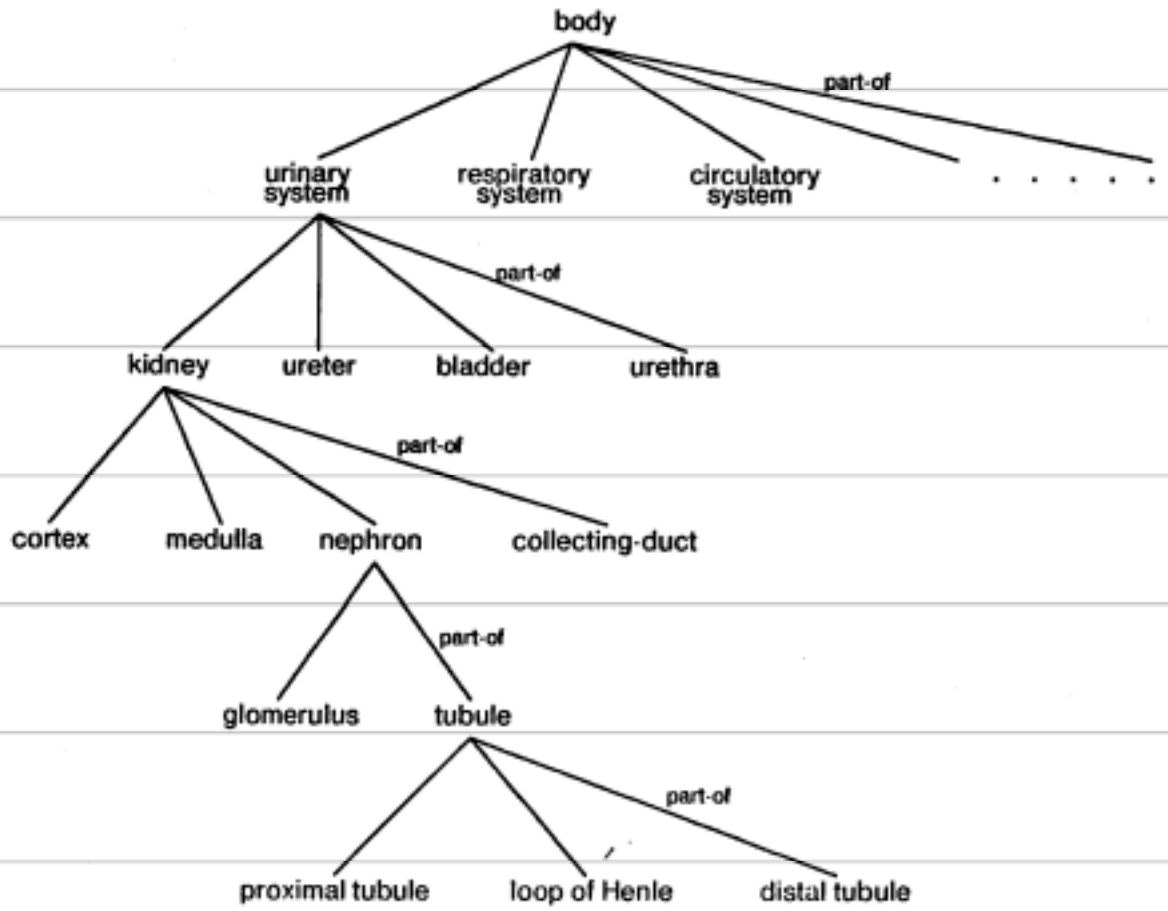
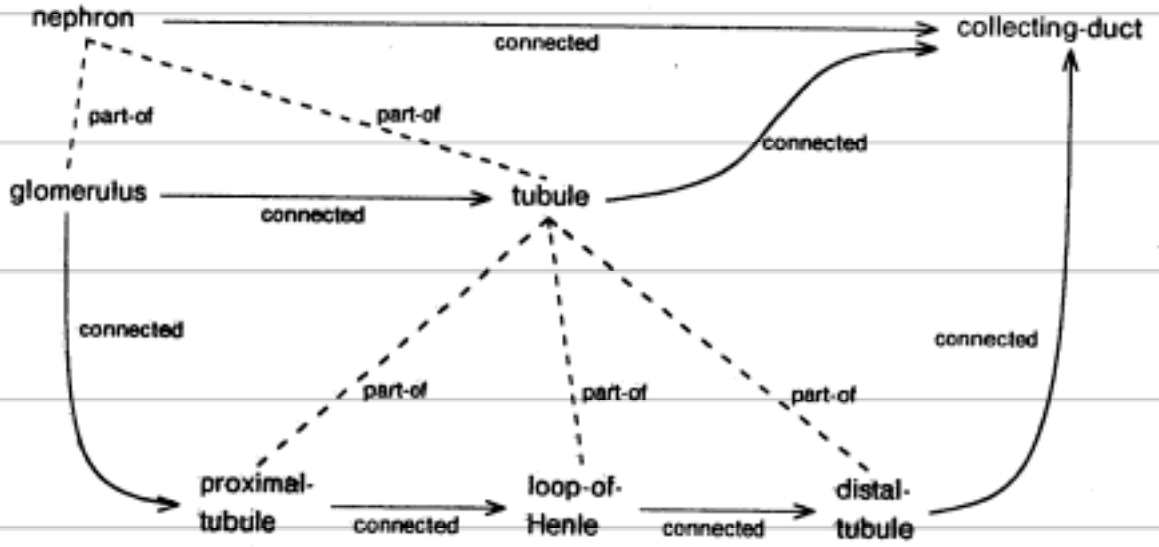


Fig. 8. Material flow relations

```
[((connected*b nephron)*e collecting-duct)]
[((connected*b glomerulus)*e tubule)]
[((connected*b tubule)*e collecting-duct)]

[((connected*b glomerulus)*e proximal-tubule)]
[((connected*b proximal-tubule)*e loop-of-henle)]
[((connected*b loop-of-henle)*e distal-tubule)]
[((connected*b distal-tubule)*e collecting-duct)]
```

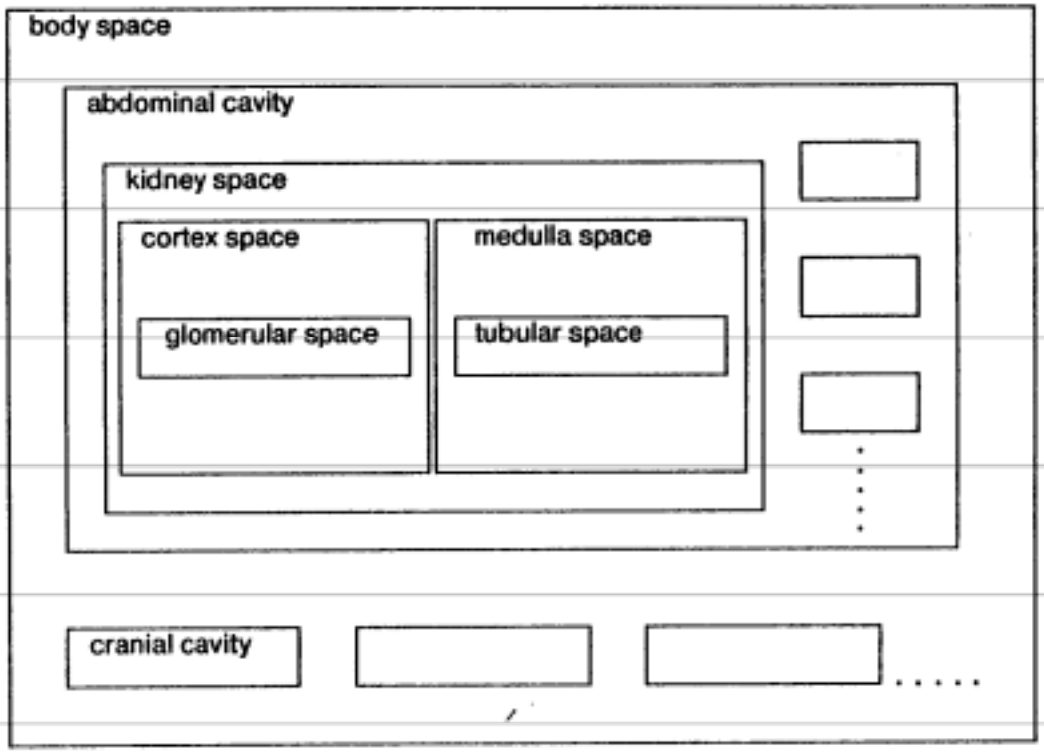


3.1.3 Anatomical Spaces

Various anatomical parts of the body are distributed in different spaces. These spaces are generally isolated from one another by membrane barriers which prevent the free flow of various electrolytes, proteins etc. Thus, the composition of the fluid surrounding organs in a given compartment can be different from that in other compartments. These general characteristics of the compartment can be useful in diagnosis and management of various diseases. Examples of such a compartment are the *cranial-cavity* and the *peritoneal-cavity*. Although the anatomical part-of relation and spatial containment relation are very similar, a distinction between the two must be made. For example, the cortex and the nephron are two different parts of the kidney and the nephron has two parts, the glomerulus and the tubule; however, the glomerulus is contained in the anatomical space of the cortex while the tubule is contained in the anatomical space of the medulla. A graphical representation of this can be seen in figure 9.

Fig. 9. The containment relation

```
[body-space
    [((contains*b body-space)*e cranial-cavity)]
    [((contains*b body-space)*e abdominal-cavity)]
    [((contains*b body-space)*e oropharynx-cavity)]
    [((contains*b body-space)*e thoracic-cavity)]
     .... ]

[abdominal-cavity
    [((contains*b abdominal-cavity)*e stomach-space)]
    [((contains*b abdominal-cavity)*e spleen-space)]
    [((contains*b abdominal-cavity)*e liver-space)]
    [((contains*b abdominal-cavity)*e kidney-space)]
     .... ]

[kidney-space
    [((contains*b kidney-space)*e cortex-space)]
    [((contains*b kidney-space)*e medulla-space)]]

[cortex-space
    [((contains*b cortex-space)*e glomerular-space)]]

[medulla-space
    [((contains*b medulla-space)*e tubular-space)]]
```

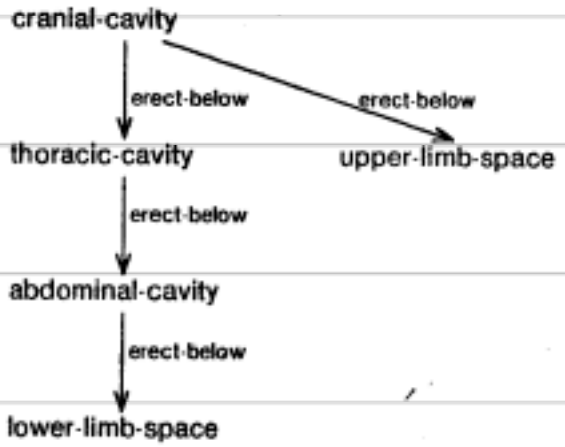


3.1.4 Miscellaneous Gross Anatomical Relations

A few additional anatomical relations are useful in common sense reasoning in medicine. An example of such a relation is the relative positioning of various anatomical spaces in supine position (lying face up in bed), erect position (standing up or ambulatory), etc. The use of this information can be illustrated by the following example. Let us consider a patient with nephrotic syndrome. A common symptom in nephrotic syndrome is periorbital edema (accumulation of fluid under the skin around the eye). In ambulatory patients, the periorbital edema can be observed only in the morning (after the patient has been lying down for some period of time); this accumulation of fluid can gravitationally move into other spaces once the patient has been up and around for some hours in the day. Thus the symptom is observable only in the morning and tends to disappear later in the day. Exactly an opposite effect can be observed in the case of pedal edema (accumulation of fluid in the feet) which tends to appear towards the evening and disappear in the mornings. This information can be used to explain away the absence of pedal edema in an edematous patient who is comatose. This information is encoded in the program with the use of positional relations as shown in figure 10.

We would like to note that the use of the anatomical knowledge in the current implementation of ABEL is limited to the use of anatomic taxonomy. However, we believe that the knowledge described here will be useful for further development of the diagnostic component as well as the therapy and prognosis components of the project.

Fig. 10. Gross anatomical relations

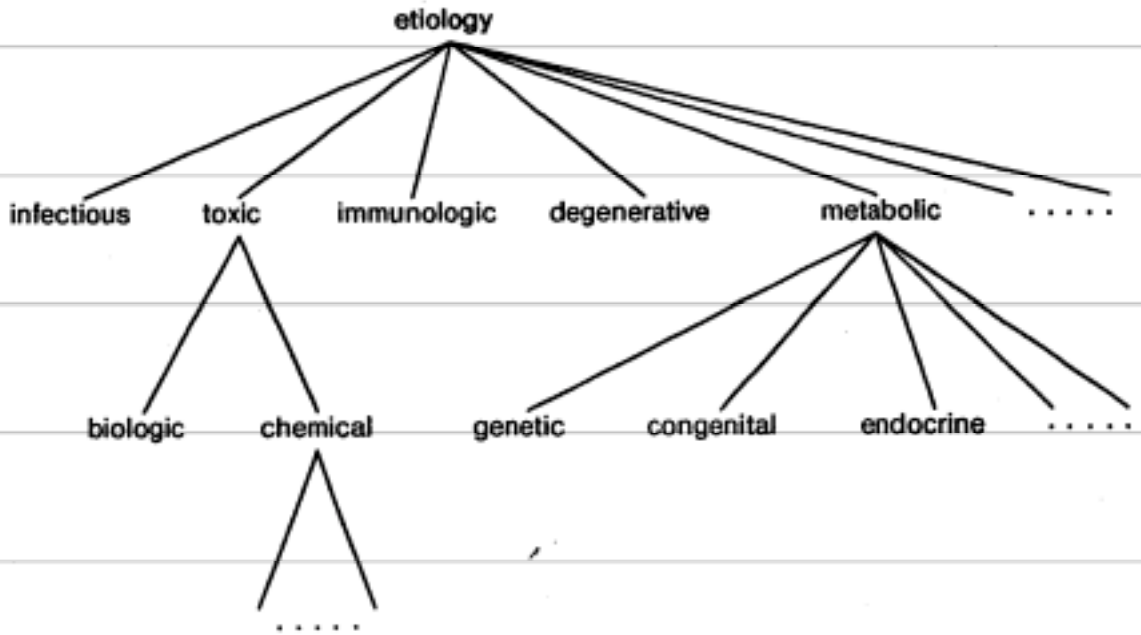


3.2 Etiological Knowledge

Disease categories are primarily organized around the organ systems; e.g., renal diseases, pulmonary diseases, liver diseases. In the previous section we have provided the basic framework of anatomical knowledge needed to provide such a categorization. The diseases of a given organ system tend to produce many symptoms associated with the loss of function of that system. For example, regardless of the cause of renal failure, all the diseases causing renal failure share common symptoms.

Fig. 11. Etiological hierarchy

```
[etiology = (medical-entity*s "etiology")
    [infectious = (etiology*s "infectious")]
    [immunologic = (etiology*s "immunologic")]
    [degenerative = (etiology*s "degenerative")]
    [toxic = (etiology*s "toxic")
        [biologic-toxins = (toxic*s "biologic-toxins")]
      [chemical-toxins = (toxic*s "chemical-toxins")]]
    [metabolic = (etiology*s "metabolic")
        [genetic = (metabolic*s ")]
      [congenital = (metabolic*s ")]
      [endocrine = (metabolic*s ")]]
    . . . . . ]
```



Another important criterion for organizing diseases is the underlying mechanism causing the clinical disorder, i.e., the etiology of the disease. Similar to the anatomical categorization, the diseases with common etiology share symptoms common to the disease mechanism. For example, most infectious diseases cause fever. The taxonomy of etiologies in the program is shown in figure 11.

3.3 Physiological Knowledge

Knowledge about the normal functioning of the body and its adaptive response to abnormalities in body function plays an important role in the understanding and recognition of diseases. The need for this understanding is even more acutely felt in complex clinical settings involving the simultaneous presence of multiple abnormalities. Emphasizing this need, Dr. Jordan Cohen notes:

> *"The recognition of how common mixed disturbances are in complex clinical settings has served to emphasize the value of recognizing the limits of the physiologic response to simple disturbances, because frequently by reference to these limits one can recognize when a complex disturbance involving more than one simple abnormality is present."*
>
> *— New Concepts of Acid-Base Balance [Cohen77, page 1].*

In the physiological component of the program we have concentrated on the knowledge necessary in dealing with fluid, electrolyte and acid-base disorders. The physiological knowledge about fluids and electrolytes in the program deals with fluid compartments of the body and the distribution of body fluids in various fluid compartments, the composition of fluid in each compartment, the space of distribution of solutes, exchange of fluid and electrolytes between these compartments, and the homeostatic mechanism for regulating the quantity and composition of the body fluids.

For example, let us look at the definition of the Serum-Potassium concentration:

```
[body = ((anatomical-entity*s "body")*u patient)
    [body-fluid = ((fluid*s "body-fluid")*u body)
        [ecf = ((fluid*s "extracellular")*u body-fluid)
            [ecf-k = (K*u ecf)
                [serum-k = (concentration*u ecf-k)
                    [low-serum-k = (serum-k*f low)
                        [default*u #v !3.0]
                        [range*u #c !'(between 2.0 3.5)]
                        #s (standard-error*t !1.0,!1.0)]
                    [high-serum-k = (serum-m*f high)
                        .... ]
                    [normal-serum-k = (serum-k*f normal)
                        .... ]]]]]]
```

The above expression defines serum-K (serum potassium) to be the concentration of potassium

ion (K) in the extracellular fluid compartment (ecf), which is one of the components of the body fluid (body-fluid). The serum-K is further categorized as being either low (i.e., [(serum-k*f low)]), normal or high. Each of these categories is also associated with its default value, range and the acceptable amount of variance associated with its value (standard-error, in this case $\pm$1.0). The next example shows the encoding of the normal composition of the lower-GI-fluid. The lower-GI-fluid contains, in addition to water, Na, K, Cl and HCO_3. The quantities of these electrolytes and their variations are further specified in terms of the total quantity of the fluid. For example, the quantity of K is specified to be equal to 40.0 $\pm$ 10.0 meq/L of water in the lower-GI-fluid.

```
[lower-gi-fluid
    [water:u #c (quantity*u lower-gi-fluid)]
    [K:u #c (times*c water:u,!40.0)
         #s (standard-error*t
                (times*c water:u,!10.0),(times*c water:u,!10.0))]
    [Na:u #c (times*c water:u,!110.0)
         #s (standard-error*t
                (times*c water:u,!10.0),(times*c water:u,!10.0))]
    [Cl:u #c (times*c water:u,!80.0)
         #s (standard-error*t
                (times*c water:u,!20.0),(times*c water:u,!10.0))]
    [HCO3:u #c (times*c water:u,!40.0)
         #s (standard-error*t
                (times*c water:u,!10.0),(times*c water:u,!20.0))]]
```

In the previous three sections we have described the anatomical, physiological and etiological knowledge which, along with the temporal characterization, forms a basis for the taxonomic organization of diseases discussed in the next section.

3.4 Disease Knowledge

In studying the organization of medical knowledge about diseases Pople notes

> *"There are two conceptual frameworks that are used to organize medical knowledge, ... One of these employs the concept of causality or pathophysiology to establish a network of interrelated pathological states that might arise in the course of a disease. The other type of structure is the taxonomy of diseases, also called a "nosology", which is used to classify disease entities on the basis of anatomical locus, etiological agent, or other unifying principle."*
>
> — Structuring Medical Diagnosis [Pople81]

This section deals with the use of anatomical, physiological, etiological and temporal knowledge in defining a taxonomic disease hierarchy. With this taxonomic hierarchy in place, we will have completed the study of the basic medical concepts needed in ABEL for the description of disease pathophysiology.

A disease is defined in terms of its anatomical involvement, its temporal characteristic, its etiologic characterization and its pathophysiology. As each of the anatomic, etiologic, and physiologic knowledge is hierarchically organized, the locus of a disease along each of these dimensions can be selected at an appropriate level. A hierarchic organization for the disease definitions can then be derived from these loci.[9] For example, acute renal failure caused by nephrotoxic drugs could be specified as

```
[renal-disease = (disease*anat renal-system)
    [renal-failure = (renal-disease*object (urine-volume*f low))
        [acute-renal-failure = (renal-failure*tempch acute)
            [drug-induced-acute-renal-failure
              = (acute-renal-failure*etiology chemical-toxins)]]]]
```

The example above defines renal-disease to be a disease of the renal-system (anatomical locus). Renal-failure is then defined as a renal disease characterized by low urine output (physiological locus). Acute-renal-failure is defined to be renal-failure with an acute temporal characteristic, and finally, the drug-induced-acute-renal-failure is defined to be acute-renal-failure of chemically-toxic etiology. Note that each step of the above definition defines a disease which is further specialized by one of its primary characterizations. This provides a more specific placing of the diseases in the taxonomic hierarchy. In the next example we show how the disease definitions can be taxonomically organized along a single locus:

```
[GI-disturbance = (disturbance*anat gastrointestinal-system)
    [lower-GI-disturbance =  (GI-disturbance*anat lower-gi-tract)]
        [disturbance-of-colon = (lower-GI-Disturbance*anat colon)]]]

[renal-disease = (disease*anat renal-system)
    [nephritis = (renal-disease*anat nephron)
        [glomerular-nephritis = (nephritis*anat glomerulus)]]]
```

As can be seen from the above two examples, the basic medical knowledge about anatomy, etiology etc., provides us with a framework for describing and organizing the disease hierarchy. We believe in the need for such a knowledge structure in the organization of any medical consulting program capable of expert level performance. However, we must note that this development is tentative and the details of the knowledge representation described above are likely to evolve considerably as its use in the diagnostic and therapeutic algorithms is better understood.[10]

---

9. Each hierarchy, such as the anatomic taxonomy, provides us with a tree structured partial order. The tree structure for the disease definitions is then derived from these partial orders.

10. In the current implementation of ABEL, this knowledge is used only for grouping different findings and diseases in diagnostic problem solving. However, the knowledge representation described in this chapter is capable of supporting a substantially wider variety of uses.

In the next section we will study the representation mechanisms for describing the causal (pathophysiological) knowledge relating different diseases.

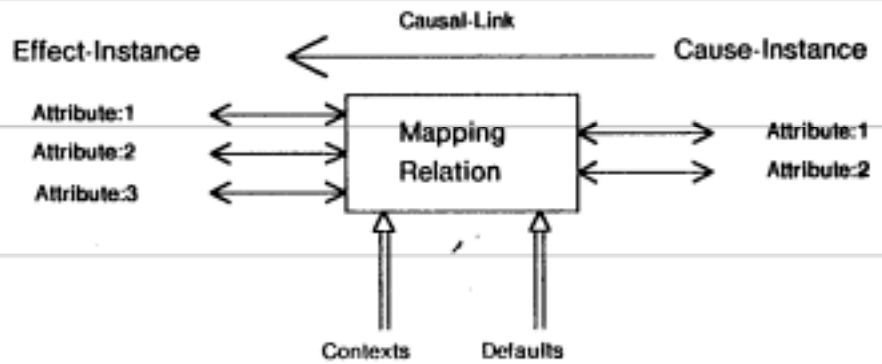
3.5 Causal Link

A causal link specifies the cause-effect relation between the cause (the antecedent) and the effect (the consequent) states. In the previous generation of programs (i.e., PIP, INTERNIST and GLAUCOMA), causal relations were described by links specifying the type of causality (e.g., *may-be-caused-by*, *complication-of*, etc.), and a number or a set of numbers representing in some form the likelihood (conditional probability), importance, etc., of observing the effect given the cause or vice versa. We believe that this simple representation of the relation between states is inadequate. The form of presentation of an effect and the likelihood of observing it depend upon various aspects of the presentation of the cause instance such as severity and duration, as well as on other factors in the context in which the causal phenomenon is manifested (such as the patient's age, sex and weight, and the current hypothesis about the patient's illness). To illustrate this, let us consider a (simplified) causal relation between diarrhea and dehydration. A rule-based description of this causal relation can be specified as follows:

```
IF diarrhea is severe,
    and its duration is greater than two days,
  THEN
    IF the patient has not received fluid replacement therapy
      THEN the patient is likely to have moderately severe dehydration
      ELSE the patient may have mild dehydration
```

From the above simple example, it is apparent that the conditional probability of observing dehydration and its severity and duration depend on the severity and duration of diarrhea and the fluid replacement therapy. Even this simplified example clearly demonstrates the need for information on how a cause relates to an effect, as well as other contextual information

Fig. 12. Schematic description of a causal link



influencing the causal relation. To capture this information, the description of a causal link has associated with it a *multivariate relation* between attributes of the cause and the effect, the context, and the assumptions which constrain the causal relation. A schematic description of a causal link and its representation in the data-base are shown in figure 12.

An example of the causal relation between total extracellular stores of potassium (ecf-K) and its serum concentration (serum-K) is described below.

```
[((caused-by*b ecf-k)*e serum-k)
    [context:1 #v total-ecf-water]
    [source:u
        [value*u #c (times*c (value*c (value*u destination:u)),
                             (value*c context:1))]
      [start-time*u #c (value*c (start-time*u destination:u))]]
    [destination:u
        [value*u #c (quotient*c (value*c (value*u source:u)),
                                (value*c context:1))]
      [start-time*u #c (value*c (start-time*u source:u))]]]
```

The causal relation between ecf-k stores and serum-k is specified by a causal link with cause (source) ecf-K and effect (destination) serum-k. The mapping relation describing this link is divided into two parts. The first part is associated with the source of the link and describes procedures for computing the attributes of the source (cause) given the attributes of the destination (effect), and the second part is associated with the destination given the attributes of the source. For example, the total quantity of potassium in the extracellular compartment (value of the source) is characterized as being the product of the quantity of the extracellular water (value of the context, total-ecf-water) and the concentration of the potassium ion in it (destination, serum-k).

Strictly speaking, it would not be appropriate to call all relations of this kind "causal," as some of the relationships are more matters of definition or association than cause. A more rigorous analysis, perhaps following the lines of [Rieger77], would further distinguish potential cause from actual, enabling conditions from true causations, etc. Such an expansion would, however, be orthogonal to our present argument, that any such link must connect several aspects of its source and destination.

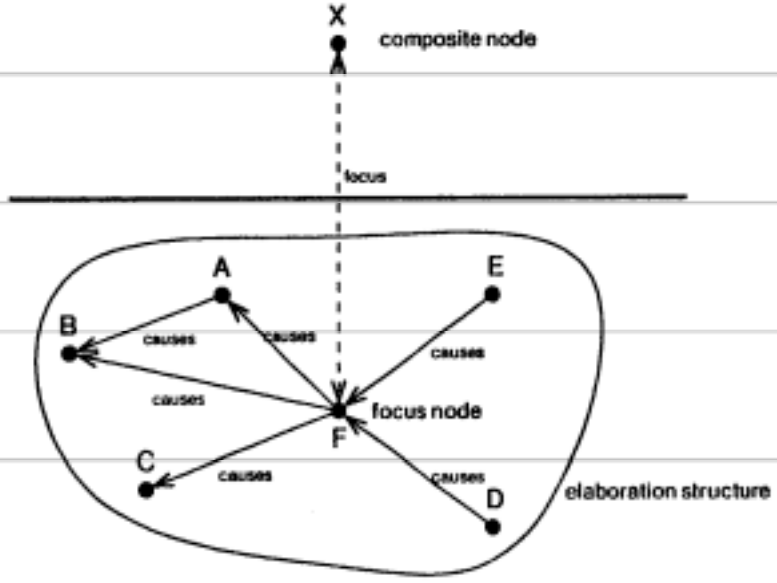
3.6 Multi-Level Causal Description

Medical knowledge about different diseases and their pathophysiology is understood to varying degrees of detail. Our understanding of medical expert reasoning also suggests that an expert physician may have an understanding of a difficult case in terms of several levels of detail. For example, *"serum creatinine concentration of 1.2 mg per cent"* is at a distinctly different level

than *"high serum creatinine"*,[11] and *"lower gastrointestinal loss"* than *"salmonellosis"*. For our program to reason at a sophisticated level of competence, it will need to share such a range of representations. In order to be effective the program must be able to describe the problem briefly yet still be able to take low level detail into consideration. We have attacked this problem by representing the program's medical and case-specific knowledge at five distinct levels of detail,[12] ranging from a pathophysiological level to a phenomenological level of knowledge.

The patient description developed here provides us with the ability to describe the patient's illness at various levels of detail. Each level of the description can be viewed as a semantic net describing a network of relations between diseases and findings. Each node represents a normal or abnormal state of a physiological parameter and each link represents some relation (causal, associational, etc.) between different states. A state in the system is represented as a node in the causal network. Associated with each node is a set of attributes describing its temporal characteristics, severity or value, and other relevant attributes. A node is called *primitive* if it does not contain internal structure and is called *composite* if it can be defined in terms of a causal network of states at the next more detailed level of description. One of the nodes at that more detailed level is designated as the *focus node* and the causal network is called the *elaboration*

Fig. 13. Schematic description of the node structure



11. A serum creatinine of 1.2 mg per cent can be interpreted in more than one way. For example we can assume this to be normal for a muscular male patient. But, for a average built female patient this could be an early indication of loss of as much as 1/3 of the renal function.

12. The number *five* does not have any medical or cognitive significance; it was chosen for purely engineering reasons.

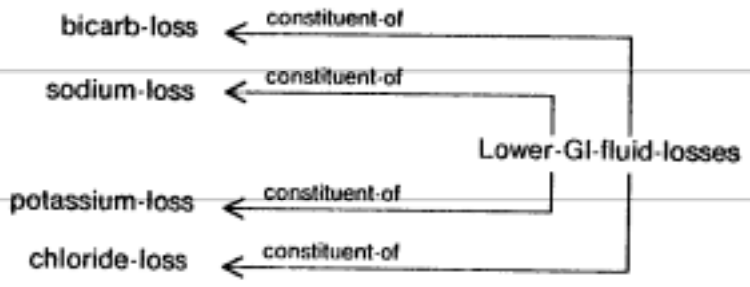
structure of the composite node. Figure 13 shows a schematic of the elaboration structure for a composite node labeled X. Nodes A through F and links between them form X's elaboration structure. Node X and F are connected together by a *focus link* making F the focus of the elaboration structure. The focus node identifies the essential part of the causal structure of the node above it. The collection of focal nodes acts to align the causal networks represented by different levels of the PSM. We note that very often a composite node and its focal description at the next level share the same name.[13] Nodes that do not play a role as the focal definition of any node at a higher level are called *non-aggregable nodes*. They represent a detailed aspect of the causal model which is subsumed under other nodes with different foci at less detailed levels of description.

To illustrate the description of a state at various levels of aggregation, let us consider the electrolyte and acid-base disturbances that occur with salmonellosis, which causes excessive loss of lower gastrointestinal fluid (lower GI fluid loss). In comparison with plasma, the lower GI fluid is rich in bicarbonate (HCO3) and potassium (K) and is deficient in sodium (Na) and chloride (Cl). The composition of lower gastrointestinal fluid and plasma are shown in figure 14. The loss

Fig. 14. Comparision of lower GI fluid and of plasma

	Lower GI fluid	Plasma	
Na	100 - 110	138 - 145	mEq/L
K	30 - 40	4 - 5	mEq/L
Cl	60 - 90	100 - 110	mEq/L
HCO3	30 - 60	24 - 28	mEq/L

Fig. 15. The loss of electrolytes in lower GI fluid



13. This is typical in English, where the level of detail of place names, for example, is often obtained from context and not encoded in the name used.

of lower GI fluid leads to the loss of corresponding quantities of its constituents as shown in figure 15. Therefore, an excessive loss of lower GI fluid without adequate replacement of fluid and electrolytes leads to a net reduction in the total quantity of fluid in the extracellular compartment (called *hypovolemia*). Because the concentration of K and HCO3 in the lower GI fluid is greater than in the plasma, there is also a corresponding reduction in the serum concentration of K (called *hypokalemia*) and HCO3 (called *hypobicarbonatemia*) in the extracellular fluid. Finally, because the concentration of Cl and Na in the lower GI fluid is lower than that in the plasma, there is corresponding increase in the concentration of Cl (called *hyperchloremia*) and Na (called *hypernatremia*) in the extracellular fluid. A graphic representation of this information at the next higher level of aggregation is shown in figure 16. Figure 17 shows the aggregation of this information along with some additional causes and consequences of lower GI loss at the next more aggregate level of detail. Hypobicarbonatemia is interpreted as metabolic acidosis at the next higher level of detail. Note that the hypernatremia and hyperchloremia have not been encoded at this level.[14] The hyperchloremia was not encoded because it is not clinically significant. The hypernatremia, however, is not encoded because it is not a common finding in the presentation of lower GI loss. The lower GI loss at this level is a non-aggregable state and therefore does not have a focal aggregation at the next level above. Figure 18 shows the

Fig. 16. Consequences of lower Gi loss described at next higher level

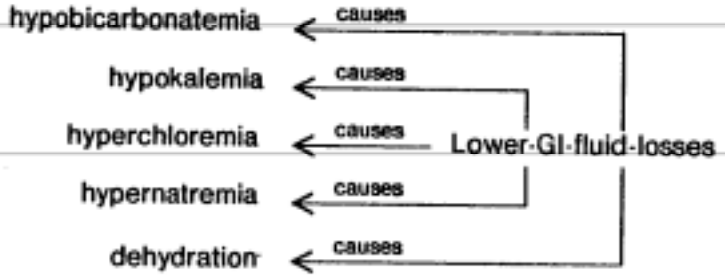
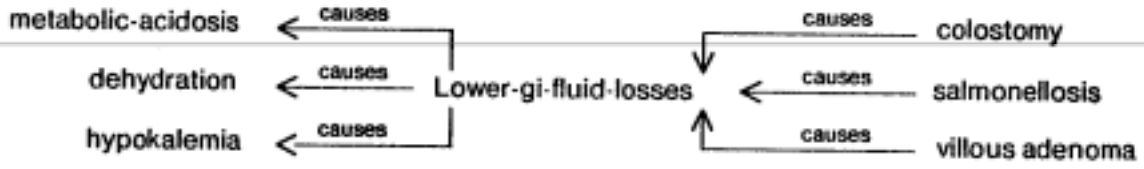
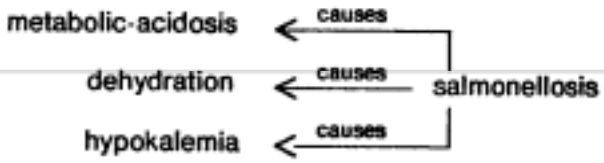


Fig. 17. Lower Gi loss expressed at an intermediate level



14. The causal knowledge described here is encoded by hand, and represents the program's general medical knowledge. A similar multi-level description built by the program to describe a specific patient illness will be discussed in chapter 4.

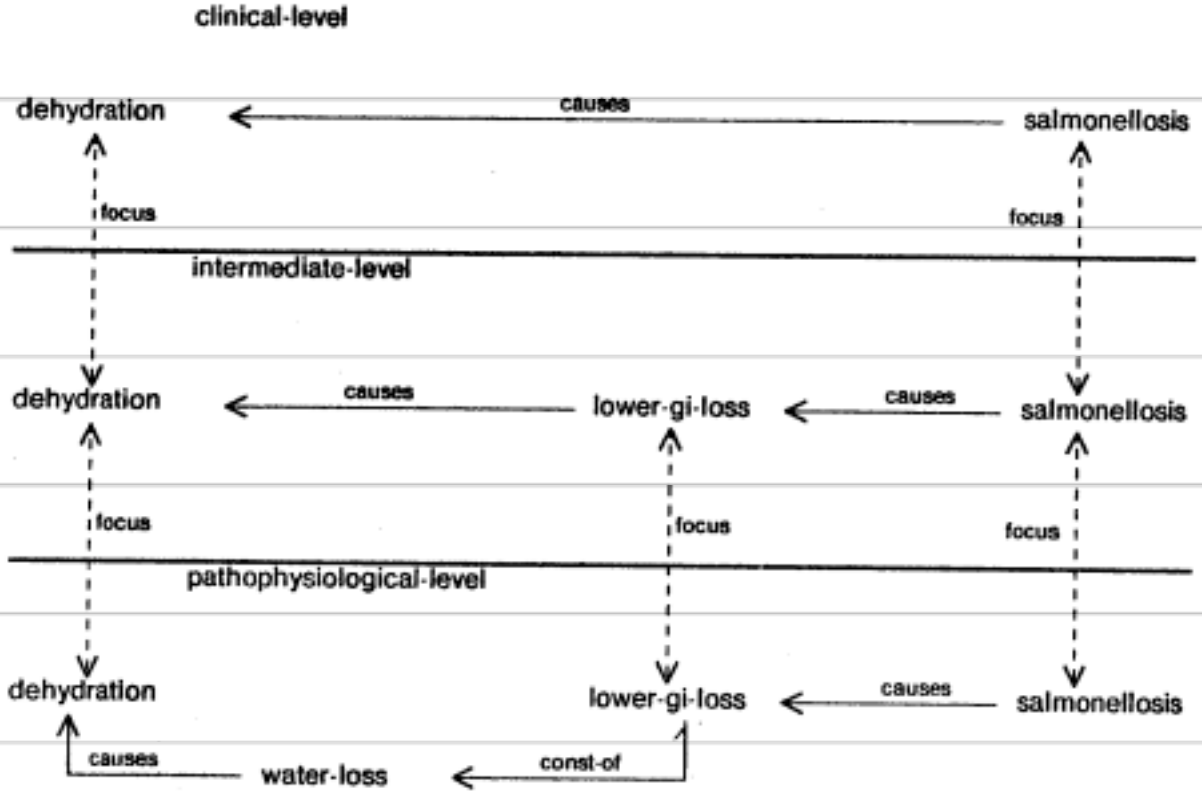
Fig. 18. Salmonellosis and its consequences expressed at the clinical level



description of the aggregate effects of salmonellosis (one of the causes of lower GI loss).

Links can be categorized into two types, as nodes are: the *primitive* links and the *composite* links. To illustrate the concept of elaborating causal links to form a causal chain, let us consider the causal relation between salmonellosis and dehydration shown in figure 19. The causal mechanism of dehydration caused by salmonellosis can be elaborated as follows: salmonellosis causes lower GI loss, which in turn causes dehydration. Expressed at the next level of greater detail, the lower GI loss leads to water loss which results in reduction in the extracellular volume.

Fig. 19. Layered description of link: salmonellosis causes dehydration



The state of reduced extracellular volume is called dehydration.

Because the causal relations specified by links are not guaranteed to be true under all circumstances (they represent strong associations, not logical truth), the validity of deductions degrades with every additional intermediate link. That is, a causal pathway containing a large number of links is less likely to be valid than one using only a few links. Therefore, in order to explore a large diagnostic space, we must reduce the lengths of commonly occurring chains of causal relations. One way of achieving this is through the multi-level description proposed in this chapter. The multi-level description scheme allows us to aggregate the diagnostic space to a level where each link represents an aggregate causal phenomenon covering large distances and thus minimizing the possibility of error in the deduction.

However, the multi-level description proposed above can not solve this problem completely. For example, there are situations where all the intermediate nodes in a given causal chain cannot be suppressed due to limited number of levels of description. Stated differently, because of the fixed number of levels in the multi-level description, the programs ability to aggregate causal description is limited. To overcome this problem we introduce the notion of a compiled link which represents a causal pathway.[15] The compiled links provide us with the ability to selectively explore commonly occurring causal paths more deeply than others without degrading the quality of deduction. This also provides us with the additional ability to activate[16] nodes which are not immediate neighbors of the node under consideration. For example, severe salmonellosis causes dehydration sufficient to cause hypotension (lowering of blood pressure). This fact can be represented in the data base by the compiled causal link as shown in the figure 20.

An important function of diagnostic reasoning is to relate causally the diseases and symptoms observed in a patient. These causal relations play a central role in identifying clusters that can be meaningfully aggregated in developing coherent diagnoses. The presence or absence of a causal relation between a pair of states can change their diagnostic and prognostic interpretations. Therefore, the system should and does have the capability of hypothesizing the presence or absence of a causal relation. This is the primary reason why links are considered

Fig. 20. Compiled link

```
[((caused-by*b salmonellosis)*e hypotension)
    #path [((caused-by*b salmonellosis)*e dehydration)],
          [((caused-by*b dehydration)*e hypotension)]]
```

15. During the exploration of a diagnostic space, traversing a compiled link is equivalent to traversing the predefined path associated with the compiled link in a single step.
16. This is similar to "triggering" a disease in PIP.

objects in their own right rather than simply an ordered pair of states.

In this section we have described the representation of the anatomic, physiologic and etiologic medical knowledge around which the disease taxonomy and pathophysiology is organized. We have also discussed a multi-level hierarchic description of causal knowledge. In the next section we will study the operations that use this knowledge in describing the individual patient's illness by patient-specific instantiation of relevant medical knowledge and by combining the effects of multiple disease phenomena.

4. Structure Building Operations

This chapter deals with the operations for building causal models (called PSMs; the Patient Specific Models) that can explain a patient's illness. A PSM is created by instantiating portions of ABEL's general medical knowledge. The creation of a PSM requires establishing and maintaining a correspondence between the medical knowledge and the observations, so that the information from each source can be added together. Much of the meaning of an observation depends on the context provided by the PSM; conversely, the PSM is created by assimilating many observations. As the PSM is multi-level, this assimilation requires the ability to summarize a detailed description into aggregate global summaries and the ability to disaggregate a summary into detailed description. This can be achieved based on the observation that the cognitive maps can be organized around *local landmarks* (focal nodes described in the previous chapter). The local topology surrounding a landmark can be described relative to the landmark and the landmarks then related to each other to construct the next level summary. It is possible to maintain sufficient mapping between adjacent levels for efficient use of this map for problem solving, if the summarization is carried out gradually using small steps, and in strict adherence to the principle of locality. Finally, note that detailed descriptions are likely to be much more accurate than global ones; detailed physiological descriptions tend to be much more accurate than global syndromic descriptions. Furthermore, local inconsistencies are easy to detect and correct, and are usually attributed to particular observations. Global inconsistencies, however, are much more difficult to pin down and are usually due to systematic errors in the interpretation of local observations and unwarranted extensions of local observations. Therefore, in building the PSM we interpret observations at the most detailed level possible and resolve inconsistencies arising at an aggregate level by using more detailed levels.

4.1 Structure of a PSM

A PSM is a multi-level causal model, each level of which attempts to give an account of the program's understanding of the patient's case. Each PSM contains all the diseases and findings that have been observed or concluded in a given patient along with hypothesized diseases, findings, and their interrelationships, which together form a *coherent* explanation. Within each PSM, the known and hypothesized diseases, findings and their interrelationships are mutually complementary, while the alternate PSMs provide alternate explanations which are mutually exclusive and are competing to explain a patient's illness. Note that considering a PSM as a hypothesis for a patient's illness avoids the problem faced by the previous programs which considered each possible individual disease as a complete hypothesis, as discussed in chapter 1.

The PSMs are implemented using a *Patient Specific Data structure* (called PSD). The PSDs are organized in a tree. The PSD in the root position of the tree contains observed findings and the structure common to all the PSMs. Differing interpretations of the observed findings are described by creating inferior PSDs each containing incremental changes (additions as well as

deletions) to their superior PSD. Each PSD in the tree inherits from its superiors all the structure present in them except that which is explicitly deleted.[17] The structure visible from each leaf node of the PSD tree corresponds to an individual PSM. The list of PSMs at any given instant of diagnosis is called causal hypothesis list (CH-list).

Each PSD is implemented as a record structure containing a record for each level of description, a list of deleted elements and a pointer to the superior PSD containing it. The record structure of a PSD is:

(⟨level-0⟩,⟨level-1⟩,⟨level-2⟩,⟨level-3⟩,⟨level-4⟩,⟨deleted-elements⟩,⟨superior⟩)

The description of each level is implemented as a record structure consisting of a set of nodes, a set of links describing the relations between the nodes at the given level, and two sets of focal links connecting the description at the current level to the description at the adjacent lower and upper levels. The record structure of a level is:

(⟨nodes⟩, ⟨links⟩, ⟨focal-links up⟩, ⟨focal-links down⟩)

The tree structure of the PSDs allows different PSMs to share structure common between them, providing efficiency in storage as well as in comparison of the structures of different PSMs. All the new information received is always added to the root PSD, the PSD common to every PSM. However, if this new information can be explained in more than one way in the context of a given PSM, the leaf PSD corresponding to the PSM is expanded to represent each of these explanations separately.

The PSMs are created and augmented using structure building operations described in this section. These operations are *initial formulation* to create the initial set of PSMs from the presenting complaints and lab results, *aggregation* to summarize the description at a given level of detail to the next more aggregate level, *elaboration* to disaggregate the description at a given level to the next more detailed level, *projection* to hypothesize associated findings and diseases suggested by states in the PSM, and *constituent summation* and *decomposition* to evaluate the combined effects of multiple etiologies and to evaluate the unaccounted components of partially accounted findings.

4.2 Initial Formulation

One of the most startling observations uncovered from the study of clinical problem solving is the physician's response to the presenting complaints [Pauker76, Elstein77, Kassirer78].

---

17. The use of PSD tree is similar to the use of a "context tree" in CONNIVER [Sussman72].

> *"The most striking aspect of the history-taking process revealed by the protocols is the sharp focus of the clinicians' problem-solving behavior. The subjects generated one or more working hypotheses early in the history-taking process when relatively few facts were known about the patient. At a time when the clinician was aware only of the age, sex, and presenting complaints of the patient, he often immediately introduced a hypothesis, ...*
>
> *The process of hypothesis activation dominated the early part of the diagnostic session as the physicians searched for some explanation of the findings and for a context in which to proceed. Later in the session the emphasis was on hypothesis evaluation rather than hypothesis activation."*
>
> — Clinical Problem Solving [Kassirer78, pages 249 - 250]

It is useful for a program to separate, like clinicians, the initial formulation of the diagnostic problem from subsequent revisions in the diagnostic alternatives. The patient specific information available at the initial phases of diagnosis is generally limited to a few nonspecific complaints. It does not provide sufficient context for a data-driven problem solver designed to perform optimally during later stages of diagnosis. Thus, failure to recognize the differences between the initial and the subsequent stages of diagnosis may result in an unfocused diagnostic inquiry with many irrelevant questions until sufficient information can be gathered for establishing a context for an orderly inquiry. The program presented here makes such a distinction. However, substantial improvement in the initial formulation of the diagnostic problem will be required before this distinction can be effectively exploited.

When provided with the initial findings and a set of serum electrolyte values, ABEL constructs a small set of PSMs, using the following steps. First, it analyses the electrolytes and formulates the possible single or multiple acid-base disturbances that are consistent with the electrolyte values provided. It then selects from them a small set which is consistent with the initial findings. Next, it generates a pathophysiological explanation of the electrolytes based on each of the proposed acid-base disturbances. This is achieved by elaborating known clinical information to the pathophysiological level, where its relationships to the laboratory data is determined by projecting the unique causes and definite consequences of every node. The program then summarizes these pathophysiological descriptions to the clinical level by repeated application of the aggregation operations. This process results in the initial description of the patient being built at every level of detail. These descriptions form the program's initial hypotheses, and are later modified as new information becomes available. Note that each of the mechanisms, *aggregation*, *elaboration* and *projection* are used in the initial formulation of the PSM.

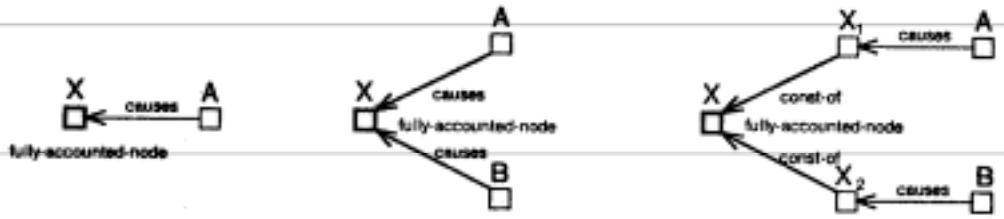
4.3 Some Definitions

This section introduces the naming conventions and definitions for describing types of nodes and their internal structures in a PSM.

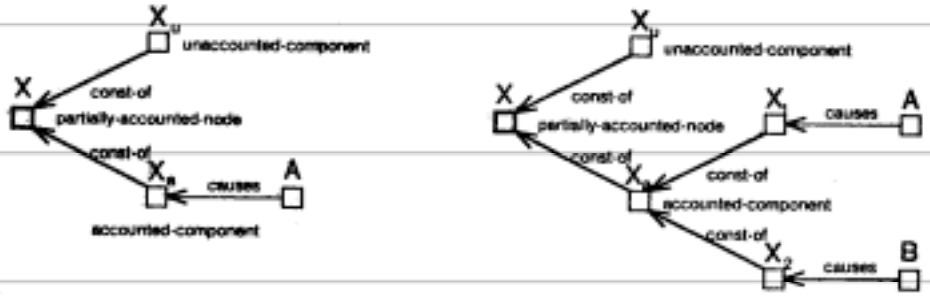
Fig. 21. Node types

(a) Fully Unaccounted Node

(b) Fully Accounted Node



(c) Partially unaccounted nodes



Prime Antecedent: A node is a prime antecedent if it does not have any link coming into it, i.e., it does not have any cause.

Ultimate Etiology: a prime antecedent is called ultimate etiology if it represents a diagnosis, i.e., a disease which does not need to be explained in the domain of application.

Unaccounted Node: a prime-antecedent which is not an ultimate etiology. It is called unaccounted because it needs to be accounted for in terms of ultimate etiologies for the diagnosis to be complete.

Fully Accounted Node: A node is said to be fully accounted if all its causes are present.

Partially Accounted Node: A node is said to be partially accounted if only some of its causes are present.

Accounted Component: the accounted component is a node which describes the sum total of the effect of all the known causes of a partially accounted node.

Unaccounted component: the unaccounted component is a node which describes that part of a partially accounted node that still remains to be accounted for. In other words it represents the difference between the partially accounted node and its accounted component.

Predecessor path: a predecessor path of a node is defined to be any causal path (with one or more links) leading into the node.

Some of these structures are also illustrated in figure 21. Figure 21(a) shows a fully unaccounted node X. Figure 21(b) shows three possible structures for fully accounted nodes. The first structure shows a fully accounted node X and its cause A. The second and third structures show a fully accounted node X with two causal predecessors A and B which together account for X. In the third structure X is a primitive node and therefore the components of X (i.e., X_1 and X_2) accounted for by each of its causes are explicitly instantiated. However, in the second structure X is a fully accounted for composite node, therefore, A and B are directly connected to X suppressing the component structure present at the greater levels of detail. Figure 21(c) shows two possible structures for partially accounted node X. X is decomposed into an accounted component X_a, and an unaccounted component X_u. X_u is an unaccounted node with structure similar to case (a) and X_a is a fully accounted node and has structure similar to case (b).

4.4 Aggregation

The aggregation process is used to summarize the description of the patient's illness at a given level to the next more aggregate level. This summarization of the causal network is achieved by identifying nodes (called *focal nodes*) which can serve as *landmarks*, summarizing each focal node and its surrounding causal relationships at the next more aggregate level (called *focal aggregation*), and by summarizing the chain of causal relations between nodes by a single causal relation between the initial cause and the final effect nodes (called *causal aggregation*).

4.4.1 Focal Aggregation

In aggregating a causal network we must first identify the nodes that form the focal points around which the causal phenomenon can be summarized. Consider a partially-constructed PSM in which some nodes at a detailed level have been instantiated. A node is a focal node if the following three conditions are satisfied. (1) In the medical knowledge-base this node is the focus of the elaboration structure of at least one node at the next more aggregate level. (2) In the PSM

at least one such higher level node already exists, or can be instantiated. (3) The aggregation is not inconsistent with the existing structure of the PSM. If the aggregate node does not exist, then both it and the focal link are instantiated. If the aggregate node exists, the focal link connecting the two is instantiated and the profiles of the focus and the aggregate nodes are updated using any additional information that can be inferred by this connection. Finally, if more than one possible candidate for aggregation is consistent with the causal structure above, the focal aggregation process is deferred until additional information can be obtained to resolve this ambiguity.

4.4.2 Causal Aggregation

Once we have determined the focal aggregations for nodes at a given level of detail we need to determine the causal relations among these aggregate nodes. This is achieved using causal aggregation. The process of causal aggregation takes a node and its causes and aggregates the relation between them according to one of three rules. First, if the node has no causal predecessors or if none of the causal paths leading into the node (*predecessor paths*) have an aggregable node, then the focal aggregation of the node does not have any causal predecessors. The focal aggregation node then is either an ultimate etiology or is an unaccounted node and no new edges need to be added to the aggregation. Figure 22 shows two examples of causal aggregation of fully unaccounted nodes. The first example shows causal aggregation of low-serum-K-1. Focally aggregating we instantiate hypokalemia-1. Next, we follow the predecessor

Fig. 22. Causal aggregation: fully unaccounted node

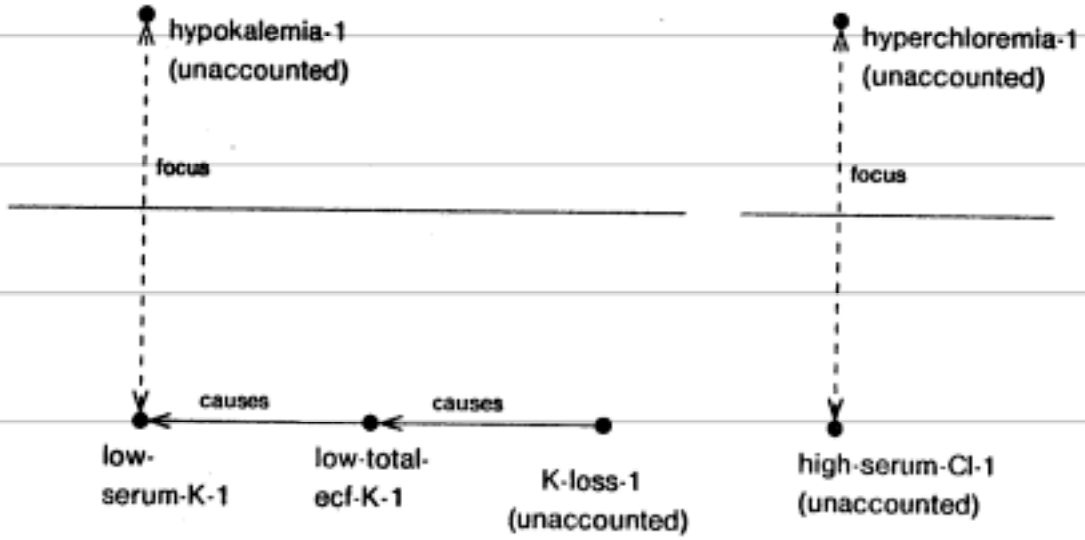
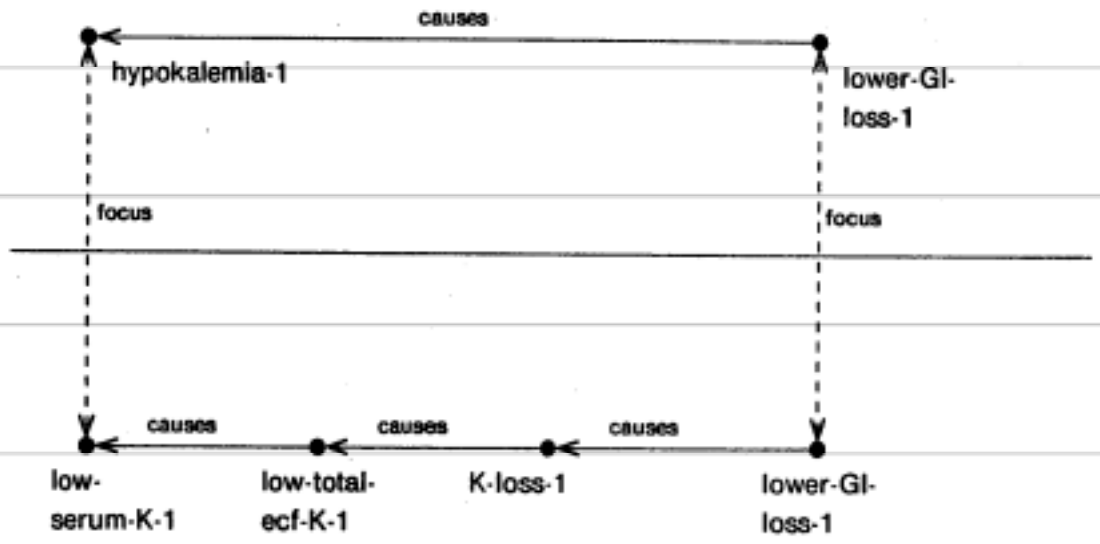


Fig. 23. Causal aggregation: fully accounted node



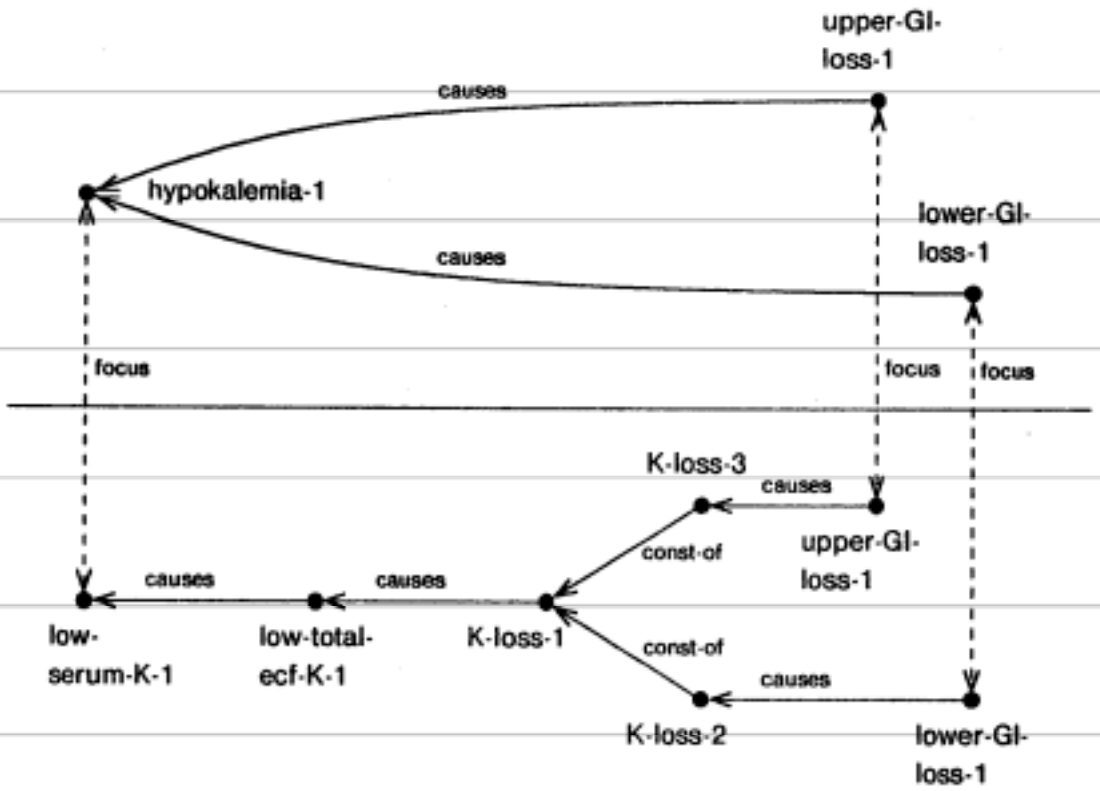
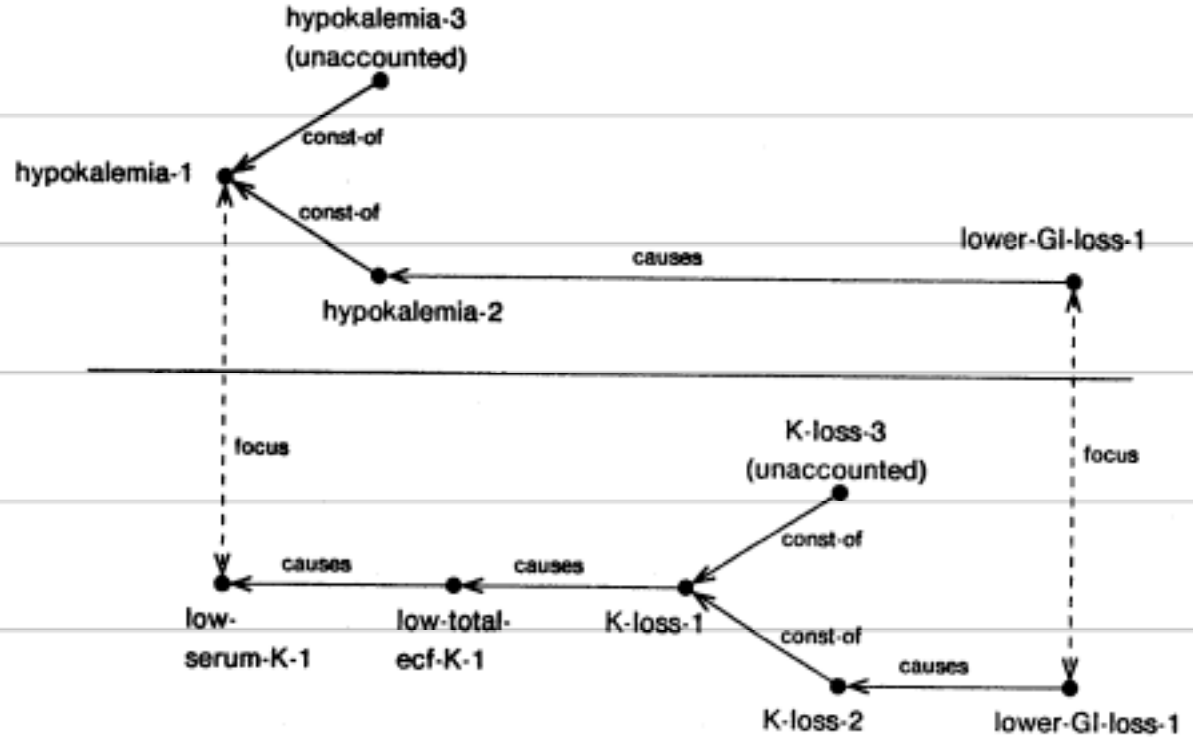


Fig. 24. Causal aggregation: partially accounted node



path of low-serum-K-1 in search of an aggregable node. As low-total-ecf-K-1 and K-loss-1 are not aggregable nodes, this search fails, and no additional structure is created. However, as the predecessor path terminates in an unaccounted node, the focal aggregation of low-serum-K-1, hypokalemia-1, is marked unaccounted. The second example shows high-serum-Cl-1. As high-serum-Cl-1 does not have any predecessor, its focal aggregation, hyperchloremia-1, does not have any causal predecessor. Furthermore, as high-serum-Cl-1 is unaccounted, hyperchloremia-1 is also unaccounted.

Second, if every predecessor path has a node with a focal aggregation then the focal aggregation of the node is fully accounted for. The causal aggregation is achieved by creating a causal link between the focal aggregation of the node and the first focal aggregation in each path. Figure 23 shows two examples of causal aggregation of fully accounted nodes. In the first example the low-serum-K-1 has one predecessor path and that predecessor path contains an aggregable node, lower-GI-loss-1. Therefore, low-serum-K-1 is a fully accounted node, and its causal aggregation is achieved by focally aggregating lower-GI-loss-1 and causally connecting hypokalemia-1 to it. In the second example, low-serum-K-1 has two predecessor paths, each containing an aggregable node. The causal aggregation is achieved by focally aggregating each of these two aggregable nodes and then causally connecting hypokalemia-1 to them as shown.

Finally, if only some of the predecessor paths have nodes with focal aggregations then the focal aggregation of this node is partially accounted for. The causal aggregation is achieved by decomposing the node into two components: (1) the *accounted component*, due to paths which have some focal aggregation, and (2) the *unaccounted component*, due to paths that do not. The focal aggregation of the node is then decomposed based on the decomposition at the present level and the two cases are treated as described above. Any new information that can be derived from the addition of causal links in the PSM is used to update the profiles of nodes involved in aggregation. Figure 24 shows an example of causal aggregation of a partially accounted node. In this example one of the two predecessor paths of low-serum-K-1 contains an aggregable node, lower-Gi-loss-1, we focally aggregate this node. The other predecessor path terminates in K-loss-3, an unaccounted node. Next, we compute the component of low-serum-K-1 that can be accounted for by lower-Gi-loss-1 and the component that remains unaccounted for because of the unaccounted K-loss-3. Then we compute the mapping of these two components at the next level of aggregation and instantiate hypokalemia-2 (the component accounted for by lower-Gi-loss-1) and hypokalemia-3 (due to unaccounted K-loss-3). We then causally connect hypokalemia-2 to lower-Gi-loss-1 and mark the hypokalemia-3 as being unaccounted for.

4.5 Elaboration

Elaboration is the dual of the aggregation operation described above. It is used to disaggregate the description of a causal network at a given level to the next more detailed level. In other words, given a summary description of a causal phenomenon, it provides a more detailed description consistent with the summary. This is achieved by instantiating the focal description of each composite node (called *focal elaboration*) and by instantiating the causal pathway between these detailed nodes corresponding to each causal link at the aggregate level (called *causal elaboration*). If the causal pathway being instantiated interacts with other causal paths in the PSM, the combined effects of the multiple causality are computed using component summation. The combined effects of this summation can then be aggregated upwards to reflect the better understanding of the causal phenomenon at the higher levels of aggregation. This is one mechanism where two aggregate phenomena may become linked, through the interaction of their detailed descriptions.

In summary, the focal aggregation and elaboration create mappings between nodes across different levels, and causal aggregation and elaborations create mappings between causal links across different levels.

4.5.1 Focal Elaboration

To elaborate a causal network we identify the nodes in the network that have been used as summary descriptions, establish their references at the next more detailed level, and establish additional nodes and links at the detailed level to describe the phenomenon described in the

aggregate network. The operation of focal elaboration deals with the first two of the three steps mentioned above.

A node can be focally elaborated if it is a composite node, and if a node corresponding to its focus already exists or can be instantiated in the PSM. If the focus node does not exist, then both it and the focal link are instantiated. If the new node is inconsistent with the detailed level, the detailed level is modified to re-establish overall consistency. If the focus node exists and is consistent, then the focal link connecting the two is created and the profiles of the node and its focus are updated using any additional information that can be inferred by this symbiosis. Finally, if more than one possible candidate for focal elaboration is consistent with the causal structure above, the focal elaboration process is deferred until additional information to resolve this ambiguity can be obtained.

4.5.2 Causal Elaboration

Causal elaboration is used to determine the causal relations between nodes at a detailed level based on the causal relations between the nodes at the next more aggregate level. Causal elaboration is centered around the composite causal link and the chain of causal links that describe each composite causal link. To elaborate a composite link, the program matches the causal path associated with the link, against existing paths in the PSM. If some part of this pathway is not present, the program recurs on each missing link in the pathway (starting from the focus node of the cause) until the link being elaborated is a primitive. When the link being elaborated is primitive it is instantiated under one of the following conditions.

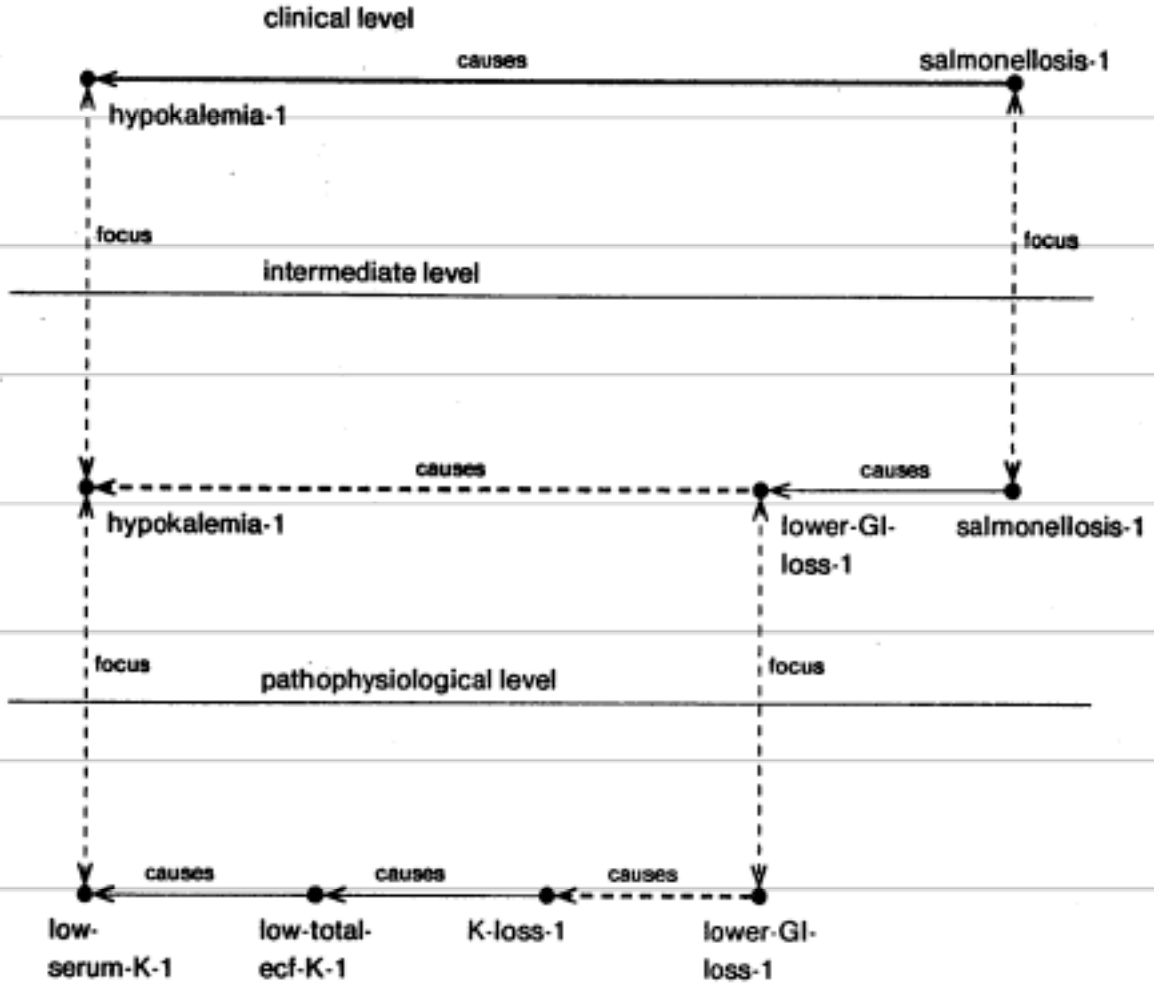
(1) If the effect node is not present in the PSM, the effect node and the link are instantiated.

(2) If the effect node is present, and the constraints on the link are satisfied and it is not causally inconsistent, then the link is instantiated connecting the cause and the effect node.

(3) If the effect node is present but is partially or fully accounted by some other cause, the effects of this additional cause are combined with the existing structure using the component summation and decomposition operation and this combined effect is propagated further as needed.

The program also updates the profiles of the nodes in the causal pathway using any additional information that can be inferred by addition of the pathway. Finally, the aggregation operation is used to revise the description of the next more aggregate level to reflect the addition of the causal pathway.

Fig. 25. An example of the elaboration process



This process is illustrated with the help of the simple example shown in figure 25. Let us consider a patient with hypokalemia and salmonellosis. For the example, let us also assume that by some reasoning process we have established a causal link between salmonellosis and hypokalemia. The elaboration operation can then be used to establish this relation at more detailed levels. The pre-existing structure in the PSM is shown in solid lines, the link being causally elaborated (between hypokalemia and salmonellosis) is shown in solid bold and the links added by the process of elaboration are shown in bold broken lines. The elaboration process attempts to match the causal path corresponding to the link between salmonellosis and hypokalemia at the next level of detail, namely, salmonellosis —causes—> lower-GI-loss —causes—> hypokalemia. The link between salmonellosis and lower-GI-loss already exists. However, the link between lower-GI-loss and hypokalemia does not and must be created and elaborated further. Similarly, at the next level, the link between lower-GI-loss and K-loss does not

exist. As this link is primitive the recursion terminates with creation of this link. Furthermore, as the attributes of K-loss and Lower-GI-loss are compatible and the two are causally consistent, this link can be established by simply adding its instantiation to the PSM. Having established this link the program aggregates this causal path to propagate the effects of the elaboration back to the higher levels of aggregation.

4.6 Projection

The projection operation is used to hypothesize and explain the associated findings and diseases suggested by the states in a PSM. The projection operation is very similar to elaboration. It differs from elaboration in that the causal relation being projected is hypothetical and therefore is not present in the PSM. Furthermore, the projection operation fails if the causal description of the hypothesized link is inconsistent with the description in the PSM at any level of detail. As a result, the application of the projection operation cannot result in the decomposition of a fully accounted node, creating an additional unaccounted component and therefore degrading the quality of explanation.

As stated above the projection operation is not an essential component of the structure building operations. However, it plays an important role in the diagnostic problem solver in exploring diagnostic possibilities, evaluating their validity and in generating expectations about the consequences of hypothesized diagnoses.

4.7 Component Summation and Decomposition

One of the important mechanisms in developing an understanding of the patient's illness is the evaluation of the effects of more than one disease present in the patient simultaneously, especially when one of the diseases alters the presentation of the others. To deal with such a situation competently, the program must have the ability to identify the effect of each cause individually, and the ability to combine these effects together. In this section we present the component summation and decomposition operations. Component summation combines attributes of the components to generate the attributes of the joint node; component decomposition identifies the unaccounted component by noting differences between the joint node and its existing components. These operations enrich the PSM by instantiating and unifying component nodes when the case demands them. This occurs whenever multiple causes contribute jointly to a single effect. An important case of this arises whenever feedback is modeled, because in any feedback loop there is at least one node acted on both by an outside factor and by the feedback loop itself. Finally, the decomposition of an effect with multiple causes into its causal components will also provide us with valuable information for evaluating the prognosis and formulating therapeutic interventions.

As the PSM is built, component summation and decomposition operations can cause a node in the program's general knowledge to be instantiated as a node and its several components in the PSM. If a node is primitive and there are multiple causes, the contribution of each cause is instantiated separately. Then the profile of the combination is computed using component summation. The combined effect is then instantiated and connected to its constituents by constituent links.

Because components are defined only for primitive nodes, the instantiation of composite nodes which involve component summation must be in terms of the summation of components in the node's elaboration structure. If the node is composite then we elaborate the constituent nodes around their focal nodes until we reach the primitive nodes associated with them. Then we combine these primitive nodes and aggregate their effects back. For example, if we know that a patient has hypobicarbonatemia and hypocapnia causing acidemia (figure 26), we can evaluate their combined effect as follows: (1) compute the component of acidemia caused by hypobicarbonatemia and hypocapnia individually, (2) focally elaborate these two components until each component can be described in terms of change in serum-pH (a primitive node), (3) sum the two components using component summation, and (4) aggregate the joint effect to derive the actual severity of acidemia.

As mentioned above, the mechanism of component summation allows us to represent feedback explicitly by representing the primary component of the change (the forward path) and the secondary feedback component (the response of the homeostatic mechanism in defense of the parameter being changed) as components to be summed to yield the whole. Figure 27 shows

Fig. 26. An example of component summation/decomposition

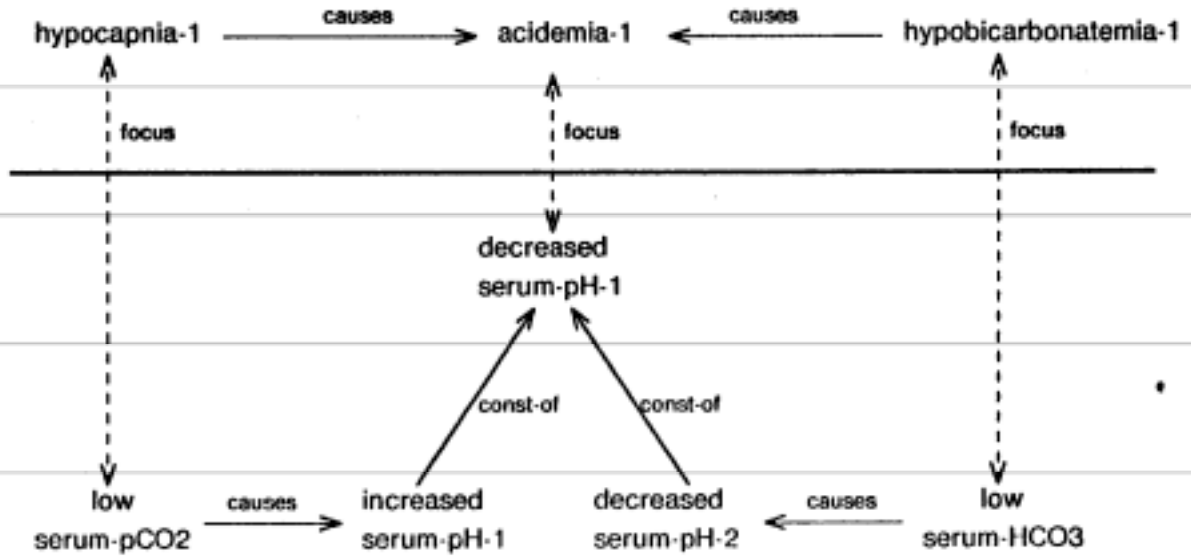
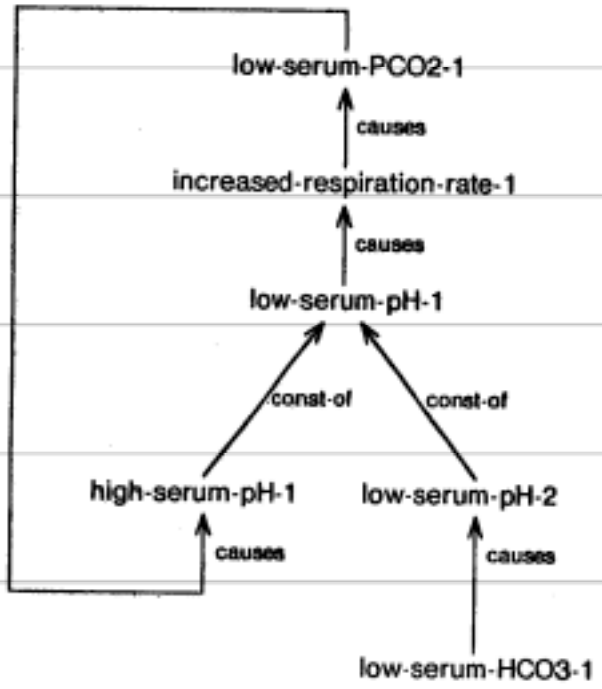


Fig. 27. Feedback loop represented using component summation



the primary change in serum pH caused by low serum bicarbonate and the response of the respiratory system to the change in serum pH. Read the example as follows: the lowering of the concentration of serum bicarbonate causes a reduction in serum pH, which causes hyperventilation and thus reduces the pCO_2, which in turn causes an increase in the serum-pH (negative feedback). This increase is less than the initial reduction, causing a net reduction in serum pH.

These operations deal not only with the magnitude of some disorder but also with other attributes such as duration. They are implemented by associating with each primitive node a multivariate relation that constrains the attributes of the node and its components. This mapping function is used by component summation in computing the attributes of the joint node from the attributes of the component nodes and by component decomposition in computing the attributes of the unaccounted component from the attributes of the joint node and its existing components. An example of the constraints is shown in the next example.

```
[(concentration*u electrolyte)
    [union:u
        [value*u #c (combine-electrolyte-value*c
                     (value*c (value*u component:1)),
                     (value*c (value*u component:2)),
                     (default*c component:1))]
        [start-time*u
            #c (min*c (value*c (start-time*u component:1)),
                      (vlaue*c (start-time*u component:2)))]
        [duration*u
            #c (max*c (value*c (duration*u component:1)),
                      (value*c (duration*u component:2)))]
        [belief*u
            #c (min*c (value*c (belief*u component:1)),
                      (value*c (belieif*u component:2)))]
    [component:1
         [value*u #c (component-electrolyte-value*c
                      (value*c (value*u union:u)),
                      (value*c (value*u component:2)),
                      (default*c union:u))]
         [start-time*u #c (value*c (start-time*u union:u))]
         [duration*u #c (value*c (duration*u union:u))]
         [belief*u #c (value*c (belief*u union:u))] ]]
```

The above example describes the multivariate relation between the components and their summation for the concentration of electrolytes. This relation is divided into two parts; the first part (associated with slot "union:u") describes procedures for combining the attributes of the two components ("component:1" and "component:2"). In particular, it states that the value of the joint-state (union) is determined from the values of the two components and the default value of the electrolyte concentration using a lisp function "combine-electrolyte-value". It further states that the belief in the joint-state is equal to the lesser of the beliefs in the components.[18] Similarly, the "start-time" of the joint-node is the earlier of the two start times and the duration of the joint-state is the longer of the two durations. A similar set of procedures for computing the difference (component:1) between the joint-state and a given component state (component:2) is described in the second part of the example shown above. This mapping relation can be used for computing the component summation/decomposition of electrolyte concentrations in any one of the different fluids in the body such as extra-cellular fluid, intracellular fluid, and urine.

The component operations are activated when a node is added to the PSM where another node in the same class is already present. These operations incorporate the new node into the structure of the PSM and delete any structure in the PSM that is no longer valid due the the addition of the new node. These operations can be divided broadly into three cases based on the

---

18. This is consistent with our view that "the belief in an explanation is equal to the belief of its weakest link". This belief computation is similar to that used in Glaucoma/CASNET program [Weiss78] and in fuzzy set theory [Zadeh65, Gaines76].

properties of the node already present and the new node being added: (1) both the new and pre-existing nodes are both unsupported by observation; (2) the new node being added is supported by observation and the pre-existing node is not; and (3) the new node is not supported by observation and the pre-existing node is. A node is said to be supported by observation if the node is either an observed node or is a causal predecessor of an observed node which is fully accounted for. The details of the three cases:

> Case 1: Neither the new nor the pre-existing node is supported by observation. In this case the joint effect of the two nodes is computed and the two nodes are connected to the joint effect using component links. If the pre-existing node already has component structure, the new node is directly connected to the pre-existing joint effect and the attributes of the joint effect are revised to be consistent with this addition. Any of the successors of the two nodes which are consistent with the joint effect are rerouted through the joint effect and those which are not consistent are deleted and the effects of these deletions are propagated.
>
> Case 2: The new node being added is supported by observation and the pre-existing node is not. In this case the joint effect of the resulting structure (upon application of the component operation) must be same as the new node. If the pre-existing node and the new node are consistent with one another then the pre-existing node is replaced with the new node and the operation is complete. If they are not, the difference between the observed and the unobserved is computed, and a node corresponding to the difference (called unaccounted-component) is instantiated. Next the pre-existing (accounted-component) node and the unaccounted component to the new (joint-effect) node are connected using component links. Any of the successors of the pre-existing node that are consistent with the joint-effect are rerouted through it and those that are not consistent are deleted and the effects of these deletions are propagated.
>
> Case 3: The pre-existing node is supported by observation while the new node being added is not. As in the case 2, the observed node is the designated joint effect. This case is somewhat more complex, because the pre-existing node is observed and may have constituents of any possible form, i.e., may be fully accounted for, partly accounted for, or fully unaccounted for. In each case the new node is added to the pre-existing structure as a constituent as shown in the figure 28.

Figure 28 shows subcases of case 3 where the pre-existing node (bold square) is supported by observation while the new node (crossed square) being added is not. The left side of the figure shows the situation before the component summation and the right side shows possible situations after the component summation. Figure 28(a) shows the operation for a fully unaccounted

Fig. 28. Component summation/decomposition: Case 3

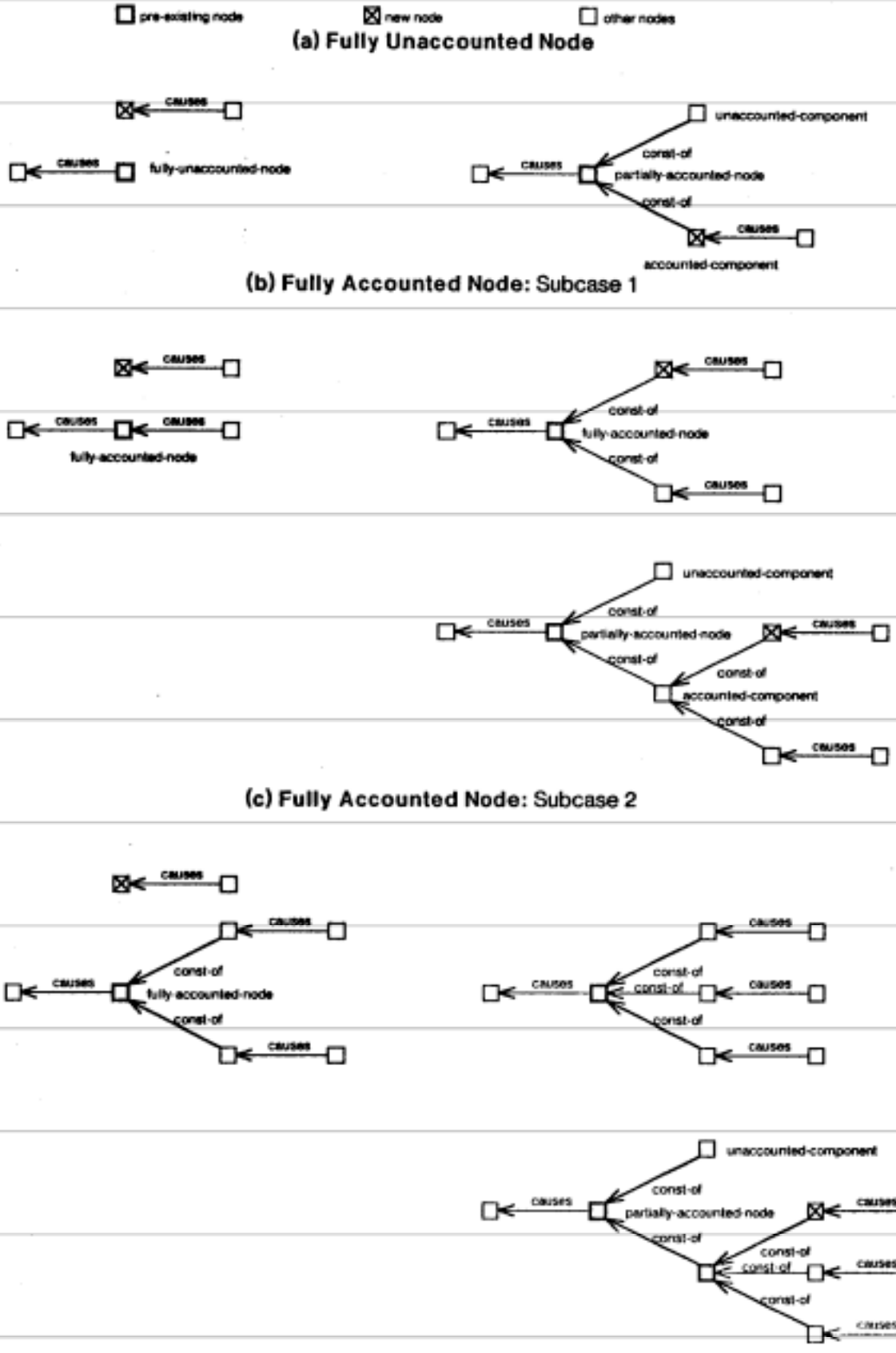
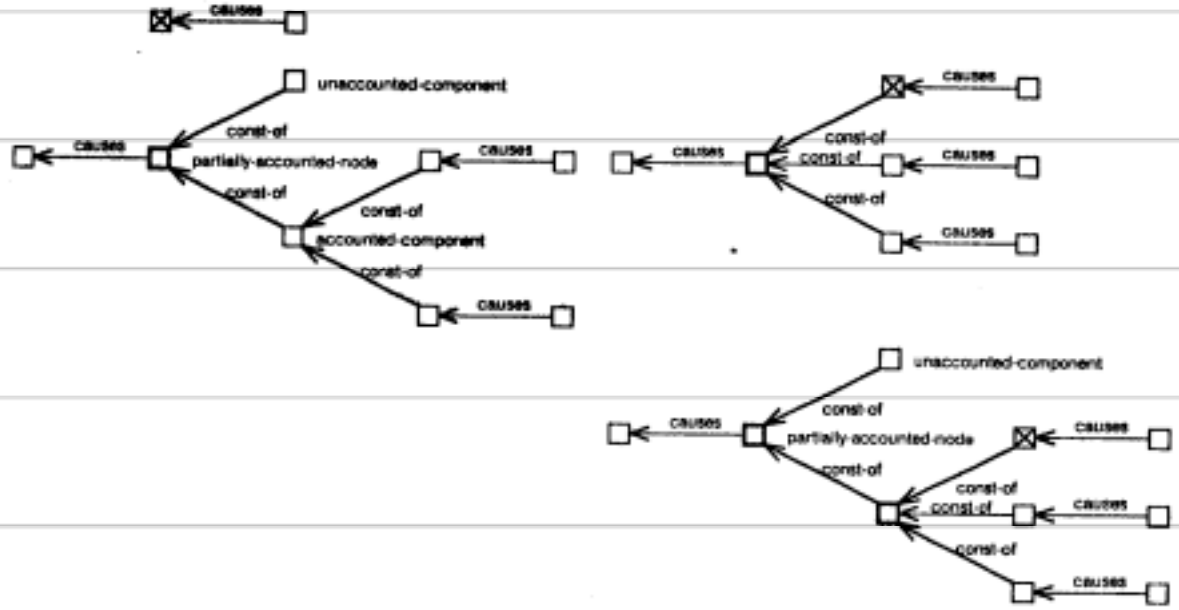
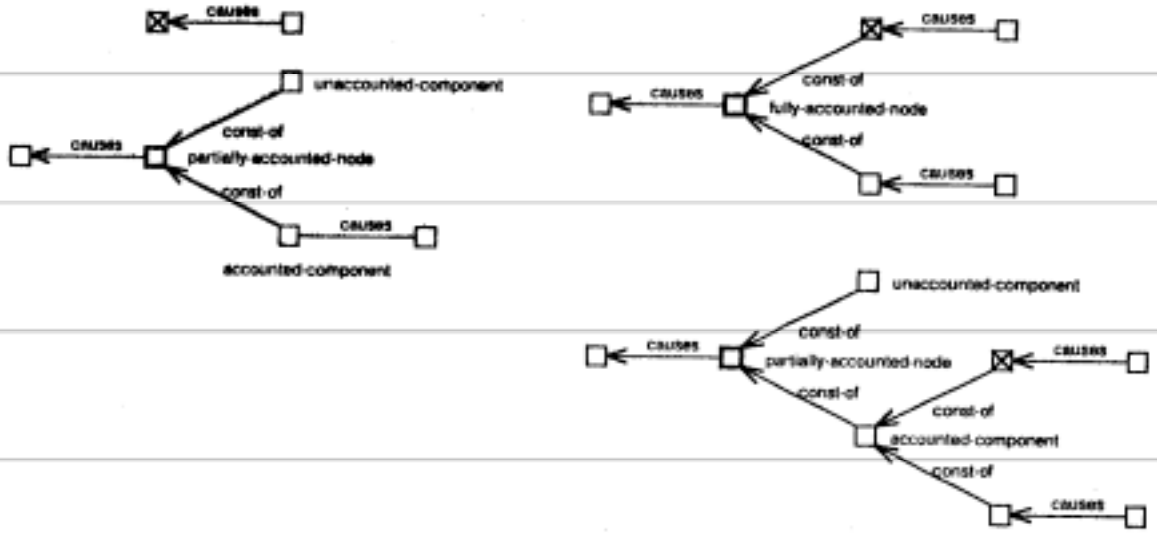


Fig. 28. (continued)

(d) Partially Accounted Node: Subcase 1



(e) Partially Accounted Node: Subcase 2



pre-existing node. Figure 28(b) shows the operation for a fully accounted pre-existing node with one cause. The first structure on the right shows the situation when the effects of existing cause and the the new node are still consistent with the pre-existing node. In this situation the components of each of the two causes are instantiated and connected as shown in the figure. The second structure shows the situation when the sum of the new node and the effect of the existing cause is not consistent with the pre-existing node. In this situation the pre-existing node is decomposed into an accounted and an unaccounted component. The accounted component is dealt with similar to the first structure and the unaccounted component is marked as being unaccounted. Figure 28(c) shows the operation for a fully accounted pre-existing node with multiple causes. This case is handled similar to that in figure 28(b). Figure 28(d) and (e) show the operation for partially accounted pre-existing node. If the new node matches the unaccounted component of the pre-existing structure, the resulting structure is fully accounted for, if it does not the accounted and unaccounted components of the pre-existing node are recomputed and the new node is connected to the accounted components.

In this section we have developed a knowledge representation formalism and operations for dealing with effects with multiple causes and feedback loops common in the physiological regulation of the body's vital functions. The mechanism developed here is intended for symbolic description for reasoning with and explaining the abnormalities in physiological regulation in a patient, not for predicting the behavior of physiological parameters over time using dynamic simulation techniques.

5. Diagnostic Problem Formulation and Information Gathering

The patient specific model (PSM) developed in chapter 4 was designed to provide the program with the capability of expressing its understanding about the patient's illness. However, due to the lack of complete knowledge about the patient and due to uncertainties in the medical knowledge, this understanding may be imprecise and incomplete. Our task is to identify these weaknesses and gather information that will help reduce or eliminate them. Viewed differently, these weaknesses identify a set of problems, all of which need to be solved in the process of diagnosis. The availability of a set of problems to work on simultaneously provides the problem solver with an opportunity to be efficient by abstracting common aspects of problems and by selecting an efficient order in which the problems are to be solved. This chapter examines several issues: (1) the process of identifying these weaknesses and formulating a diagnostic problem based on them, (2) the representation of this diagnostic problem and its decomposition into simpler problems, and (3) the evaluation of newly acquired information for apparent and real discrepancies.

The general medical knowledge in the program contains disease prototypes. However, given the facts about a patient along with a possible explanation, this prototypical information can be substantially constrained. For example, knowing that the patient has moderately severe metabolic acidosis, we can constrain the diseases hypothesized to account for the metabolic acidosis to be consistent with it, e.g., if salmonellosis is a hypothesized cause of this metabolic acidosis, it must be moderately severe and must have a duration of greater than two days. Secondly, only a small portion of the medical knowledge is relevant to any given diagnostic situation. For example, knowing that the patient's anion gap is normal, all the causes of metabolic acidosis that are not consistent with normal anion gap can be ruled out as being irrelevant to the diagnosis.[19] We therefore introduce the notion of a *diagnostic closure* (called DC) which contains the medical knowledge local to the diagnostic situation, extracted from the medical data-base and made specific to the PSM. The DC is constructed by hypothetically projecting forward the states of a PSM to identify the consequences predicted by the states of the PSM and by projecting backwards the unaccounted for states of the PSM to identify diseases that can account for these states. Note that within each PSM all the findings and diseases complement each other in forming a single coherent explanation, while different PSMs provides alternate explanations which are mutually exclusive. Further, each DC contains alternatives within the context of the PSM associated with it. Thus, the diagnostic alternatives themselves are divided into groups, each group being consistent with a partially complete explanation of the patient's illness, and the different groups represent alternatives consistent with markedly different possible explanations.

---

19. A similar distinction is also made in PIP and Internist, where any disease which is not currently active can be considered to be irrelevant to the current diagnostic activity.

We have argued that the ability to identify discrepancies in incoming information plays a crucial role in the diagnostic process. For example, in studying the problem solving behavior of clinicians, Kassirer and Gorry note:

> *"The physician appeared to use ... his concept of a disease (hypothesis), a state, or a complication as a model with which to evaluate new data from the patient. Such a model provides a basis for expectation; it identifies the relevant clinical features that should prove fruitful for further investigation."*
>
> — Clinical Problem Solving [Kassirer78, page 250]

The ability to evaluate the implications of the incoming information is an important part of clinical practice, where the accuracy or the completeness of information cannot be taken for granted. We may be presented with a questionable finding which, if accepted, may require reformulation of the currently held diagnosis with far-reaching implications. However, it may be unwise to act on any such information unless it can be substantially corroborated, and its validity as a diagnostic sign checked out. For example, upon unexpectedly finding "substantial weight increase" in a patient it is wise to check if the two weights were taken on the same scale before jumping to the conclusion that the patient is "retaining water". Inability to do so poses a serious problem for programs such as PIP. The problem arises because accepting such a finding may strongly favor hypotheses which erroneously predict the finding and against those hypotheses which correctly do not predict it, possibly causing the correct hypotheses to be dropped from further consideration. Thus, the program may not be able to come back and ask a simple question that could save it from taking a "garden path".

The diagnostic closure discussed above provides the program with an ability to evaluate the consistency of a finding before it decides to accept it. For example, as new information is gathered, if the profile of the new information is consistent with that present in a DC, we know that this information is consistent with the PSM and lends positive support to the diagnosis under consideration. By the same token, if some information is not consistent with a DC under consideration, we know that this information can not be assimilated into the PSM without some modification. Finally, if the incoming information is not consistent with any of the DCs then we know that our entire line of reasoning is under question, and if the information is true, a major re-analysis of the program's understanding will have to be undertaken. Because such a situation can be identified, the program has an opportunity to suspend the global diagnostic processing and revert to local processing to validate the finding or to justify ignoring it.

The problem solvers in PIP and INTERNIST-I alternate between gathering a fact (based on their hypothesis lists) and re-evaluating the hypothesis lists (based on the new fact). Each fact is treated as an independent inquiry; the program does not group facts in a clinically meaningful and focused pursuit of diagnosis. This causes the information acquisition to become erratic and

vulnerable to incomplete specification of information.[20] Furthermore, the lack of commitment in pursuing any given information gathering strategy (e.g., discriminate, confirm) to completion diminishes their effectiveness. This problem can be solved by allowing the diagnostic problem solver to plan a group of questions focused around a single diagnostic task. The diagnostic closure already provides the dependencies necessary for such diagnostic planning. Diagnostic planning generally begins with the global task of discriminating between the alternate explanations provided by the set of PSMs. This task is successively decomposed into smaller tasks using diagnostic strategies of confirm, differentiate, rule-out, group-and-differentiate and explore. This results in a set of questions which, if answered, would help the program in solving the problem at hand.

It is common among physicians to *"think out loud"* while discussing a medical case with their colleague. For example, in analyzing protocols of medical diagnosis, Sussman notes:

> *"Thus, we have heard doctors react to new facts with such phrases as: "I expected that.", "(It is) consistent with my assumptions.", "I did not expect that ...", "This new fact is making me very unhappy with my diagnosis.". Among the most important reactions are ones of the form: "this does not really fit in. Perhaps he has...."."*
>
> — Some Aspects of Medical Diagnosis [Sussman73]

This *thinking out loud* plays an important role in communication between physicians. We require the program to have not only a similar ability evaluate the incoming information in comparison with its expectation, but also the ability to *think out loud*, which is essential in allowing the user physician to get a feel for the program's reasoning and understanding. The diagnostic closure allows the program to explain the spectrum of diagnostic alternatives consistent with a PSM, and the planned goal oriented diagnostic questioning allows the program to justify the motivation of the diagnostic reasoner in asking the questions, its expectations about the information being sought, and how this information relates to the hypotheses under consideration.

5.1 Global Diagnostic Cycle

The diagnostic algorithm for the ABEL system is:

---

20. On the other hand, de novo generation of the hypothesis list prevents the program from taking "garden paths".

(1) **Presenting Complaints:** The serum analysis and the initial complaints are analyzed. A small set of initial PSMs is created and added to the list of causal hypotheses (the CH-list).

(2) **Rank Ordering Hypotheses:** All PSMs in the CH-list are scored for the quality of explanation they provide for the patient's illness. The leading one or two of these PSMs are selected as possible explanations.

(3) **Computing Diagnostic Closure:** Diagnostic closures for the selected PSMs are computed and disease hypotheses in each DC are scored.

(4) **Termination:** if the diagnostic closures for all PSMs are null or if some PSM provides a complete and coherent account for the patient's illness then the current phase of diagnosis is complete.

(5) **Diagnostic Information Gathering:** Based on the number of DCs (i.e., the PSMs selected in step 2), a top level confirm or differentiate goal is formulated. Using diagnostic strategies, this goal is successively decomposed into simpler subproblems until individual questions are formulated.

(6) **Re-structuring the PSM:** If step 5 results in any new finding being known, then that finding is incorporated into the each of the PSMs by extending the structure of the PSMs to take the observed finding into account. Finally, this process is repeated starting at step 2.

In the remaining sections of this chapter we will study the individual steps of this algorithm.

5.2 Diagnostic Closure of a Hypothesis

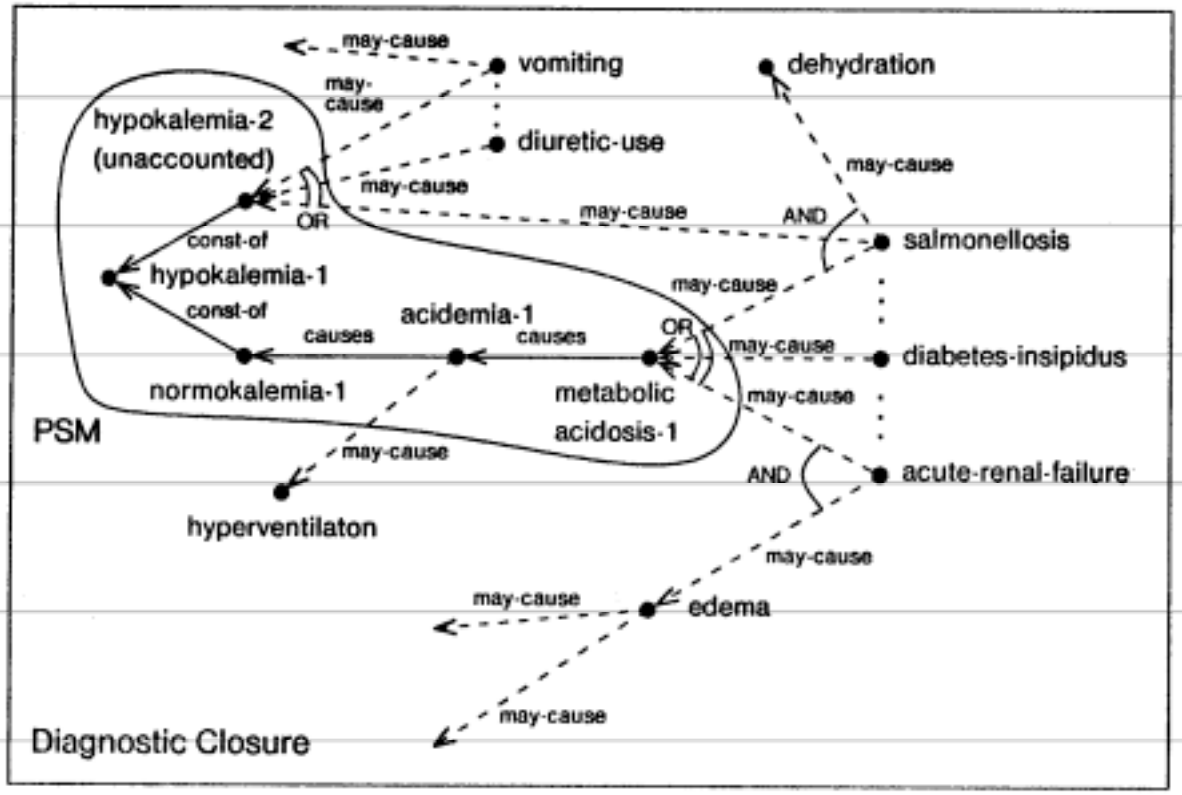
A diagnostic closure (DC) describes that part of the medical knowledge that is directly relevant to the diagnostic exploration of a PSM. It contains, in addition to the PSM, causal pathways from the unaccounted findings in the PSM to some of the possible diseases (ultimate etiologies) that can account for them, and causal pathways from some of the states in the PSM and the hypothesized diseases to (predicted) observable findings. Stated differently, a DC contains alternative extensions needed to adequately complete the explanation provided by the PSM. The DC associated with a PSM is initially created by hypothetically projecting the states of the PSM. During the process of diagnostic planning, new DCs may be created by copying parts of an existing DC,[21] and by further projecting the diseases or findings under consideration. Furthermore, when some new information is received during the execution of a diagnostic plan,

21. For example, in order to differentiate between alternative hypotheses contained in a DC the program may create a set of disjunctive DCs, one for each alternative.

the alternatives which are not consistent with the finding may be pruned from a DC. Figure 29 shows an example of a DC for a PSM with unaccounted metabolic acidosis and partially accounted hypokalemia. Note that metabolic acidosis and hypokalemia both can be accounted for by a single disease hypothesis: salmonellosis. However, if we assume that the unaccounted component of hypokalemia is caused by vomiting, we must find some other cause for the metabolic acidosis, e.g., acute renal failure or diabetes insipidus.

The diagnostic closure of a PSM provides us with the attributes of the hypothesized diseases and findings that are consistent with the PSM. It describes the program's diagnostic expectations against which the incoming information can be evaluated. Furthermore, by tracing the causal pathway from the hypothesized finding to the states in the PSM, we can determine how this finding relates to the PSM, and what intermediate assumptions are needed to assimilate this finding into the PSM. On the other hand, if the new finding is not consistent with any of the DCs under consideration then we know that this information is inconsistent with the program's current understanding. To accommodate a contradiction with the currently held hypothesis requires some major revision in the structure of the PSM. This process is computationally expensive and,

Fig. 29. An example of diagnostic closure



if possible, should be avoided. As described above, ABEL has the ability to identify situations requiring a major revision, and to ask further questions to validate or invalidate the contradictory finding. However, when a contradictory finding is validated, ABEL abandons its current line of diagnostic inquiry and revises its PSMs. Clinical studies have shown that a physician when faced with a similar situation also attempts to avoid revising his diagnostic hypotheses. He attempts to disprove the offending piece of information or reconcile it by finding a sufficient excuse for ignoring it. On occasions, even after the validity of the contradictory finding is established, a physician may choose to ignore the finding until the current line of diagnostic questioning is completed. ABEL however, abandons its current line of diagnostic inquiry and revises its PSMs if a contradictory finding is validated. It does not have the ability to postpone consideration of any contradictory finding.

5.3 Scoring the PSM

The score of a PSM measures the degree of incompleteness of the PSM as an explanation of the patient's illness. It is computed by summing the severities of partially and fully unaccounted states in the PSM. The scoring algorithm could be further improved by taking into consideration the need of a finding to be accounted for by an acceptable diagnosis. Furthermore, the program currently does not take into account the degree of explainability of a PSM. For example, a PSM may have a large number of unaccounted findings that can be accounted for by a single etiology, while another PSM may have only a few unaccounted findings but may require the invocation of multiple etiologies to account for them. Clearly, diagnoses with multiple etiologies are less desirable and much less frequent than diagnosis with a single etiology. The degree of explainability of a PSM is an important measure and should eventually be taken into account while scoring a PSM.[22] Although the current method for computation of the score is primitive and should be extended using the additional factors discussed above, it appears to provide an acceptable level of discrimination between PSMs.

5.4 Scoring a Disease Hypothesis

Diseases are hypothesized to explain findings left unaccounted for by the PSM: a new disease is hypothesized only when it is capable of explaining some of the unaccounted findings. In this section we will consider a mechanism for scoring these hypotheses.

When a disease is hypothesized it may predict some consequences which may not fit well with the PSM, giving rise to new unexplained states. These additional unexplained consequences reduce the desirability of the hypothesis being considered. Furthermore, the hypothesized

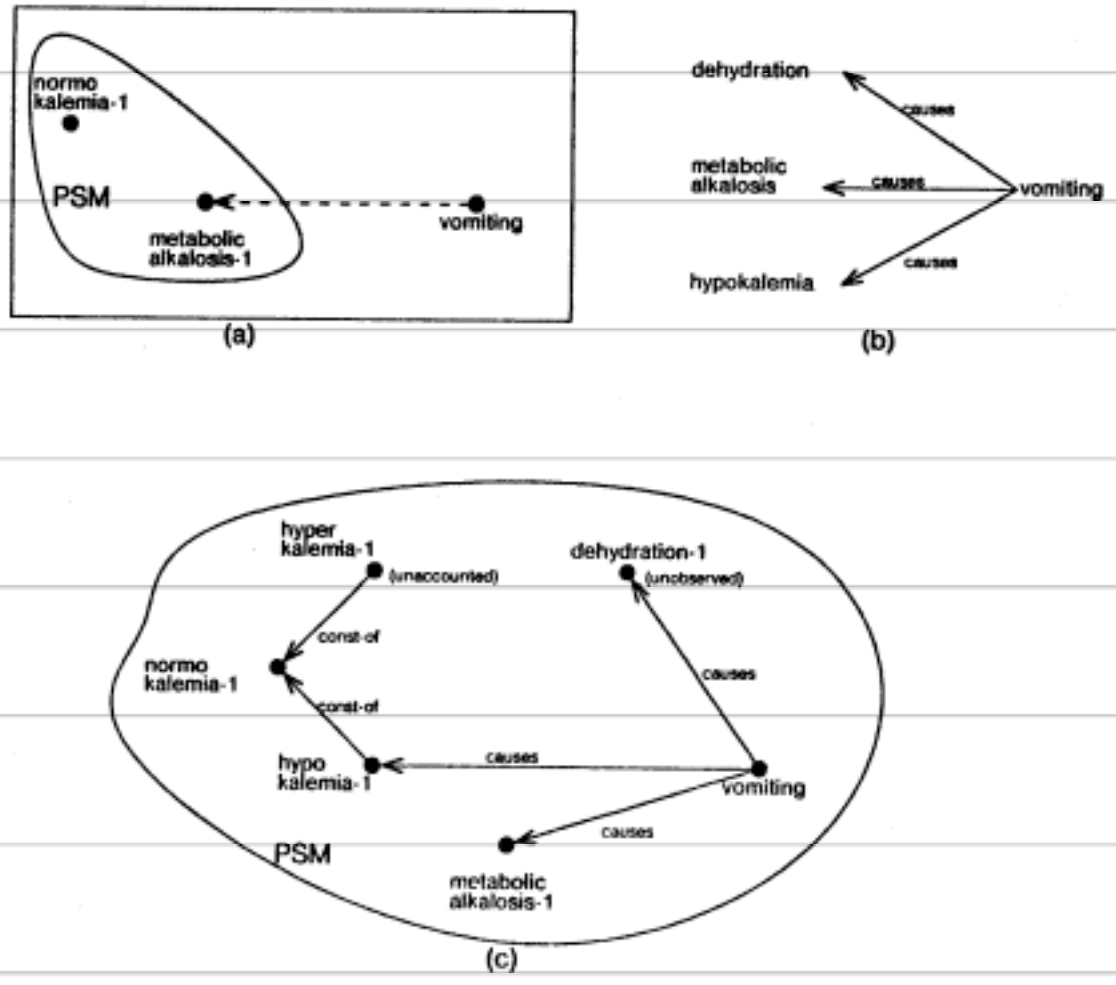
---

22. It can be done if we compute the smallest number of etiologies that cover all unaccounted findings (using the DC) in each PSM before scoring them. This measure, however, has not been implemented as it is computationally prohibitive in the current implementation of the program.

disease may predict some consequences which are as yet unobserved. These unobserved findings identify the additional information that can be used to confirm the disease hypothesis. For example, figure 30(a) shows a PSM with metabolic alkalosis and normokalemia, and vomiting hypothesized to account for the metabolic alkalosis. Figure 30(b) shows the findings predicted by the hypothesized vomiting. Figure 30(c) shows the consequences of adding the hypothesized vomiting to the PSM. The vomiting hypothesized in figure 30(a) explains an unaccounted for node, metabolic alkalosis, gives rise to a new unexplained node, hyperkalemia, and predicts an as yet unobserved finding, dehydration.

The usefulness of a disease hypothesis depends (ultimately) on its potential of being confirmed. This usefulness can be estimated using the explained, unexplained and unobserved findings associated with the hypothesis. Note, however, that the disease scores are computed for the purpose of ordering the diagnostic search, i.e., they provide a heuristic for performing a best-first search. The score of a disease hypothesis does not reflect the belief in the likelihood of the

Fig. 30. An example of explained, unexplained and unaccounted findings



given disease being the correct diagnosis, but an estimate of its heuristic search utility. That is, given the available information, pursuing that disease hypothesis will lead efficiently to the final diagnosis. Although the two measures are similar and have often been confused with one another, they can be substantially different as more and more sophisticated search and error recovery techniques are used. In most of the previous programs this distinction was not made; thus even if a particular disease was a useful hypothesis, it could not be considered if most of its findings were as yet unknown. Further, it prevented these programs from accepting a working hypothesis which, *even while having a low probability of being right*, could lead efficiently to the right "ball-park", which when reached would allow them to resort to more specific criteria to explore the restricted space.

In ABEL the disease hypotheses are ordered in two steps. First, they are grouped according to the number of unaccounted findings that can be accounted for by each hypothesis. Second, among those hypotheses that can account for the same number of findings, the diseases are rank-ordered by a score computed from the three factors discussed above. They are: (1) match, the number of causes and findings in the PSM that are consistent with the disease hypotheses;[23] (2) mismatch, the number of causes and findings in the PSM that are inconsistent with the disease hypotheses; and finally (3) unknown, the number of unobserved findings predicted by the hypothesis which are not inconsistent with the PSM. A disease hypothesis is eliminated from immediate consideration (for one cycle of diagnostic inquiry) if the difference of match and mismatch is below an arbitrary threshold. The match combined with the unknown corresponds to the maximum possible score attainable by a given disease hypothesis. If this score goes below a threshold, the hypothesis can not be confirmed even if all the remaining unknown findings are resolved in favor of the hypothesis.

The above criterion for scoring the disease hypotheses is purely structural. It does not take probabilities of occurrences of different diseases into account. Incorporation of probabilities as a secondary scoring criterion should substantially improve the quality of the scoring mechanism. However, we believe that the criterion for evaluation of the heuristic value of the disease hypothesis as well as belief in a diagnosis should be primarily structural. Probabilistic scoring can be used effectively in differentiating between structurally similar hypotheses. However, primary reliance on probabilistic scoring without structural considerations (such as adequacy, coherence, match and mismatch), as has been the case with the first generation programs, is inadequate. Some of these inadequacies have been discussed in chapter 1.

---

23. Note that a finding that is fully accounted for in the PSM can still be consistent with the new hypothesis if the addition of the hypothesis does not cause the finding to be over-compensated, resulting in an unaccounted component.

5.5 Information Gathering Strategy

The process of diagnosis can be viewed as the process of discriminating between diagnostic alternatives. A strategy commonly used to achieve this called the *differentiation strategy*. Using the results of protocol analysis, researchers [Pople75a, Miller75, Elstein78, Kassirer79] have identified a larger class of diagnostic strategies which in addition to differentiate include confirm, rule-out, explore, refine etc. Although these additional strategies can be considered to be special cases of the differentiation strategy, in special situations they can provide substantial improvement in processing over differentiation.

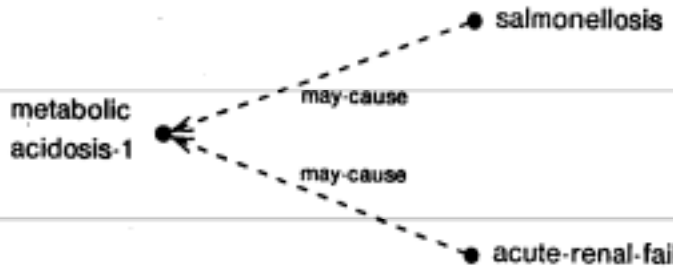
The selection of an appropriate strategy is based primarily upon the syntactic structure of the diagnostic problem.[24] One measure commonly used is the number of alternate hypotheses under consideration and their relative strength. The confirmation strategy is used when only one hypothesis is under consideration, or when one hypothesis is much more likely than all others. The rule-out strategy is the inverse of the confirm strategy; it is used to eliminate some hypothesis which is substantially less likely than all the others. Its major utility is in allowing final confirmation of some hypothesis, such as essential hypertension, by eliminating all other less likely alternatives or cutting a large group down to where differentiate strategy can be used. The differentiation strategy is used to discriminate between two (or three) hypotheses with similar belief factors. The above strategies are all used in the Internist-I program.

The remaining strategies, such as group-and-differentiate and refine, reformulate the diagnostic problem. The group-and-differentiate strategy is used when we have a large number of alternate hypotheses with similar belief factors. Here we need to discard a large number of hypotheses rapidly in order to focus our attention on a small number of alternatives. This can be achieved by partitioning the alternatives into a small number of groups according to some common characterization (e.g., common organ system involvement, etiology, temporal characteristic or pathophysiology) and then applying a differentiation strategy to rule in or rule out one of the groups, thus narrowing the hypothesis set substantially. The refinement strategy is used to refine a hypothesis about a general class of diseases into more specific hypotheses. Refinement results in a disjunctive set of hypotheses. Hence, refinement and, as we have seen group-and-differentiate are commonly followed by differentiation. Finally, the explore strategy is used when the patient description does not provide any well-defined diagnostic problem to solve. In such a situation we explore the findings systematically (e.g. review of systems) to uncover sufficient evidence to formulate a specific diagnostic problem.[25]

---

24. In certain situations though, the general syntactic mechanism may be overruled by more important considerations. For example, if one of the alternatives has life-threatening consequences, we may first want to get a definitive ruling on it rather than differentiating among all the possibilities.
25. An example of this is the *review of systems*, a detailed exploration of every part of the body in search of abnormalities.

Fig. 31. Initial diagnostic closure for salmonellosis and acute renal failure



The use of these strategies in the first generation programs has been limited to a single application to identify the most useful finding. In this document we advocate viewing these strategies as decomposition operators that reformulate the diagnostic problem into a group of simpler problems. With this formulation we can repeatedly apply the diagnostic strategies to the top level diagnostic problem, successively decomposing it, until we reach subproblems that can be solved directly by asking single questions.

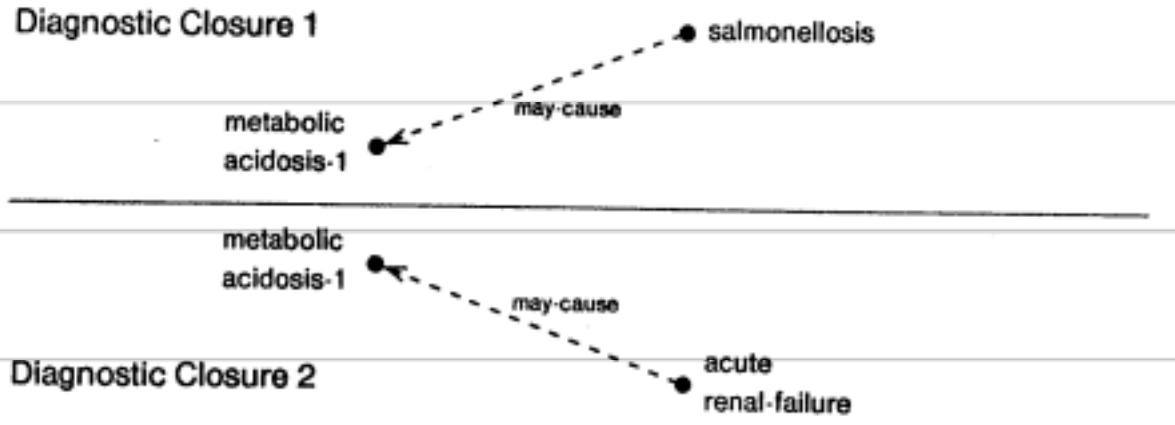
Consider the following simple example. Assume that we have a patient with moderately severe metabolic acidosis and are considering two possible causes of this metabolic acidosis, namely salmonellosis and acute renal failure.[26] The diagnostic closure consistent with this situation is shown in figure 31. We pursue this diagnostic closure by setting up a diagnostic problem as shown below.

```
Goal 1: differentiate Salmonellosis acute-renal-failure
              salmonellosis
                  belief: likely
                  severity: moderate
                  duration: greater-than 2 days
              acute-renal-failure
                  belief: possible
                  severity: moderate
                  duration: greater-than 1 week
```

To differentiate between salmonellosis and acute renal failure the program sets up a diagnostic closure for each of the possibilities (shown in figure 32). The first DC is constructed with the assumption that salmonellosis is the true cause of the observed metabolic acidosis, and the second with the assumption that acute renal failure is the true cause. The program then explores the consequences of its assumption in each case by projecting the disease hypotheses forward

26. We are using an unrealistically simple example for the purpose of illustration. For this example we have assumed that the patient has received fair quantity of IV fluid. Furthermore, we assume that the electrolyte concentrations in urine are not available; the differentiation is trivial if the urine electrolytes are available.

Fig. 32. Diagnostic closure separated for each possibility



(shown in figure 33) and compares the two projections. From the projections it observes that salmonellosis and the acute renal failure predict different states of hydration for the patient. Based on this observation it formulates the next diagnostic problem shown below.

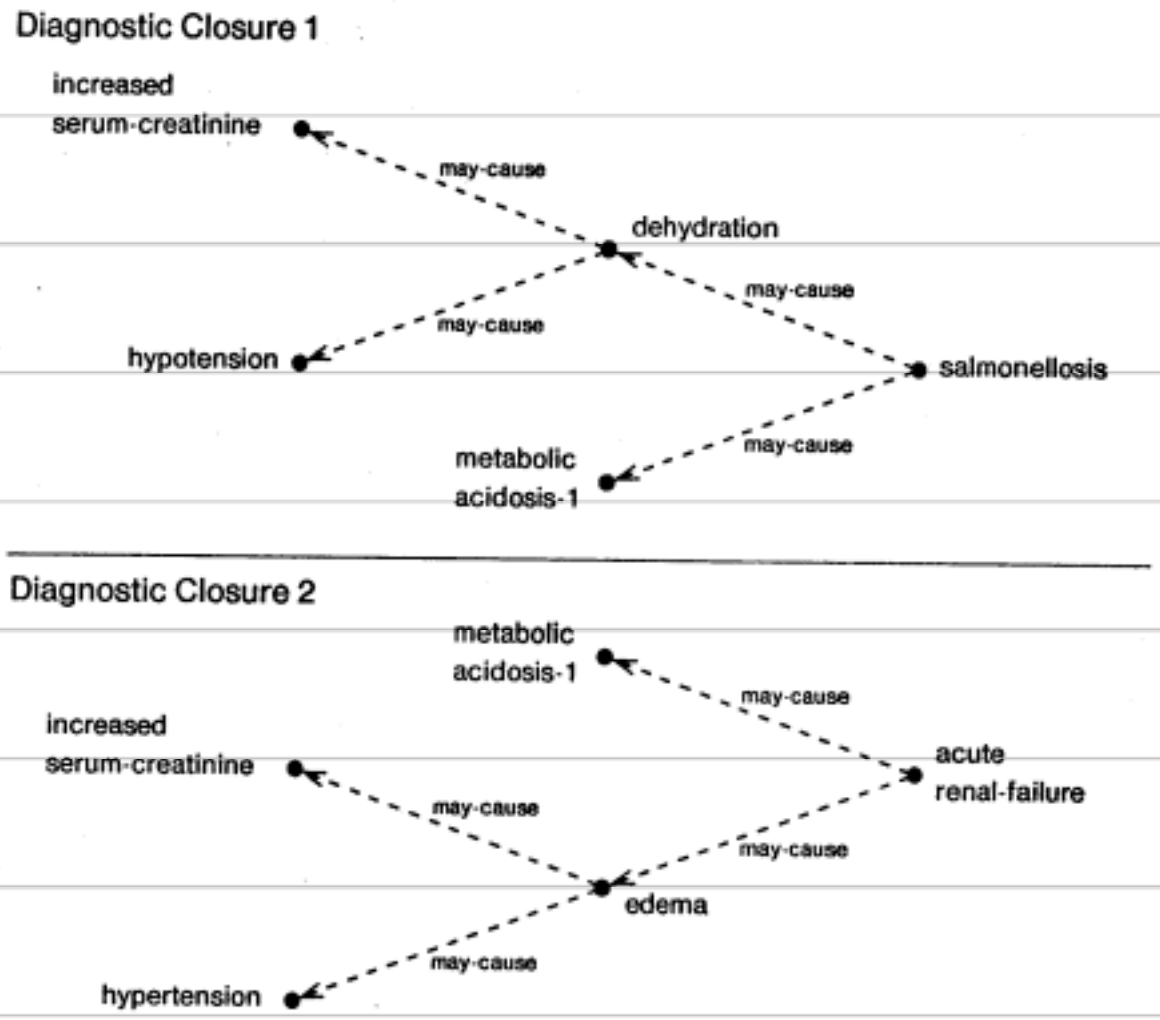
```
Goal 2: differentiate dehydration edema
              dehydration
                  caused-by: salmonellosis
                  belief: likely
                  severity: moderate
              edema
                  caused-by: acute-renal-failure
                  belief: possible
                  severity: moderate
```

Let us assume that the state of hydration cannot be directly ascertained by inquiry and the program decides to decompose this goal into two subgoals, one each for confirming dehydration and edema.

```
Goal 3: confirm dehydration
              dehydration
                  caused-by: salmonellosis
                  belief: likely
                  severity: moderate

Goal 4: confirm edema
              edema
                  caused-by: acute-renal-failure
                  belief: possible
                  severity: moderate
```

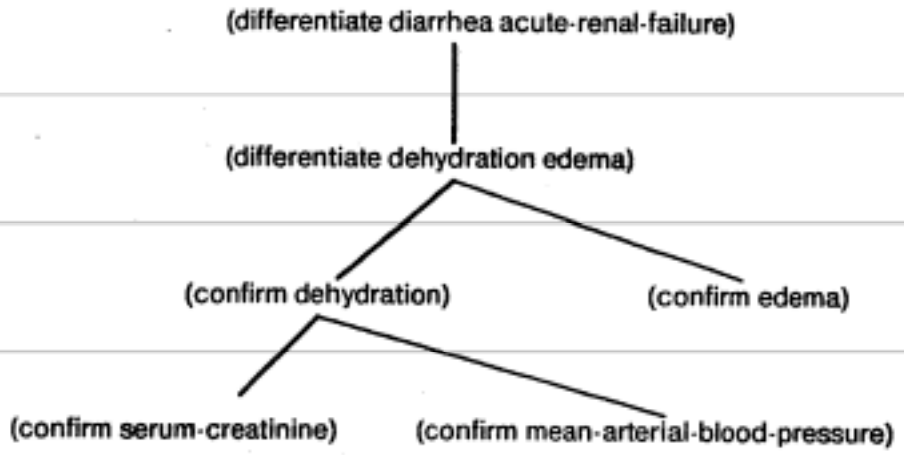
As dehydration is the more likely of the two (resulting from our initial assumption that salmonellosis is more likely than acute renal failure), the program chooses to pursue dehydration first. Since we have assumed that the state of hydration is unknown, the program must attempt to

Fig. 33. Diagnostic closures for each possibility projected forward



confirm it by gathering information like increased serum creatinine, hypotension, and poor tissue turgor. However, while formulating the goal for confirming serum creatinine, the program notices (using the second DC, figure 33) that the increased serum creatinine is also predicted by acute renal failure. The program incorporates this information in its goal structure. The subgoals formulated by the program in this situation are shown next.

Fig. 34. The goal tree




```
Goal 5: confirm serum-creatinine
              serum-creatinine
                  caused-by: dehydration
                  belief: likely
                  value: between 2 and 4
              serum-creatinine
                  caused-by: acute-renal-failure
                  belief: possible
                  value: between 3 and 7

Goal 6: confirm mean-arterial-blood-pressure
              mean-arterial-blood-pressure
                  caused-by: dehydration
                  belief: likely
                  value: low
              mean-arterial-blood-pressure
                  caused-by: acute-renal-failure
                  belief: possible
                  value: high
```

The goal structure of the program when inquiring about the serum creatinine is shown in figure 34 (the bold lines indicate the flow of control). The goal structure encodes the program's rationale for asking the question: it explicitly encodes the program's reason for asking the question and the context in which the question is being asked. Therefore, if the user chooses to ask for an explanation at this point it is possible for the program to provide the following types of explanations. (The explanation provided here is a paraphrasing, in better English, of the program's actual explanation, which is produced by a very simple English generator [Swartout80].)

Explain: I am expecting the patient to have mild elevation in serum creatinine. Increase in serum creatinine may be caused by dehydration, which may be caused by salmonellosis. The salmonellosis may account for the observed metabolic acidosis. It is also the leading cause of metabolic acidosis under consideration. Increase in serum creatinine may also be caused by acute renal failure, which may cause metabolic acidosis.

Justify: I am exploring the cause of metabolic acidosis. I am differentiating between the two leading causes of metabolic acidosis, namely salmonellosis and acute renal failure. I am differentiating between dehydration and edema. The dehydration may be caused by salmonellosis and the edema by acute renal failure. I am pursuing dehydration. I am pursuing serum creatinine. Increase in serum creatinine may be caused by dehydration. Increase in serum creatinine may be caused by acute renal failure.

Viewing the individual diagnostic strategies as problem decomposition operators allows the program to set up the diagnostic goal structure described above. This goal structure not only allows the program to explain and justify its diagnostic behavior, but also provides a framework for evaluating the user response locally in the context of the expectations. It allows the program to react locally when a discrepancy is detected or when further exploration of the finding is needed, gracefully integrating the program's global disease-centered processing with the local symptom-centered processing.[27]

Each top level diagnostic inquiry, described above, is followed by incorporation of all the information gathered into the existing PSMs (using the structure building operators described in chapter 4), and the formulation of a new diagnostic problem. This process is repeated until an adequate diagnosis of the patient's illness is achieved or until all the information relevant to the diagnosis is exhausted.

Summarizing, in this chapter we have introduced the notion of a diagnostic closure, which contains the hypothesized diseases, findings and causal relations relevant to the diagnostic task at hand. A diagnostic closure is created by projecting appropriate states in the PSM or hypothesized diseases forward to identify their predicted consequences and backwards to identify their possible causes. Once we have this knowledge for each diagnostic possibility, we have the dependencies necessary to do diagnostic planning.

---

27. We have just begun to exploit all the capabilities afforded by this mechanism. Although the current program does not make sophisticated use of these capabilities, we believe that extending the program to do so is possible given the current understanding of the process.

Diagnostic problems are generated by identifying the places were two or more hypotheses differ from one another in the interpretation of the findings. The set of problems identified is used in formulating a top level diagnostic goal for one cycle of diagnostic problem solving. The problem solver then generates a tree structured plan by successively decomposing this goal using strategies such as differentiate, confirm, group-and-differentiate, and rule-out. The diagnostic plan, in conjunction with the diagnostic closure, provides the context in which a question is asked, the program's reason for asking the question and its expectations about the possible responses to the question. This knowledge is used to guide the diagnostic inquiry as well as to provide explanation for the program's behavior.

Each cycle of diagnostic problem solving is viewed as an integral operation. During this cycle, the problem solver focuses on one top level diagnostic problem and attempts to solve it. This provides a focus for the interaction between the user physician and the program.

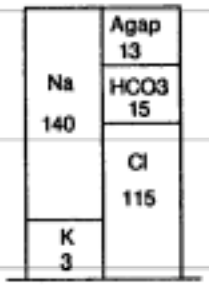
Finally, the information gathering process of each diagnostic cycle is followed by the revision of the structure of each PSM, making it consistent with the newly available information. Thus, at the end of each cycle of diagnostic inquiry, the PSMs are internally consistent, allowing the program to relinquish control to the superior management program (not implemented, see chapter 1) which could review the progress of diagnosis and possible therapies to decide between further diagnosis and immediate therapeutic intervention.

6. Examples Revisited

In this chapter we will consider in detail the two examples described in chapter 2. We will examine *how* the program accomplishes the tasks involved in these examples. Recall that the first example discusses a 40 year old 70 Kg male patient who has been suffering from moderately severe salmonellosis, and as a result, has developed moderately severe metabolic acidosis and hypokalemia. Recall also that the laboratory analysis of the patient's blood sample is:

Fig. 35. Serum electrolytes and the bar diagram

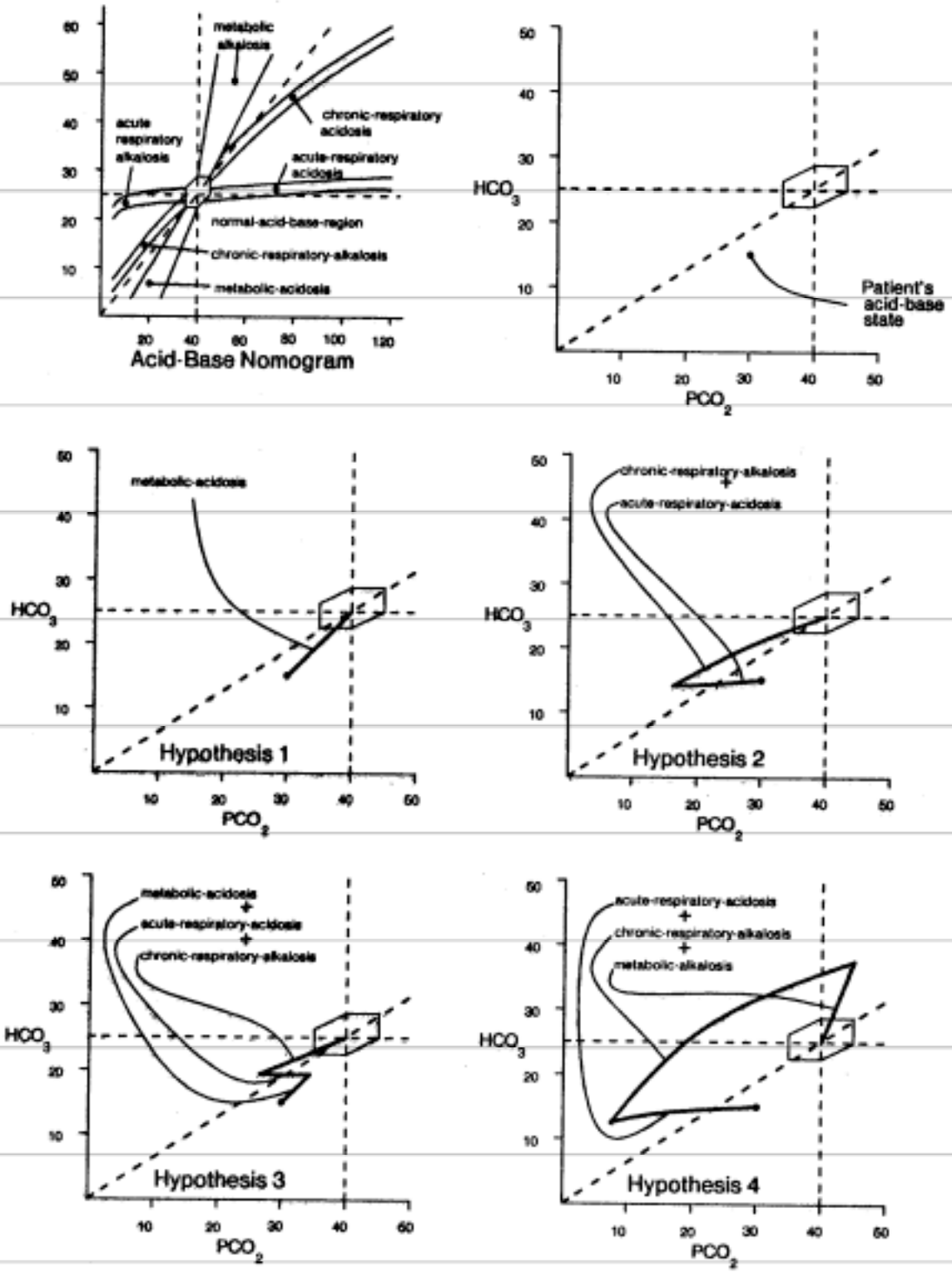
```
Time:   0
 Sex:   male
  Na:   142  meq/l
   K:   3    meq/l
  Cl:   115  meq/l
HCO3:   15   meq/l
PCO2:   30   mmHg
```



The program creates a top level PSD (the root node of the PSD tree) and instantiates the electrolytes in it. This PSD also corresponds to a PSM as it is the only PSD in the PSD tree. Next, the program generates possible acid-base disturbances that can account for the laboratory data. The acid-base analysis is based on the regression equations for the 95 per cent confidence intervals for acid-base disturbances [Schwartz65, Cohen66]. The nomogram of acid-base disturbances, the patient's acid-base state and the possible acid-base disturbances are shown in figure 36. The list of these disturbances is rank-ordered and pruned. The rank-ordering is performed in two stages: first, by the complexity of the disturbance, and second, among the disturbances with same complexity by their severities. For example, the complexity of the second acid-base disturbance is 2 (the number of components in the disturbance) and its severity is $0.75 = (0.68^2 + 0.32^2)^{0.5}$. The rank-ordered list is pruned to remove all the disturbances with more than two components from consideration during the initial formulation.[28] The rank-ordered list of the likely disturbances is:

28. Triple disturbances, although possible, are rare and should be considered only when sufficient evidence demands consideration of triple disturbance, generally after one of the components has been confirmed and the acid-base profile after compensating for the known disturbance still requires at least two further disturbances for proper accounting. Quadruple disturbances are almost never considered in clinical practice.

Fig. 36. Graphical description of acid-base disturbances



```
---- Patient Acid-Base Profile ----
1. metabolic-acidosis [sev: 0.4]                          very likely
2. chronic-respiratory-alkalosis [sev: 0.68]
   + acute-respiratory-acidosis [sev: 0.32]               unlikely
```

The two possible acid-base disturbances provide competing explanations of the serum electrolyte values. The program creates two inferior PSDs under the root PSD. It instantiates, at the clinical level, the nodes corresponding to metabolic-acidosis in the first, and chronic respiratory alkalosis and acute respiratory acidosis in the second (shown in figures 37 and 38). Next, it focally elaborates these nodes to the physiological level (the level at which the instances of electrolyte data are present). For example, in the first PSM the program focally elaborates the metabolic acidosis through the intermediate levels until it reaches the pathophysiological level and identifies the amount of HCO_3 loss corresponding to the severity of the metabolic acidosis. Based on this information and the laboratory data, ABEL instantiates the feedback loop corresponding to the acid-base homeostatic mechanism. Next, it projects backward each node whose cause can be uniquely determined and projects forward the definite consequences of each node in the PSM.[29] We now have the pathophysiological level explanation of the electrolyte abnormalities for each of the two likely acid-base disturbances (shown in figures 37 and 38).

After the pathophysiological description is completed, it is aggregated, one level at a time, to the clinical level of detail. To illustrate this process let us consider the aggregation of the low-serum-K-1 node in PSM 1. Focally aggregating this node, we instantiate hypokalemia-1 as shown in figure 39. Next, we observe that one of the predecessor paths of low-serum-K-1 has an aggregable node on it, namely low-pH-1.[30] We focally aggregate this node to instantiate one of the causes of hypokalemia-1 (acidemia-1) at the next higher level. Note that the other predecessor path from low-serum-K-1 does not have an aggregable node, therefore the component of low-serum-K-1 caused by this path must remain unaccounted for at the next higher level. Next, we compute the component of low-serum-K-1 that can be accounted for by low-pH-1 and the component that remains unaccounted because of the unaccounted ECF-K-loss-2. Then we compute the mapping of these two components at the next level of aggregation and instantiate normokalemia-1 (the component accounted for by low-pH-1) and hypokalemia-2 (due to unaccounted ECF-K-loss-2). We then causally connect the normokalemia-1 to acidemia-1 and mark the hypokalemia-2 as unaccounted. The structure added by the operations described above is shown in bold in figure 39.

---

29. Note here that since we are at the pathophysiological level, each link being projected is primitive. Thus, projecting back a node at this level is equivalent to instantiating the cause and the link connecting the cause and the effect node.

30. A node is aggregable if in the medical knowledge-base it is the focus of the elaboration structure of some node at the next higher level which can be instantiated within the PSM. Otherwise, the node is not aggregable.

Fig. 37. Hypothesis 1

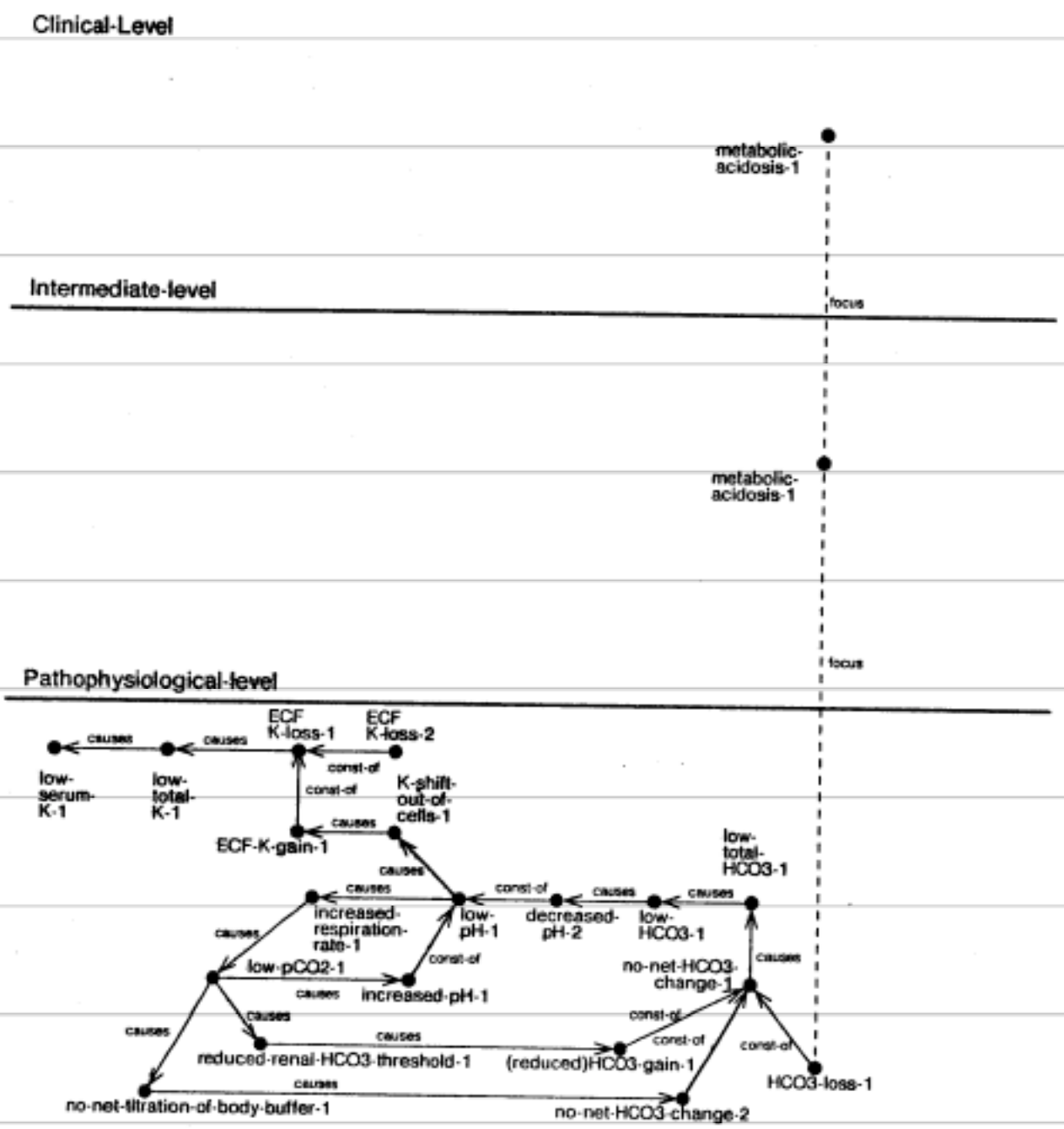


Fig. 38. Hypothesis 2

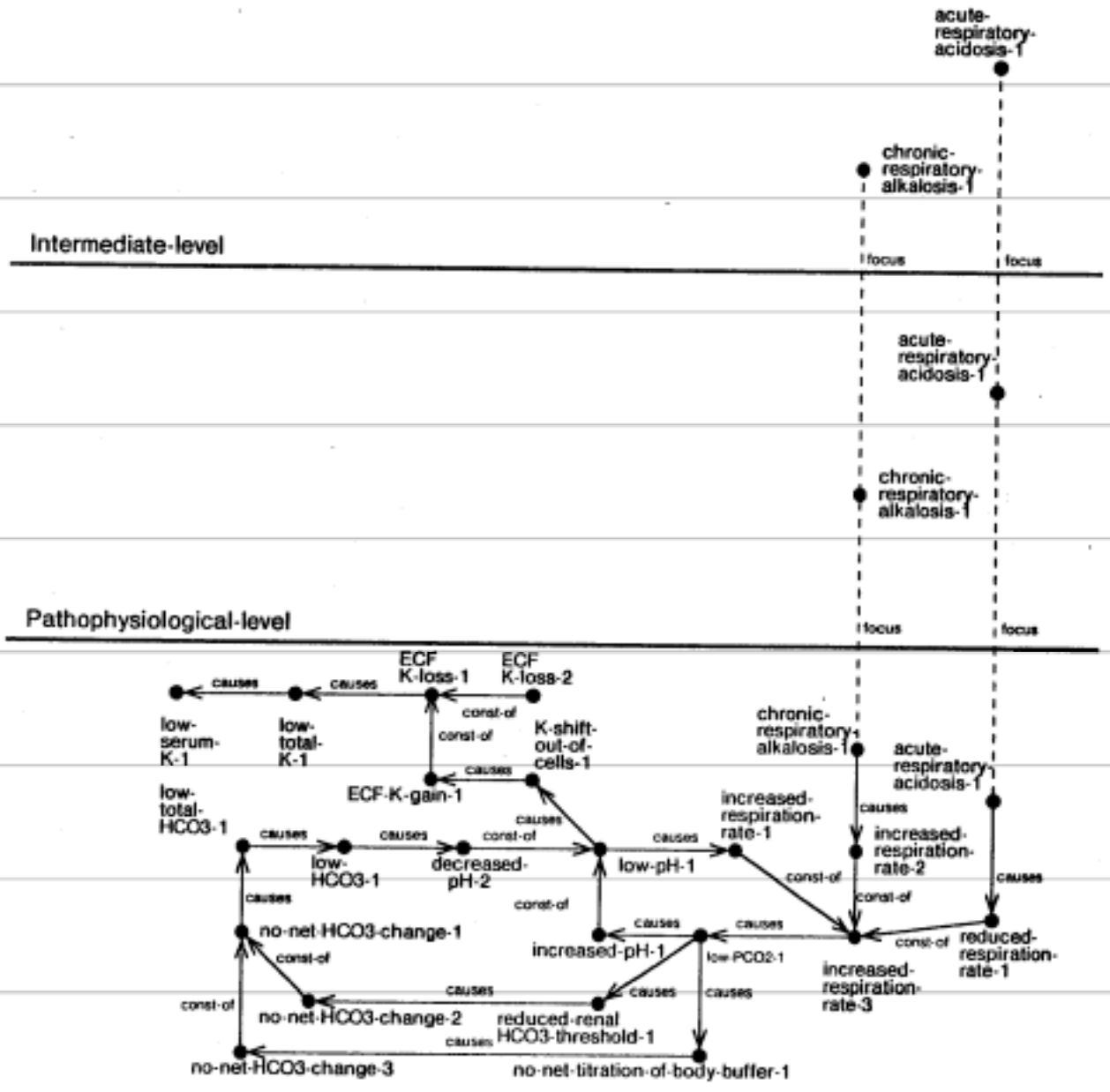


Fig. 39. aggregation of low-serum-K-1

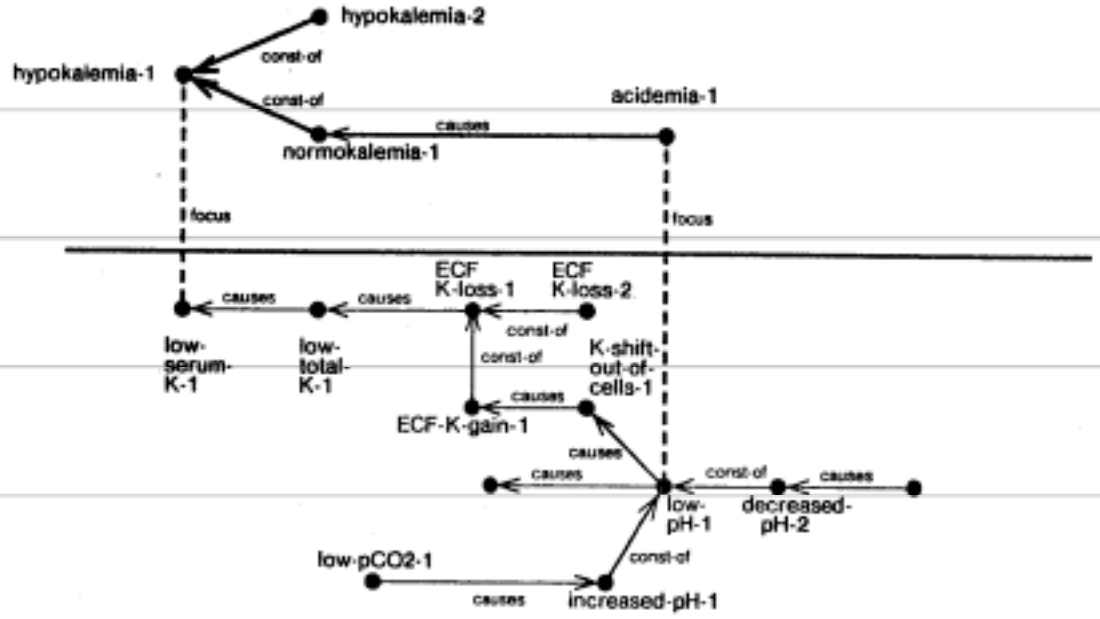


Fig. 40. Aggregation of low-pH-1

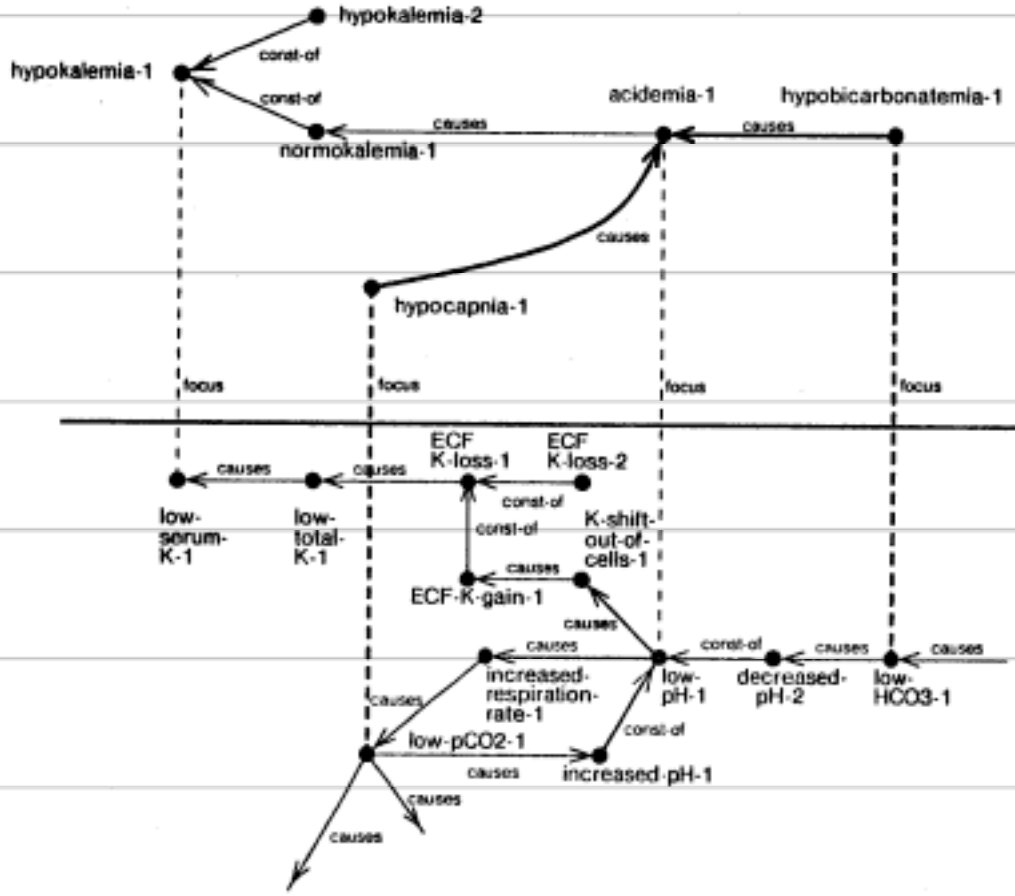


Fig. 41. PSM for hypothesis 1

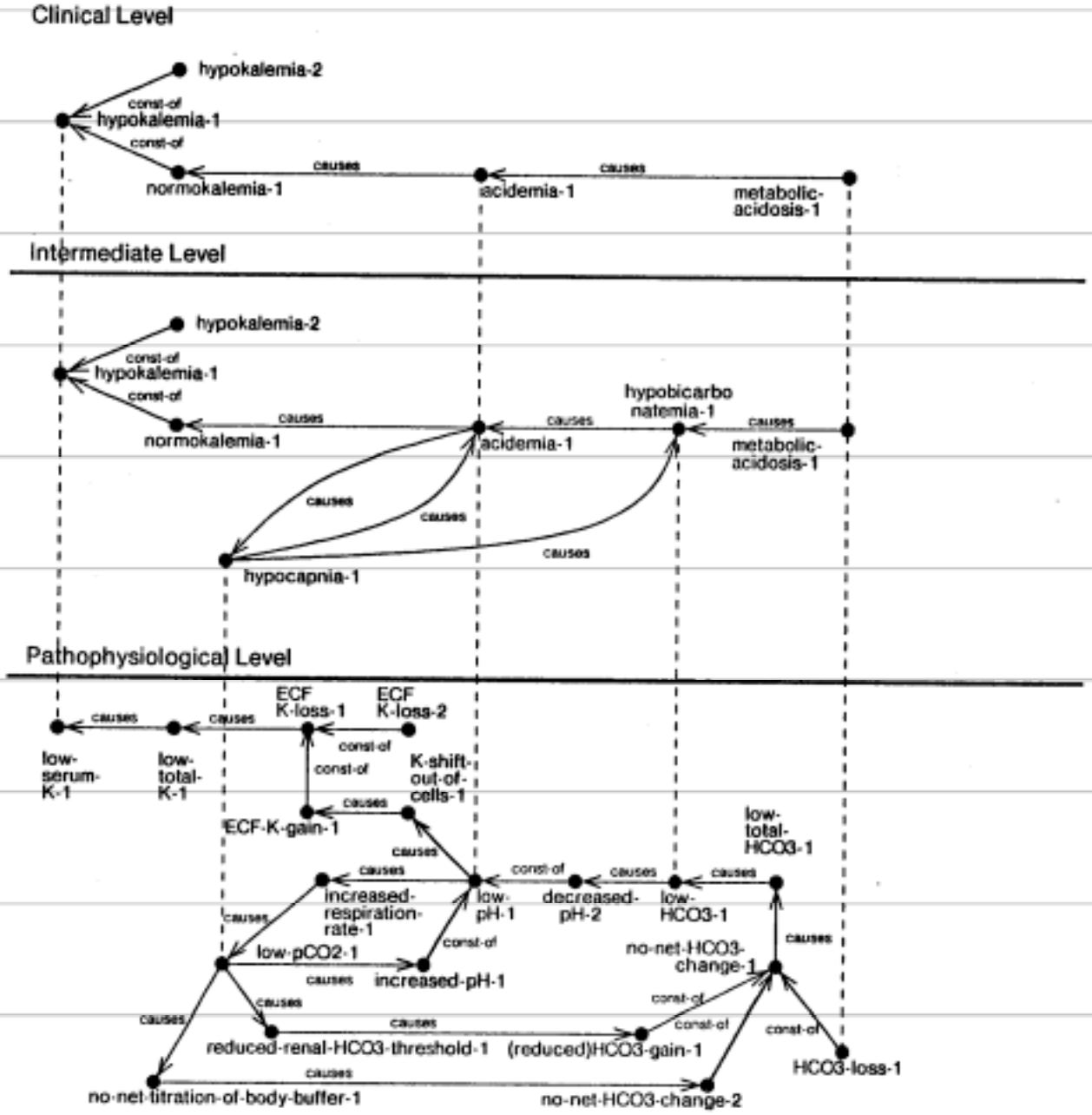
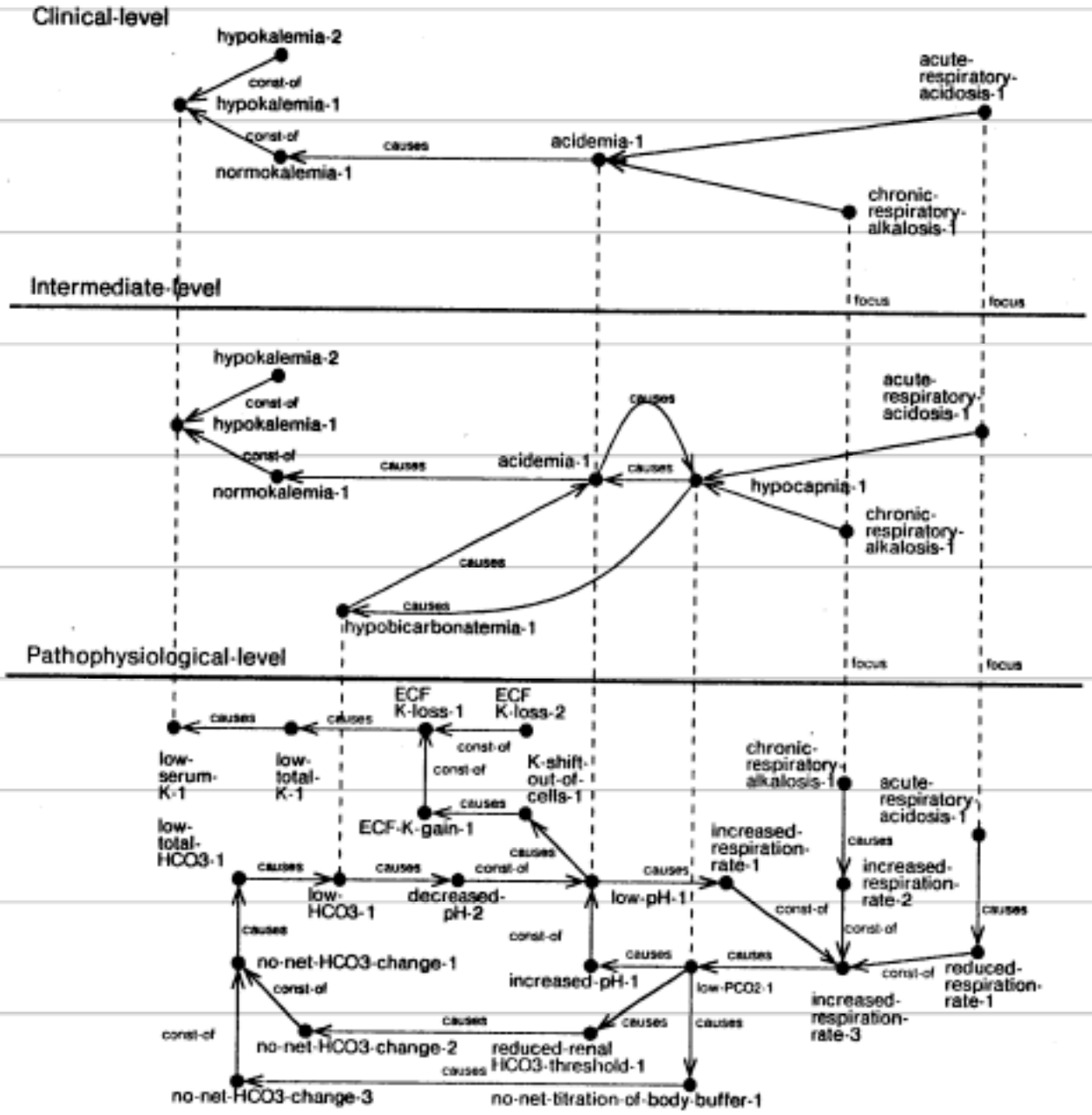


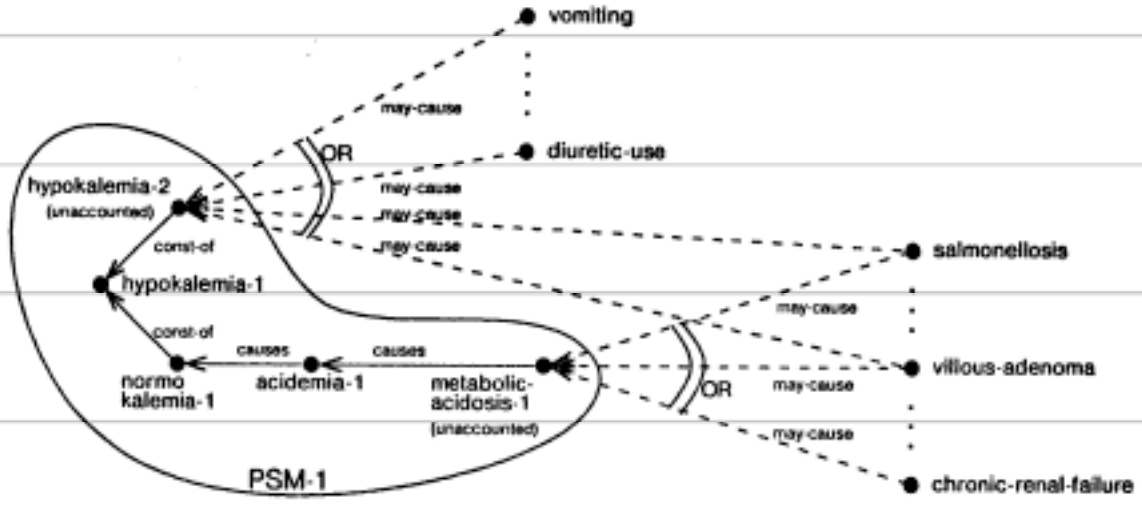
Fig. 42. PSM for hypothesis 2



Next, let us consider the causal aggregation of low-pH-1 shown in figure 40. As each of the paths leading back from low-pH-1 has an aggregable node (low-pCO2-1 and low-HCO3-1),[31] the focal aggregation of low-pH-1 (acidemia-1) is a fully accounted node. The causal aggregation is achieved by focally aggregating the low-pCO2-1 and low-HCO3-1 into hypocapnia-1 and hypobicarbonatemia-1 respectively, and by causally connecting hypocapnia-1 and hypobicarbonatemia-1 to acidemia-1. This process is repeated for each aggregable node at the current (pathophysiological) level and then the whole process is repeated at the next level until we reach the clinical level of aggregation. The resulting structures for the two acid-base hypotheses (encoded by the two PSMs) are shown in figures 41 and 42.

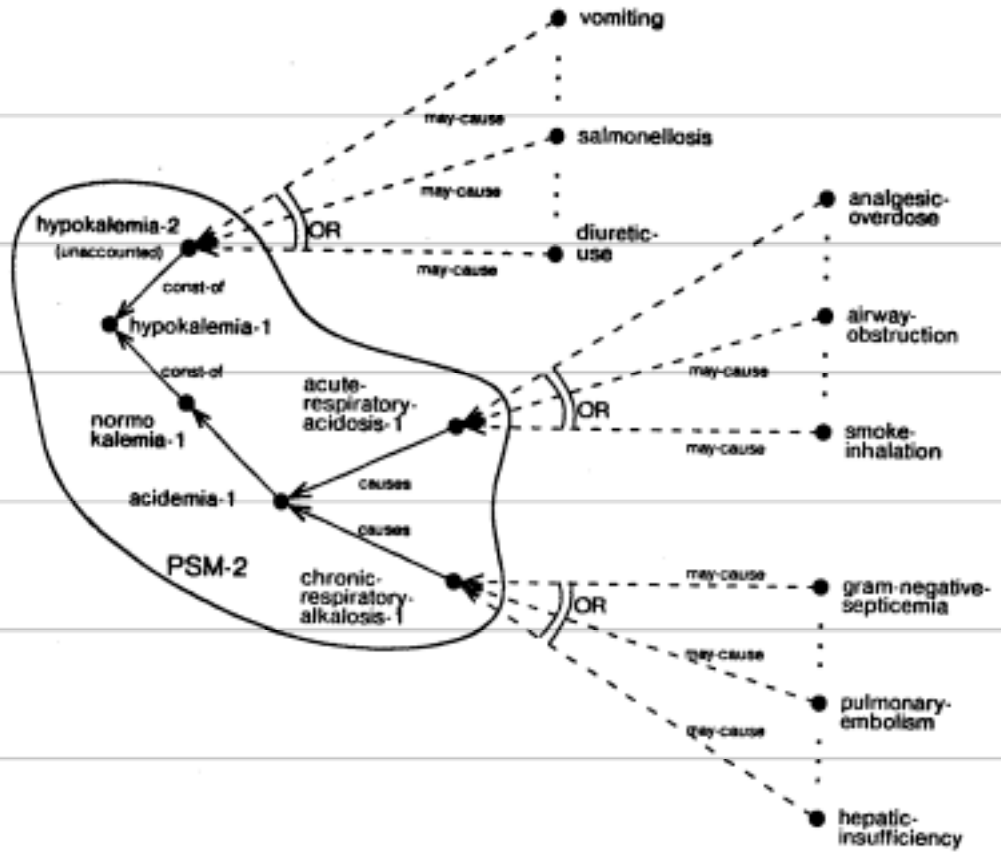
As discussed in chapter 2, a comparison of the clinical level explanations shows that the two PSMs share the structure involving hypokalemia and acidemia. They differ in their accounting for acidemia. Note that the acid-base feedback cycles present at the pathophysiological and intermediate levels have been abstracted away by the aggregation process and the two clinical level descriptions are fairly simple. A comparison of the intermediate level descriptions shows that they differ principally in the way the acid-base feedback cycle is perturbed. In the first case, the change in acid-base state is a consequence of addition of H^+ to the body which causes hypobicarbonatemia, whereas in the second it enters as primary disturbance in ventilation which alters the CO_2 tension. The pathophysiological level differences between the two cases can be identified similarly by comparing the two pathophysiological level descriptions. Finally, note that the first PSM has two unaccounted findings while the second PSM has three unaccounted findings.

Fig. 43. Diagnostic closure 1



31. The search terminates when the program finds the first aggregable node on each path.

Fig. 44. Diagnostic closure 2



In the context of this initial analysis the program starts its diagnostic exploration. It computes the diagnostic closures for the two hypotheses (DC-1 and DC-2 shown in figures 43 and 44), and formulates the top level goal of pursuing DC-1. One complete cycle of diagnostic inquiry is shown in figure 45.

As a first step towards exploring DC-1, the program groups the disease hypotheses according to the number of unexplained findings each disease hypothesis can explain. For example, the salmonellosis hypothesized to account for moderately severe metabolic acidosis can also account for the hypokalemia. Therefore, the hypothesized salmonellosis can account for all the unaccounted findings in PSM-1. However, if the patient had very severe metabolic acidosis and mild hypokalemia, the salmonellosis hypothesized to account for metabolic acidosis would not have been consistent with hypokalemia. In such a case we would have had to hypothesize two separate instances of salmonellosis, each accounting for only one of the two unaccounted findings. Subsequently, each of the two instances of salmonellosis would have been grouped with disease hypotheses accounting for only one unaccounted finding.

Fig. 45. One complete cycle of diagnostic inquiry

```
Differentiating between the causes of                          (a)
the leading complete hypothesis.

     1    SALMONELLOSIS
     2    URETEROSIGMOIDOSTOMY
     3    VILLOUS-ADENOMA
     ----------------------------
     4    DISTAL-RTA
     5    PROXIMAL-RTA
     6    ACUTE-RENAL-FAILURE
     7    CHRONIC-RENAL-FAILURE
continue? ==> y

Does the patient have any of the following?                    (b)
     1    SALMONELLOSIS
     2    URETEROSIGMOIDOSTOMY
     3    VILLOUS-ADENOMA

Present: ==> none     Absent: ==> none     Unknown: 1 2 3

I would like to ask about the effects of SALMONELLOSIS.

Does the patient have one of the following?                    (c)
     1    DEHYDRATION
     2    EDEMA

Present: ==> none     Absent: ==> none     Unknown: 1 2

Is the value of SERUM-CREATININE known? ==> yes                (d)

Please enter the attributes of SERUM-CREATININE
    What is the VALUE of SERUM-CREATININE ? ==> 3
    What is the START-TIME of SERUM-CREATININE ? ==> 0

Is the value of MEAN-ARTERIAL-BLOOD-PRESSURE known? ==> yes    (e)

Please enter the attributes of MEAN-ARTERIAL-BLOOD-PRESSURE
    What is the value of MEAN-ARTERIAL-BLOOD-PRESSURE ? ==> 75
```

Next, the diseases in the same group are rank-ordered based on their scores computed from the three factors, match, mismatch and unknown (described in section 5.4). Those hypotheses which have higher mismatch than match are not considered. For example, consider the scoring of the vomiting hypothesized to account for unaccounted hypokalemia. The vomiting so hypothesized matches the hypokalemia. However, the hypothesized vomiting predicts metabolic-alkalosis which is inconsistent with the observed metabolic acidosis. Furthermore, if vomiting were really observed in the patient, the additional amount of HCO_3 loss necessary to cause the observed state would require a very severe cause of metabolic acidosis to be present. Therefore, vomiting has a substantially higher mismatch factor as compared to the match and it is rejected. The program deletes the hypotheses that have been rejected and rank-orders the remaining as shown in figure 45(a).

Based on the categorization of the disease hypotheses, ABEL decomposes the diagnostic problem into two groups by constructing two separate diagnostic closures (DC-3 and DC-4). DC-3 (shown in figure 46) contains disease hypotheses 1 to 3, and DC-4 contains disease hypotheses 4 to 7. It projects forward the disease hypotheses in each of the two diagnostic closures to identify their unobserved findings. Next, it sets up a goal to differentiate between the three hypotheses in DC-3. As the first step towards this differentiation, the program asks if the user is already aware of any of the possible alternatives as shown in figure 45(b).

When none of the three hypotheses can be directly confirmed, the program pursues the task of differentiating between the three further. It sets up an individual diagnostic closure for each of

Fig. 46. Diagnostic closure 3

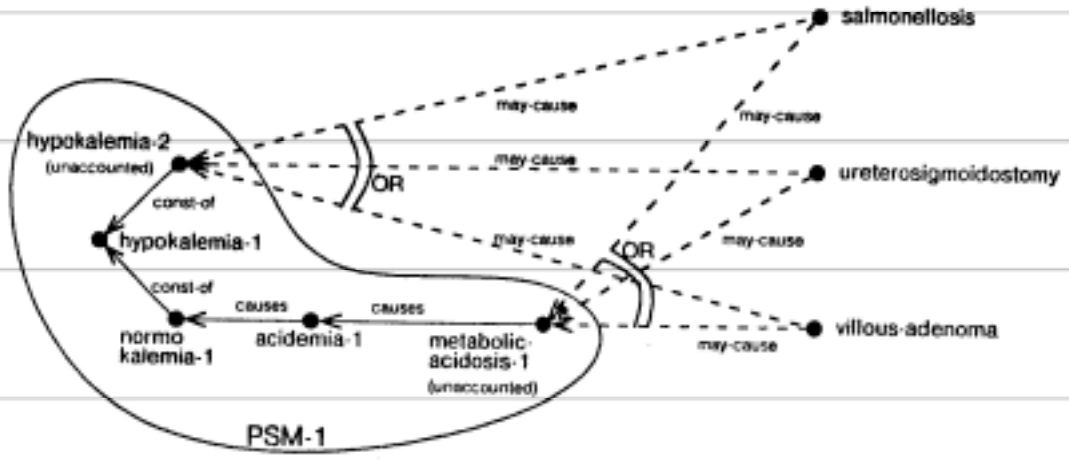


Fig. 47. Diagnostic closure 4

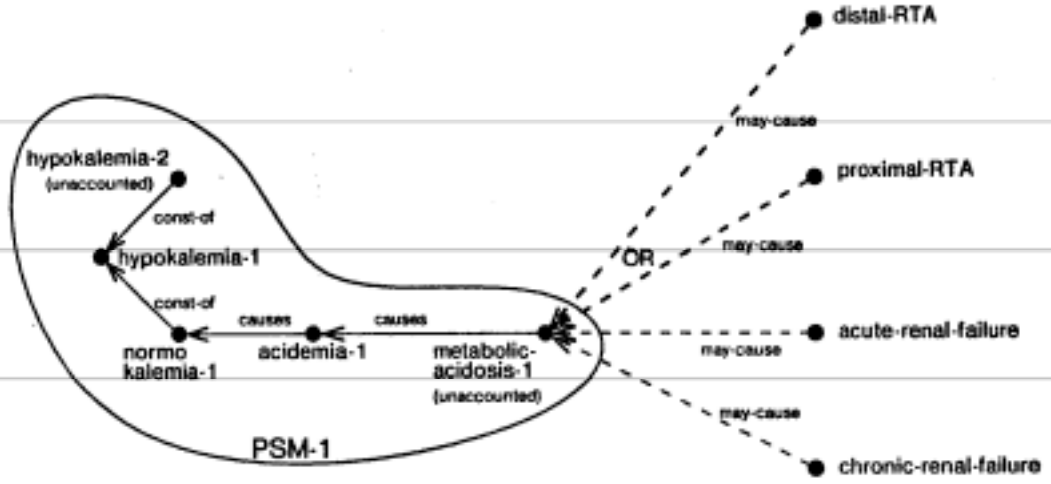
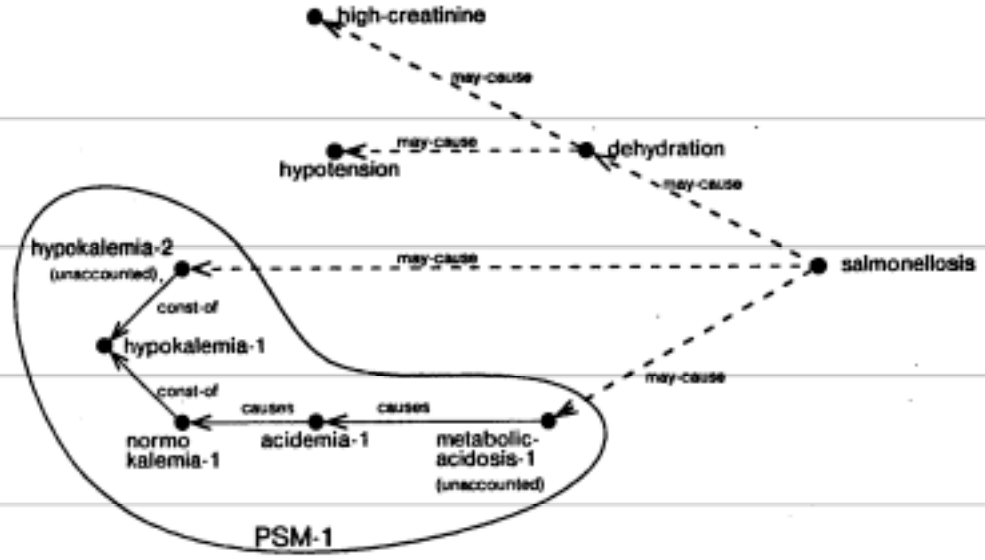


Fig. 48. Diagnostic closure 5



the three alternatives (DC-5, DC-6, and DC-7) and selects the next item (dehydration) for inquiry.[32] Note that salmonellosis, ureterosigmoidostomy and villous-adenoma all cause dehydration. However, the program also notices that some of the diseases in DC-4 (e.g., chronic-renal-failure) may have the exact opposite effect of causing edema. Therefore, while exploring dehydration

32. These three hypotheses could be differentiated very easily on the basis of history and clinical evidence. For the simplicity of the example, we assume that this information is not available to the program.

(state of extracellular fluid volume) the program includes edema in the question (shown in figure 45(c)). The program is expecting dehydration. Therefore, when it fails to confirm or deny the dehydration it pursues the finding further (figure 45 (d) and (e)).

The program has now completed one full cycle of its diagnostic inquiry. It incorporates the information gained during this cycle in both the PSMs. Note that the program has already gathered sufficient information to confirm salmonellosis. It is unable to do so because we have not implemented the criteria for confirming a disease yet.[33] Therefore, the program starts the new cycle of diagnostic planning in which it attempts to rule out all other possible causes of the acid-base disturbance. Finally, when it has exhausted all the findings relevant to the diagnosis of this case, it concludes that salmonellosis is the leading candidate and asks if the user would like to assume salmonellosis (shown in figure 49).

The program adds salmonellosis to the patient models and re-evaluates the two hypotheses. The process of assimilating salmonellosis into the PSMs is described next. Let us first consider the operation of causally connecting salmonellosis with metabolic-acidosis in PSM-1. As the observed salmonellosis is consistent with the metabolic acidosis, a causal link from salmonellosis to metabolic acidosis is established at the clinical level. The elaboration operator is used to establish this relation at the more detailed levels (the resulting structure is shown in figure 50).

Fig. 49. After all findings have been exhausted

```
     All possible etiologies that could explain the patient's
illness are unknown.  In order to proceed we must at least
hypothetically assume one of them.  Possible etiologies that could
explain the patient's illness listed in decreasing order are:

     1    SALMONELLOSIS
     ----------------------------------
     2    VILLOUS-ADENOMA
     3    URETEROSIGMOIDOSTOMY
     ----------------------------------
     ----------------------------------
     4    DIABETES-INSIPIDUS
     ----------------------------------
     5    DISTAL-RTA
     6    PROXIMAL-RTA

Would you like to assume SALMONELLOSIS ?      ==> yes

Assuming MODERATE ACUTE SALMONELLOSIS.
```

33. A simple criterion for confirming a disease similar to that in PIP or MYCIN can easily be added to the program. However, we have chosen not to do so because of two reasons: first, because the choice of threshold for confirming a disease is arbitrary and therefore, very difficult to explain, and second, in the electrolyte and acid-base program we envision this to be the task of the global decision-making module.

Fig. 50. Hypothesis 1 with salmonellosis

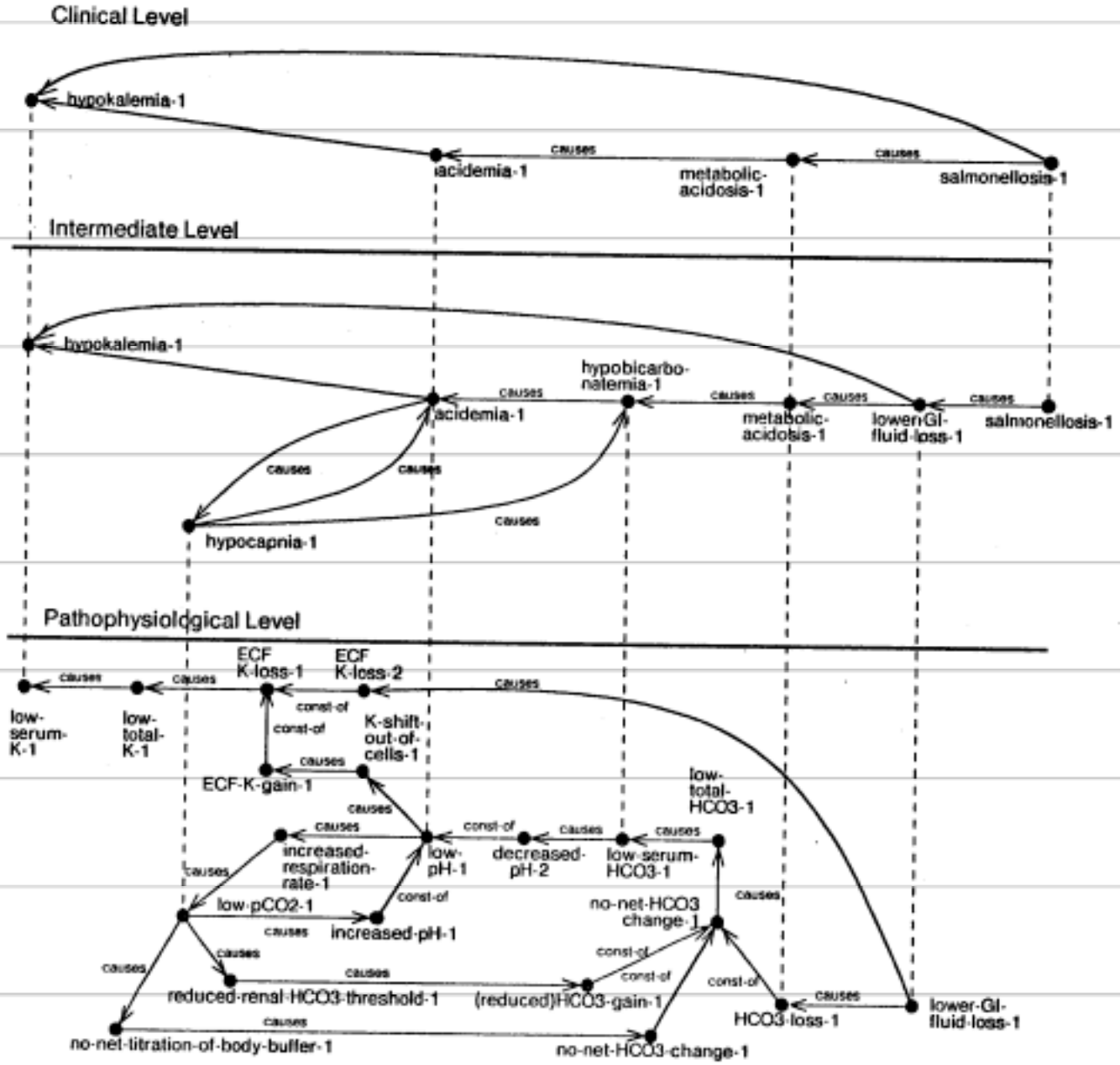
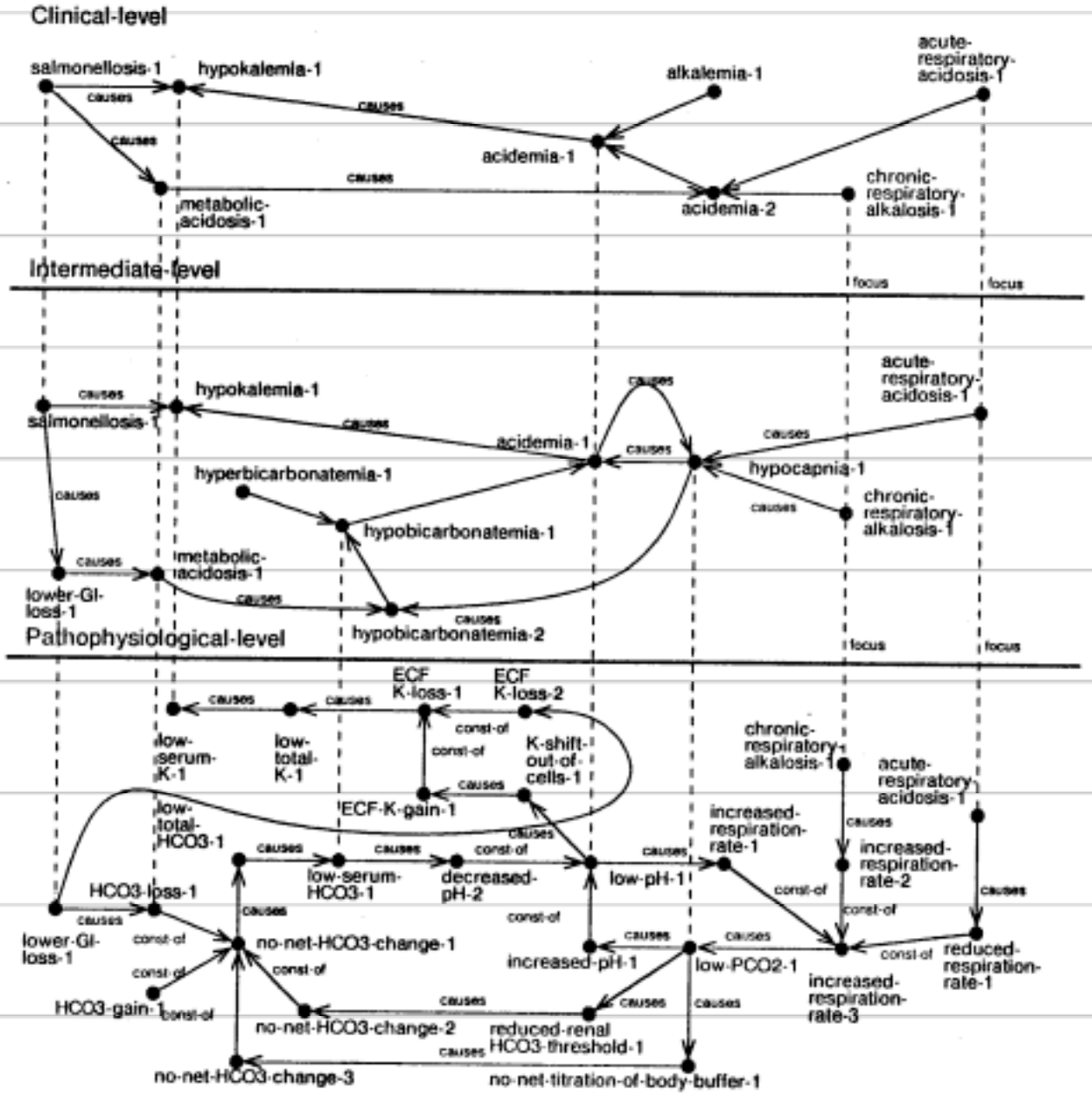


Fig. 51. Hypothesis 2 with salmonellosis



The elaboration process begins with the focal elaboration of salmonellosis to the intermediate level. The focus nodes of the source and the destination of the link being elaborated (salmonellosis —causes—> metabolic-acidosis) are now present at the next level. Next, ABEL attempts to match the causal path associated with the link at the next level of detail, namely salmonellosis —causes—> lower-GI-loss —causes—> metabolic-acidosis. As this path does not exist at the intermediate level, ABEL must establish this path and then proceed to elaborate each link in it. Let us first consider the link salmonellosis-1 —causes—> lower-GI-loss-1. As salmonellosis is a primitive node at this level (it does not have a focal node at the next lower level), the link between salmonellosis and lower-GI-loss is a primitive link and cannot be elaborated any further. The next link, lower-GI-loss-1 —causes—> metabolic-acidosis-1 however, can be elaborated further. This is done by first focally elaborating lower-GI-loss-1 at the intermediate level to the pathophysiological level, and second by connecting, at the pathophysiological level, lower-GI-loss-1 to HCO3-loss-1. As the remaining links in the causal path at this level are already present, this completes the process of elaboration. Next, the newly created structure is causally aggregated to propagate the consequences of the lower level additions back up to the clinical level. The results of assimilating salmonellosis into the two PSMs are shown in figures 50 and 51.

A comparison of PSM-1 and PSM-2 shows that PSM-1 contains only one acid-base disturbance, while PSM-2 contains three acid-base disturbances. All the findings in PSM-1 have been accounted for, while PSM-2 has three nodes that still need to be accounted for. Therefore, based on the assumption that the patient is suffering from moderately severe salmonellosis, ABEL concludes that PSM-1 provides an adequate explanation of the patient's illness. The computer generated English descriptions of the clinical levels of the two PSMs are shown in figure 52.

The second example deals with a patient suffering from moderately severe vomiting and salmonellosis. Recall that the electrolyte and acid-base disturbance in vomiting results from an excessive loss of upper gastrointestinal fluid, whereas in salmonellosis it results from an excessive loss of lower gastrointestinal fluid. The upper GI fluid is acidic while the lower Gi fluid is alkaline, therefore the two tend to have offsetting effects on the acid-base balance. However, vomiting and salmonellosis both cause hypokalemia and volume depletion, therefore they compound the effects of each other.

In this example, we will consider the presentations of vomiting and salmonellosis such that they exactly offset the acid-base effects of each other, leaving the patient with no net change in pH. We will demonstrate the program's capabilities in dealing with multiple etiologies and in reformulating its patient description when new information is provided. We will illustrate this by describing the program's understanding of the case at three points during the diagnostic process: (1) just after the initial presentation of electrolytes, (2) after the program has identified the first of the two diseases, namely vomiting, and (3) at the end of the diagnostic process.

Fig. 52. English description of the two hypotheses

The Successful Explanation

This is a 40 year old 70.0 kg male patient with moderate salmonellosis. His electrolytes are:

Na: 142.0	HCO3: 15.0	Anion Gap: 13.0
K: 3.0	pCO2: 30.0	
Cl: 115.0	pH: 7.32	Creatinine: 3.0

The salmonellosis causes moderate metabolic acidosis and moderate dehydration. The dehydration causes moderate hypotension and moderately high creatinine disturbance. The metabolic acidosis causes mild acidemia. The salmonellosis and acidemia cause mild hypokalemia. All findings have been accounted for.

The Alternate Explanation

This is a 40 year old 70.0 kg male patient with salmonellosis. His electrolytes are:

Na: 142.0	HCO3: 15.0	Anion Gap: 13.0
K: 3.0	pCO2: 30.0	
Cl: 115.0	pH: 7.32	

The salmonellosis causes moderate metabolic acidosis and moderate dehydration. The dehydration causes moderate hypotension and moderately high creatinine disturbance. Moderate acute respiratory acidosis, moderate chronic respiratory alkalosis and metabolic acidosis partly compensate the suspected mild alkalemia leading to the observed mild acidemia. The salmonellosis and acidemia cause mild hypokalemia. The chronic respiratory alkalosis and acute respiratory acidosis remain to be accounted for. The alkalemia has only been partially accounted for.

The program's initial evaluation of the patient's electrolytes is as follows:

---- Patient Acid-Base Profile ----
1. normal-acid-base-state

This is a 40 year old 70 Kg male patient with moderate hypokalemia. His electrolytes are:

Na: 143.0	HCO3: 25.0	Anion Gap: 12.0
K: 2.0	PCO2: 39.0	
Cl: 108.0	pH: 7.42	

The hypokalemia remains to be accounted for.

The initial PSM (PSM-1) created by the program is shown in figure 53. Note that the clinical level of the PSM contains only one abnormal finding, hypokalemia. Figure 54 shows the revised PSM after vomiting has been introduced. A detailed description of this process of revision is considered next.

Based on the information in the diagnostic closure the program concludes that the vomiting present cannot account fully for the hypokalemia. However, as the vomiting can partially account for the hypokalemia (leaving a smaller amount unaccounted for), the program decides to project forward, to identify the quantity of hypokalemia accounted for by it. The projection process begins with the focal elaboration of vomiting-1 from the clinical level to the intermediate level. Next, the program matches the causal path associated with the link, i.e., vomiting —causes—> upper-GI-loss —causes—> hypokalemia. As this path is not inconsistent with the PSM, the program recurs on each link in the path. The first link, vomiting —causes—> upper-GI-loss, is a primitive link. Therefore, the program instantiates the upper-GI-loss (upper-GI-loss-1) and the link connecting it upwards to vomiting-1. The second link, upper-GI-loss —causes—> hypokalemia, is a compound link. The path associated with this link at the next level is upper-GI-loss —causes—> ECF-K-loss —causes—> low-total-ecf-K —causes—> low-serum-K. Matching this path with the description in the PSM, the program finds that all but one link, upper-GI-loss —causes—> ECF-K-loss, is already present. Since this link is primitive, the program revises the component structure of ECF-K-loss-1 and instantiates the link between ECF-K-loss-4 and upper-GI-loss-1. Note that as soon as this link is instantiated the path at the pathophysiological level is complete. The program aggregates back the effects of the projection process to reflect the additions at the lower levels at the upper levels.

An important side-effect occurs when the program is reasoning (at the pathophysiological level) about the quantity of ECF-K-loss associated with the upper-GI-loss. As the ECF-K-loss is dependent on the quantity of upper gastrointestinal fluid loss, this loss must be accompanied by the loss of corresponding amounts of the other electrolytes present in the upper-GI-fluid, notably the loss of H^+.[34] This fact is incorporated into the PSM, causing the program to revise its acid-base hypothesis. This hypothesis now contains two components: an alkalemia (metabolic-alkalosis) caused by vomiting, and an acidemia (unaccounted) which cancels the effects of alkalemia leaving the patient in a normal acid-base state as shown in figure 54. Thus, the PSM after vomiting contains two unaccounted nodes: the unaccounted component of hypokalemia (less severe than before vomiting was introduced), and acidemia which must be present to offset the metabolic-alkalosis caused by vomiting.

---

34. The loss of H^+ from the extracellular fluid can be viewed as gain in HCO_3^-, because as the H^+ is removed from the carbonic acid — bicarbonate buffer an equivalent amount of HCO_3^- is released into the fluid.

Fig. 53. Initial PSM

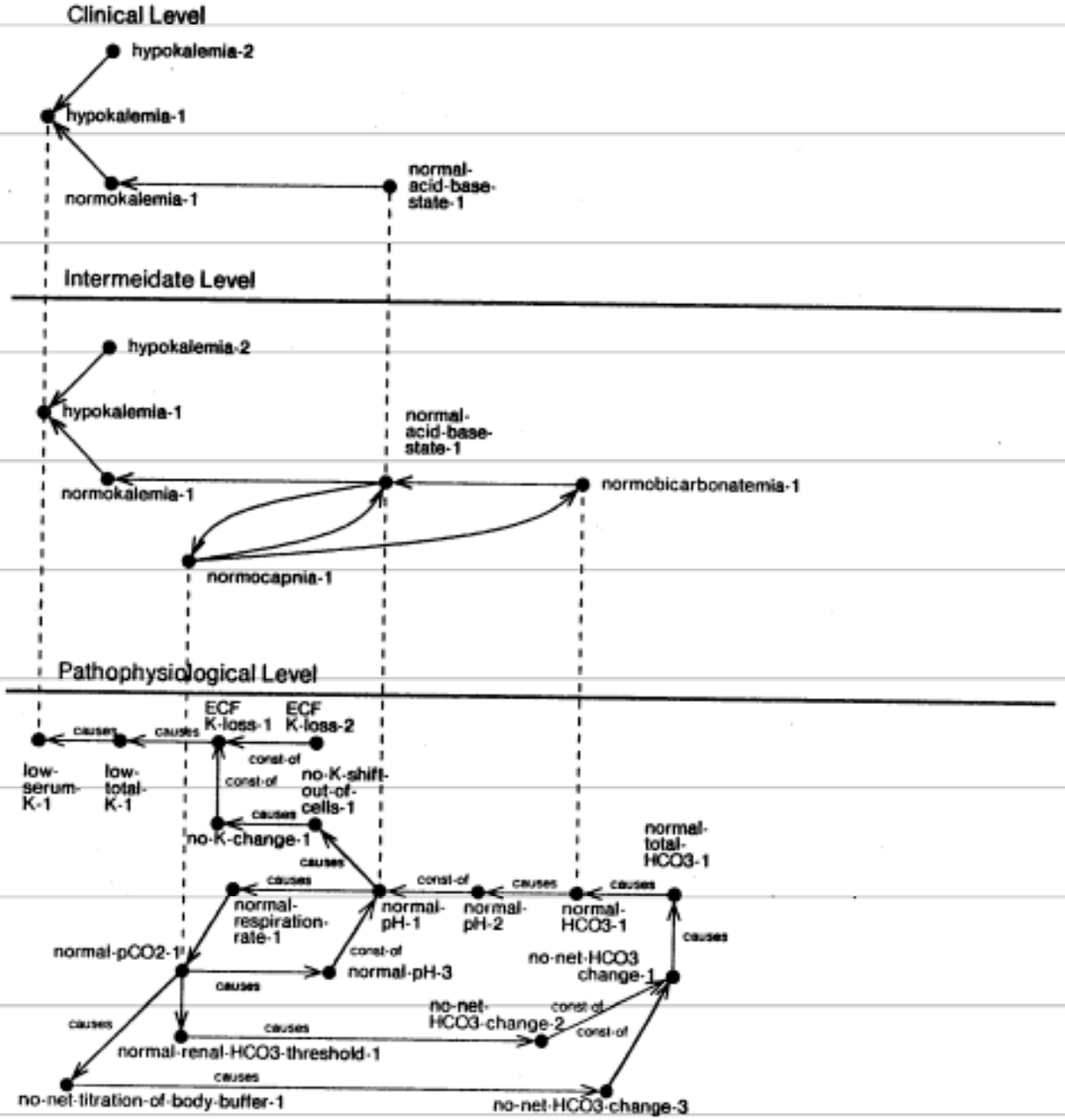


Fig. 54. Revised PSM after vomiting is entered

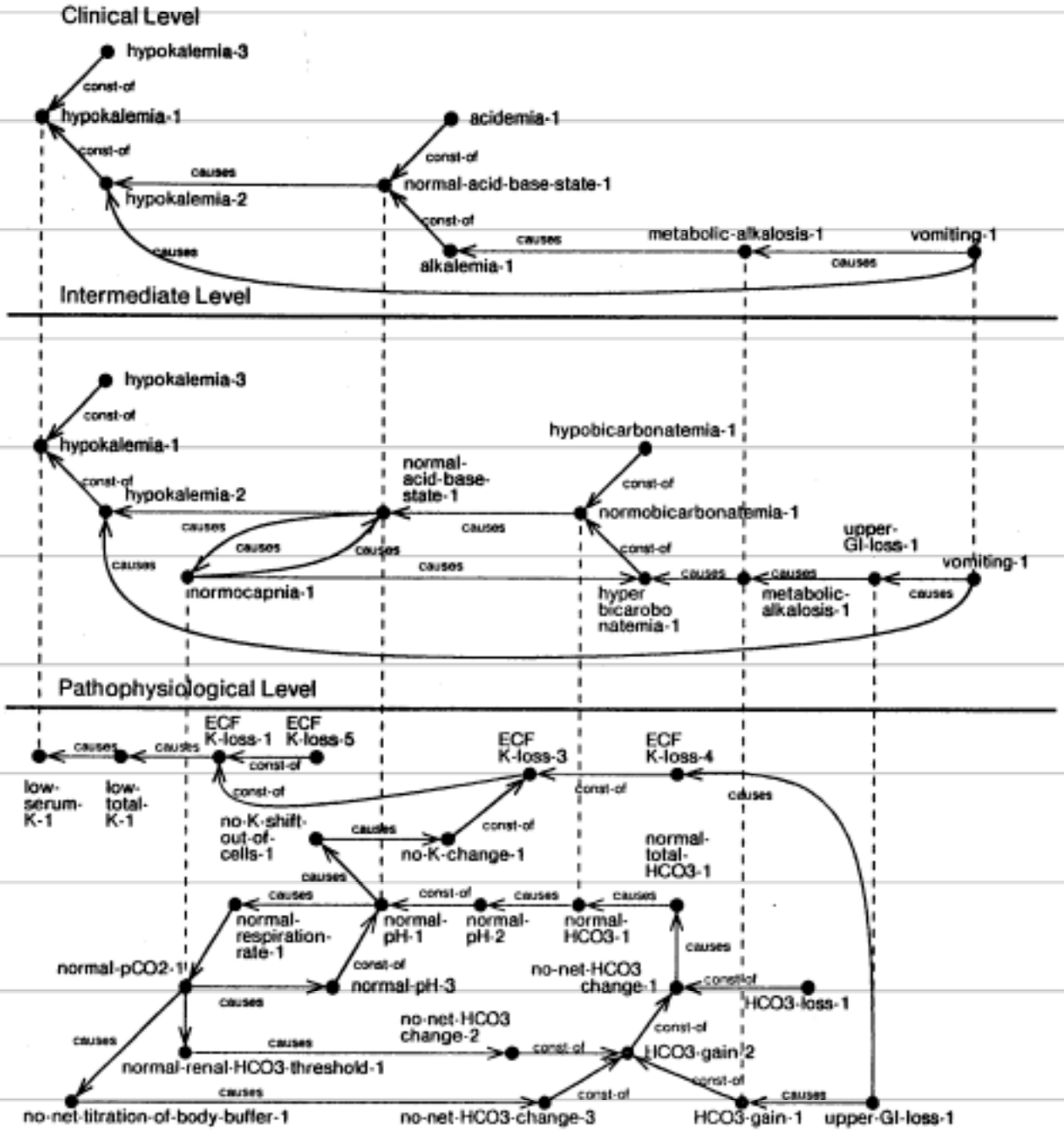


Fig. 55. Final PSM after salmonellosis is introduced

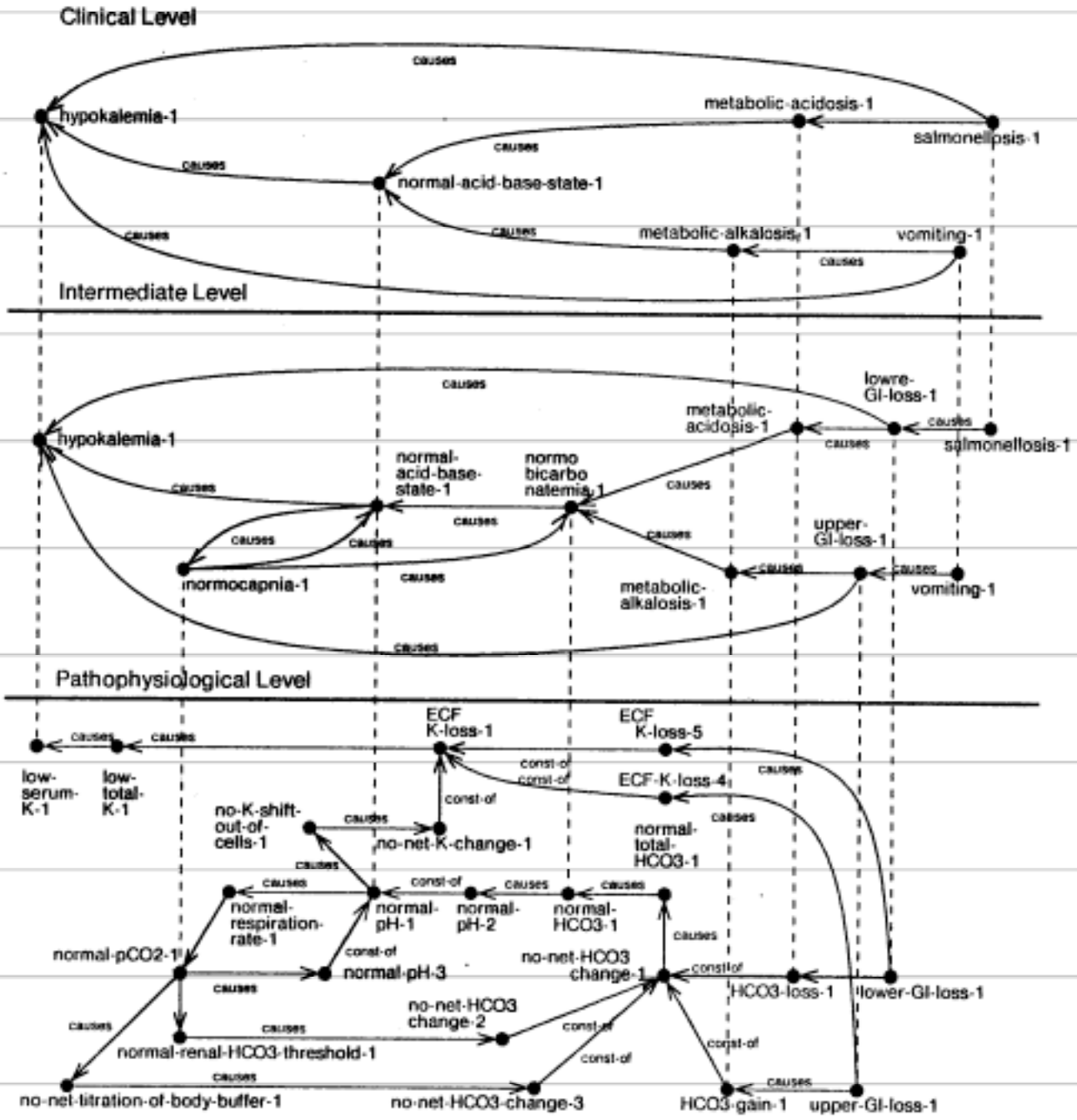


Fig. 56. English text of the final explanation

Clinical Level

```
     This is a 40 year old 70.0 kg male patient with moderate
vomiting and moderate salmonellosis.  His electrolytes are:

        Na: 143.0          HCO3: 25.0        Anion Gap: 12.0
         K: 2.0            pCO2: 39.0
        Cl: 108.0            pH: 7.42        Creatinine: 3.0

     The vomiting causes moderate metabolic alkalosis.  The salmonellosis
and vomiting cause moderate dehydration, which causes hypotension.
The dehydration also causes moderately high  creatinine disturbance.
The salmonellosis causes moderate metabolic acidosis.  The metabolic
acidosis and metabolic alkalosis cause  normal ph.  The salmonellosis,
and vomiting cause moderate  acute hypokalemia.  All findings have
been accounted for.
```

Intermediate Level

```
     This is a 40 year old 70.0 kg male patient.  His electrolytes
are: ....

     The patient has moderate vomiting and moderate salmonellosis.
The vomiting causes moderate upper gi loss, which causes moderate
metabolic alkalosis.  The salmonellosis causes moderate lower gi loss.
The lower gi loss and upper gi loss cause moderate dehydration,
which causes hypotension.  The dehydration also causes moderate
high creatinine disturbance.  The lower gi loss causes moderate
metabolic acidosis.  The metabolic acidosis and metabolic alkalosis
cause normobicarbonatemia. The normobicarbonatemia and normocapnia
cause normal ph.  The lower gi loss and upper gi loss cause moderate
hypokalemia.  All findings have been accounted for.
```

Note that the two unaccounted components of the PSM are the same as those present in PSM-1 of the first example. We have been successful in separating the effects of vomiting from the remaining disturbance (salmonellosis). As might be expected, from here on the diagnosis of this case is similar to that of the first example. The final diagnosis after salmonellosis has been added to the PSM is shown in figure 55. Figure 56 shows the program's English explanation of the final diagnosis at two different levels of detail.

7. Conclusion

7.1 Summary

Each new scientific endeavor is built on previous endeavors, consolidating their successes and learning from their shortcomings. This is no exception; we have drawn heavily from first generation AIM programs. This thesis has benefited from the studies of clinical skills of physicians, by introspection and by observing each other, development of models of diagnostic processes and their implementation using computers by Schwartz, Pauker, Gorry, Kassirer, Szolovits and others. Implementation of the Present Illness Program and analysis of its performance was an important first step for the research presented here. Experience with PIP and the other first generation AIM programs exposed the need for substantially more detailed and categorical reasoning in diagnostic programs and provided an ideal environment in which to explore the issues addressed in this thesis.

The research presented in this thesis was also influenced by the discussions of difficult diagnostic cases at the informal Electrolyte and Acid-Base rounds at the Tufts New England Medical Center Hospital. The most striking aspect of these discussions was the frequent use, by the clinicians, of the pathophysiological knowledge in evaluation and justification of diagnostic hypotheses, and the ease with which they were able to combine knowledge of global diagnostic associations such as "disease X is a common complication in a patient with a history of Y" with intricate physiological deductions such as "Na^+ and K^+ exchange in the distal tubule is coupled with the excretion of H^+, therefore increased distal delivery of Na^+ enhances..." These observations strengthened our conviction that in order to begin to approach the level of competence of an expert a computer program must possess a similar ability. It must be able to reason simultaneously with phenomenological knowledge about disease associations and with the best available pathophysiological knowledge about disease mechanisms. Much of our effort has been focused on building representational and procedural mechanisms to provide such a capability. The emphasis has been on the development of multi-level causal descriptions of a patient's illness and on the development of techniques for composing/decomposing effects with multiple causes (described in chapters 3 and 4). We believe this approach provides our program with a level of understanding of disease not present in the first generation of AIM programs.

The study of clinical problem solving activity by Elstein [Elstein78], Kassirer [Kassirer78] and others suggests that a physician's diagnostic reasoning is strongly guided by structural notions such as "coherence" and "adequacy". Each diagnostic alternative entertained by a physician is a mosaic of connected hypotheses, together accounting for the observable aspects of the patient's illness. This thesis describes the use of a coherent hypothesis as the logical unit of hypothesis representation (represented as a PSM). A PSM is a collection of causally connected disease hypotheses and findings, providing a (perhaps partial) explanation of the patient's illness. A set of alternative diagnoses consistent with a PSM is represented using a *diagnostic closure*. A

diagnostic closure unites all the dependencies and expectations necessary for diagnostic inquiry, for selecting appropriate questions, and for evaluating the information received in response to the questions.

Expert clinicians employ a variety of diagnostic strategies for an efficient exploration of the diagnostic space. Some of the first generation programs, notably INTERNIST-I, use similar strategies to guide their diagnostic exploration. However, their lack of commitment in pursuing a given strategy to completion results in unfocused and inefficient use of these strategies. This problem can be alleviated by allowing the diagnostic problem solver to plan a sequence of questions focused around a diagnostic task before embarking on an inquiry. In this thesis we have described a simple diagnostic planner which formulates a tree structured plan. The planning begins with the global task of discriminating among the set of alternative PSMs (and their associated diagnostic closures). This task is reduced to a set of questions by recursive application of diagnostic strategies: *confirm, differentiate, rule-out, group-and-differentiate* and *explore*. The diagnostic planning provides the program with a focused and efficient diagnostic behavior. In addition it serves as a framework for justifying the motivation for asking a particular question.

We have argued that for a competent medical system to be accepted, it must be able to explain its conclusions to its user. This thesis has applied some recent explanation technology [Swartout80][35] developed in a simpler domain to the much more complex domain of electrolyte and acid-base diagnosis. ABEL is capable of justifying its motivation for asking a particular question and explaining its understanding of the patient's illness at multiple levels of detail.

7.2 Limitations of ABEL and Future Directions

The research presented in this thesis has many limitations. Some are due to limitations of time and resources. More seriously, the inherent size and complexity of the domain has forced us to limit the scope of this research to just a few issues and to adopt engineering compromises.

The representation of the relation between states in ABEL is inadequate; all interactions are described using a single type of link, namely a causal link. This is unnatural when the relationship between disease states is statistical with no known causal explanation. Furthermore, we need to group states which jointly have significant diagnostic and prognostic implications even if the states are not causally or statistically related. Weaker relations, such as "associational links" and "grouping links" are needed to capture these two cases [Pauker76, Patil79].

---

35. The explanation technique developed by Swartout explores the use of *automatic programming* for encoding a *performance program's* domain knowledge and principles which are then used to explain the behavior of the performance program. ABEL, however, maintains an explicit account of its knowledge. Therefore, the use of automatic programming is not necessary to explain ABEL's reasoning or understanding.

Causal interactions are themselves complex and multi-faceted. For example, an effect may be triggered by a cause, or the presence of an effect may require continuous presence of the cause. We consider an elaborate taxonomy of causal relations (e.g., [Rieger77]) to be a necessary component in the future development of ABEL.

One primary objective of this thesis has been to explore structural criteria such as coherence and adequacy in the construction and evaluation of causal hypotheses. We have intentionally avoided probabilistic measures in order to test the full potential of this newly developed structural criteria. However, the structural and probabilistic measures complement each other; both are essential in a diagnostic system. We intend to develop probabilistic measures for evaluating coherent hypotheses based on techniques described in [Duda76] and [Pednault81].

The diagnostic problem solver in ABEL has a simple tree structured plan for controlling its diagnostic information gathering. Although it already provides the rudimentary abilities discussed above, it fails to capture the interactions between different branches of the tree. Additional inadequacies arise for the following two reasons. First, as discussed in chapter 3, the use of available knowledge of anatomy, etiology and disease taxonomy is limited. Second, the program does not ascertain the overall state of the patient's health, e.g., the vital signs, stability etc.[36] This assessment is an important component of the physician's evaluation and has considerable influence on his formulation of diagnostic goals and strategies. We believe that a similar assessment of the overall state of a patient's health should be modeled in the PSM explicitly, and used in guiding the diagnostic exploration. In coming years we envision implementing diagnostic reasoners with increasing sophistication based on the models of causal reasoning developed in this thesis and on recent advances in planning paradigms (e.g., [Sacerdoti75, Stefik81]).

A serious criticism of the work presented here could be the small size of the domain and the availability of a well defined methods for the initial formulation of the diagnostic problem. This leads to the questions; do the techniques presented here scale up? What are the problems if they are applied to medical diagonsis in a larger domain similar to that of PIP or INTERNIST-I?

The exact methods used by ABEL in the initial formulation of diagnostic problems are domain dependent. We believe that use of similar techniques to limit the size of initial problem formulation is common among clinicians [Elstin78, Kassirer78]. We believe that it is important to distinguish between the processing strategies used in the initial formulation and those used during later stages of the diagnostic process. Substantial further research is needed in identifying strategies for initial processing in larger contexts.

---

36. We have often noted clinicians describing a patient in terms such as "this is an otherwise healthy patient with chronic urinary tract infection" or "this is a very sick patient with acute bowel inflammation".

We believe, however, that the multi-level causal representation of medical and patient-specific knowledge, and the description-building processes are independent of the size of the data-base. The major difficulty in using these methods lies in the enormity of the knowledge-base that will be required to adequately cover problems of the size tackled by PIP or the INTERNIST-I.

In summary, this thesis has developed a new representational scheme, capable of capturing some of the subtlety and richness of knowledge employed by expert physicians, and we have presented a new form of diagnostic problem solver which avoids some of the problems present in the previous programs. We believe that the research presented in this thesis is a small step in the right direction. Designing expert medical programs is a difficult and challenging task; much work clearly remains before successful and acceptable expert medical systems are a reality.

The road to wisdom? — Well, it's plain
and simple to express:
Err
and err
and err again
but less
and less
and less.
Piet Hein, *Grooks*

8. References

[Bleich72]
Bleich, H. L., Computer-Based Consultation: Electrolyte and Acid-Base Disorders, *American Journal of Medicine* **53**:285 (1972).

[Bobrow76]
Bobrow, D. G., and Winograd, T., An Overview of KRL, a Knowledge Representation Language, Artificial Intelligence Laboratory, Stanford University, Technical Report AIM-293 (1976).

[Brachman78]
Brachman, R. J., Theoretical Studies in Natural Language Understanding, Bolt, Beranek, and Newman Inc., Technical Report 3888 (1978).

[Buchanan71]
Buchanan, B. G., and Lederberg, J., The Heuristic DENDRAL Program for Explaining Empirical Data, *Proceedings of the IFIP Congress 1971*, Ljubljana, Yugoslavia (1971). Also Artificial Intelligence Laboratory, Stanford University, Technical Report AIM-141 (1971).

[Church80]
Church, K. W., On Memory Limitations in Natural Language Processing, Massachusetts Institute of Technology, Laboratory for Computer Science, Technical Report MIT/LCS/TR-245 (1980).

[Cohen66]
Cohen, J. J., and Schwartz, W. B., Evaluation of Acid-Base Equilibrium in Pulmonary Insufficiency, *The American Journal of Medicine* **41**:2, 163-167 (1966).

[Cohen77]
Cohen, J. J., New Concepts of Acid-Base Balance, in A. B. Schwartz and H. Lyons, ed. *Acid-Base and Electrolyte Balance* Groune & Stratton, New York (1977).

[Davis76]
Davis, R., Applications of Meta Level Knowledge to the Construction, Maintenance and Use of Large Knowledge Bases, Ph.D. thesis, Stanford Artificial Intelligence Laboratory, Technical Memo 283 (1976).

[Davis77]
Davis, R., Buchanan B., and Shortliffe, E., Production Rules as a Representation for a Knowledge-Based Consultation Program, *Artificial Intelligence* **8**:15-45 (1977).

[deKleer79]
deKleer, J., Causal and Teleological Reasoning in Circuit Recognition, Massachusetts Institute of Technology, Artificial Intelligence Laboratory, Technical Report MIT/AI/TR-529 (1979).

[Doyle78]
Doyle, J., A Glimpse of Truth Maintenance, Massachusetts Institute of Technology, Artificial Intelligence Laboratory, Technical Memo MIT/AI/TM-461a (1978).

[Doyle80]
Doyle, J., A Model for Deliberation, Action, and Introspection, Massachusetts Institute of Technology, Artificial Intelligence Laboratory, Technical Report MIT/AI/TR-581 (1980).

[Duda76]
Duda, R. O., Hart, P. E., and Nilsson, N. J., Subjective Bayesean methods for rule-based inference systems, *Proceedings of the 1976 National Computer Conference*, AFIPS Press (1976).

[Elstein78]
Elstein, A. S., Shulman, L. A., and Sprafka S. A., *Medical Problem Solving*, Harvard University Press (1978).

[Feinstein77]
Feinstein, A. R., Clinical Biostatistics XXXIX. The haze of Bayes, the aerial palaces of decision analysis, and the computerized Ouija board, *Clinical Pharmacology and Therapeutics* **21**:4:482-496 (1977).

[Fikes72]
Fikes, R. E., Hart, P. E., Nilsson, N. J., Learning and Executing Generalized Robot Plans, *Artificial Intelligence* **3**:251-288 (1972).

[Forrester61]
Forrester, J. W., *Industrial Dynamics*, MIT Press, Cambridge (1961).

[Gaines76]
Gaines, B. R., Foundations of fuzzy reasoning, *Int. J. Man-Machine Studies* 8:623-668 (1976).

[Goldberg73]
Goldberg, M., Green S., Moss, M. L., Marbach, C. B., and Garfinkel, D., Computer-based Instruction and Diagnosis of Acid-Base Disorders, *JAMA* **223**:3:269-275 (1973).

[Gorry67]
Gorry, G. A., A System for Computer-Aided Diagnosis, Project MAC, Massachusetts Institute of Technology, Technical Report MIT/MAC/TR-44 (1967).

[Gorry68]
Gorry, G. A., and Barnett, G. O., Sequential Diagnosis by Computer, *JAMA* **205**:12:141-146 (1968).

[Gorry73]
Gorry, G. A., Kassirer, J. P., Essig, A., and Schwartz, W. B., Decision Analysis as the Basis for Computer-Aided Management of Acute Renal Failure, *The American Journal of Medicine* **55**:473-484 (1973).

[Gorry78]
Gorry, G. A., Silverman, H., and Pauker, S. G., Capturing Clinical Expertise: A Computer Program that Considers Clinical Responses to Digitalis, *American Journal of Medicine* **64**:452-460, (1978).

[Gorry78a]
Gorry, G. A., Pauker, S. G., and Schwartz, W. B., The Diagnostic Significance of the Normal Finding, *N Engl J Med* **298**:486 (1978).

[Hawkinson75]
Hawkinson, L. B., The Representation of Concepts in OWL, *Proceedings of the Fourth International Joint Conference on Artificial Intelligence*, MIT Artificial Intelligence Laboratory (1975).

[Hawkinson80]
Hawkinson, L. B., XLMS: A Linguistic Memory System, MIT Laboratory for Computer Science, Technical Memo MIT/LCS/TM-173 (1980).

[Kahneman73]
Kahneman, D., and Tversky, A. , On the Psychology of Prediction, *Psychol. Rev.* 80:237-251, (1973).

[Kassirer78]
Kassirer, J. P., and Gorry, G. A., Clinical Problem Solving: A Behavioral Analysis, *Ann. Int. Med.* 89:245, (1978).

[Kuipers78]
Kuipers, B. J., Representing Knowledge of Large-Scale Space, Massachusetts Institute of Technology, Artificial Intelligence Laboratory, Technical Report MIT/AI/TR-418 (1978).

[Kulikowski81]
Kulikowski, C. A., and Weiss, S. M., Representation of Expert Knowledge for Consultation: The Casnet and Expert Projects, in P. Szolovits, ed., *Artificial Intelligence in Medicine*, Westview Press, forthcoming (1981)

[Lynch60]
Lynch, K., *The Image of the City*, MIT Press, Cambridge, Massachusetts (1960).

[Long77]
Long, W. J. A Program Writer, Massachusetts Institute of Technology, Laboratory for Computer Science, Technical Report MIT/LCS/TR-187 (1977)

[Long81]
Long, W. J., and Szolovits, P., The Development of Clinical Expertise in the Computer, in P. Szolovits, ed., *Artificial Intelligence in Medicine*, Westview Press, forthcoming (1981).

[Marcus79]
Marcus, M., A Theory of Syntactic Recognition for Natural Language, Ph.D Thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, (1979).

[Martin77]
Martin, W. A., Descriptions and the Specialization of Concepts, *Proceedings of the Fifth International Joint Conference on Artificial Intelligence, Vol. 2*, M.I.T., Cambridge, Ma., August 1977. Pittsburgh, Pa.: Carnegie-Mellon University, 1977. Also available as Massachusetts Institute of Technology, Laboratory for Computer Science, Technical Memo MIT/LCS/TM-101, Cambridge, Ma., (1977).

[Martin79]
Martin, W. A., Roles, Co-Descriptors and the Formal Representation of Quantified English Expressions, Massachusetts Institute of Technology, Laboratory for Computer Science, Technical Memo, MIT/LCS/TM-139.

[Martin81]
Martin, W. A., Church, K. W., and Patil, R. S., Preliminary Analysis of a Breadth-First Parsing Algorithm, MIT Laboratory for Computer Science, Technical Report, M.I.T./LCS/TR-261, Cambridge, Ma., (1981).

[Miller75]
Miller, P. B., Strategy Selection in Medical Diagnosis, Project MAC, Massachusetts Institute of Technology, Technical Report MIT/LCS/TR-153 (1975).

[Minsky75]
Minsky, M., A Framework for Representing Knowledge, in P. H. Winston, ed., *The Psychology of Computer Vision*, McGraw-Hill, New York (1975).

[Moon74]
Moon D., A., MACLISP Reference Manual, Laboratory for Computer Science, Massachusetts Institute of Technology, (1974).

[Newell72]
Newell, A., and Simon, H. A., *Human Problem Solving*, Prentice Hall (1972).

[Newell73]
Newell, A., Artificial Intelligence and the Concept of Mind, in Schank, R. C., and Colby, K. M, eds., *Computer Models of Thought and Language*, W. H. Freeman and Co. (1973).

[Newell76]
Newell, A., and Simon, H. A., The Logic Theory Machine: A Complex Information Processing System, *IBM Trans. on Information Theory*, **2(3)**:61-79, (1976).

[Patil79]
Patil, R. S., Design of a Program for Expert Diagnosis of Acid Base and Electrolyte Disturbances, MIT Laboratory for Computer Science TM-132 (1979).

[Patil81]
Patil, R. S., Szolovits, P., and Schwartz, W. B., Causal Understanding of Patient Illness in Medical Diagnosis, *Proceedings of the Seventh International Joint Conference on Artificial Intelligence*, available from American Association for Artificial Intelligence, 445 Burgess Drive, Menlo Park, Ca 94025 (1981).

[Pauker75]
Pauker, S. G., and Kassirer, J. P., Therapeutic Decision Making: A Cost-Benefit Analysis, *N Engl J Med* **293**:229-234 (1975).

[Pauker76]
Pauker, S. G., Gorry, G. A., and Kassirer, J. P., and Schwartz, W. B., Toward the Simulation of Clinical Cognition: Taking a Present Illness by Computer, *The American Journal of Medicine* **60**:981-995 (1976).

[Pauker77]
Pauker, S. G., and Szolovits, P., Analyzing and Simulating Taking the History of the Present Illness: Context Formation, in Schneider/Sagvall Hein, eds., *Computational Linguistics in Medicine*, North-Holland (1977).

[Pednault81]
Pednault, E. P. D., Zucker, S. W., and Muresan, L. V., On the Independence Assumption Underlying Subjective Bayesian Updating, *Artificial Intelligence*, **16**:213-222 (1981).

[Perlman74]
Perlman, F., McCue, J. D., and Friedland, G., Urinary Tract Infection (UTI) / Vaginitis Protocol, Introduction, Ambulatory Care Project, Lincoln Laboratory, Massachusetts Institute of Technology, and Beth Israel Hospital, Harvard Medical School (1974).

[Pople75]
Pople, H. E., Jr., Artificial-Intelligence Approaches to Computer-Based Medical Consultation, IEEE Intercon Conference (1975).

[Pople75a]
Pople, H. E., Jr., Myers, J. D., and Miller, R. A., DIALOG: A Model of Diagnostic Logic for Internal Medicine, *Advance Papers of the Fourth International Joint Conference on Artificial Intelligence*, available from the Artificial Intelligence Laboratory, Massachusetts Institute of Technology (1975).

[Pople76]
Pople, H. E., Jr., Presentation of the INTERNIST System, Proceedings of the A.I.M Workshop, Rutgers Research Resource on Computers in Biomedicine, Rutgers University (1976).

[Pople77]
Pople, H. E., Jr., The Formation of Composite Hypotheses in Diagnostic Problem Solving: an Exercise in Synthetic Reasoning, *Proceedings of the Fifth International Joint Conference on Artificial Intelligence*, available from the Department of Computer Science, Carnegie-Mellon University, Pittsburgh, PA 15213 (1977).

[Pople81]
Pople, H. E., Heuristic Methods for Imposing Structure on Ill Structured Problems: The Structuring of Medical Diagnostics, in P. Szolovits, ed., *Artificial Intelligence in Medicine*, Westview Press, Forthcoming (1981).

[Raiffa70]
Raiffa, H., *Decision analysis.* Addison-Wesley Publishing Company, Reading, Massachusetts (1970).

[Richards77]
Richards, B., and Goh, A. E. S., Computer Assistance in the Treatment of Patients with Acid-Base and Electrolyte Disturbances, *MEDINFO 77* IFIP, North-Holland Publishing Company (1977).

[Rieger77]
Rieger, C., and Grinberg, M., The Declarative Representation and Procedural Simulation of Causality in Physical Mechanisms, *Proceedings of the Fifth International Joint Conference on Artificial Intelligence*, available from Department of Computer Science, Carnegie-Mellon University, Pittsburgh, PA 15213 (1977).

[Rubin75]
Rubin, A. D., Hypothesis Formation and Evaluation in Medical Diagnosis, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Technical Report MIT/AI/TR-316 (1975).

[Sacerdoti75]
Sacerdoti, E. D., *A Structure for Plans and Behavior*, North Holland/American Elsevier, New York (1977).

[Safran76]
Safran C. et al., Diagnostic planning and cancer management, Massachusetts Institute of Technology, Laboratory for Computer Science, Technical Report MIT/LCS/TR-169 (1976).

[Schwartz65]
Schwartz, W. B., Brackett, N. C., Jr. and Cohen, J. J., The response of extracellular hydrogen ion concentration to graded degrees of chronic hypercapnia: The physiologic limits of the defense of pH, *J. Clin. Invest.*, **44**:291 (1965).

[Schwartz70]
Schwartz, W. B., Medicine and the Computer: The Promise and Problems of Change, *N Engl J Med*, **283**:1257-1264 (1970).

[Schwartz73]
Schwartz, W. B., Gorry, G. A., Kassirer, J. P, and Essig, A., Decision Analysis and Clinical Judgment, *The American Journal of Medicine*, **55**:459-472 (1973).

[Shafer76]
Shafer, G., *A Mathematical Theory of Evidence*, Princeton University Press (1976).

[Shortliffe75]
Shortliffe, E. H., and Buchanan, B. G., A Model of inexact reasoning in medicine, *Mathematical Biosciences* **23**:351-379 (1975).

[Shortliffe76]
Shortliffe, E. H., *Computer Based Medical Consultations: MYCIN*, Elsevier North Holland Inc. (1976).

[Shortliffe79]
Shortliffe, E. H., Knowledge Engineering for Medical Decision Making: A Review of Computer-Based Clinical Decision Aids, *Proceeding of the IEEE*, forthcoming (1979).

[Silverman75]
Silverman, H., A Digitalis Therapy Advisor, Project MAC, Massachusetts Institute of Technology, Technical Report MIT/MAC/TR-143 (1975).

[Smith79]
Smith, B. C., A Proposal for a Computational Model of Anatomical and Physiological Reasoning, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Technical Memo MIT/AI/TM-493 (1978).

[Stedmans]
Stedman's Medical Dictionary, 20th ed., Baltimore: Williams & Wilkins (1961).

[Stefik81]
Stefik, M., Planning with Constraints (MOLGIN: Part 1), *Artificial Intelligence*, **16**:111-140 (1981).

[Stefik81a]
Stefik, M., Planning and Meta-Planning (MOLGIN: Part 2), *Artificial Intelligence*, **16**:141-170 (1981).

[Sussman72]
Sussman, G. J., and McDermott, D. V., From PLANNER to CONNIVER - A Genetic Approach, *Proceedings of the 1976 Fall Joint Computer Conference*, pp. 1171-1179, AFIPS Press (1976).

[Sussman73]
Sussman, G. J., Some Aspects of Medical Diagnosis, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Working Paper 56 (1973).

[Swartout81]
Swartout, W., R., Producing Explanations and Justifications of Expert Consulting Programs, Massachusetts Institute of Technology, Laboratory for Computer Science, Technical Report MIT/LCS/TR-251 (1981).

[Szolovits76]
Szolovits, P., and Pauker, S. G., Research on a Medical Consultation System for Taking the Present Illness, *Proceedings of the Third Illinois Conference on Medical Information Systems*, University of Illinois at Chicago Circle (1976).

[Szolovits77]
Szolovits, P., Hawkinson, L., and W. A. Martin, An Overview of the OWL Language for Knowledge Representation, *Proceedings of the Workshop on Natural Language Interaction with Databases*, International Institute for Applied Systems Analysis, Schloss Laxenburg, Austria, Jan. 10, 1977. Also appeared as MIT/LCS/TM-86 (1977).

[Szolovits78]
Szolovits, P., Pauker, S. G., Categorical and Probabilistic Reasoning in Medical Diagnosis, *Artificial Intelligence* **11**:115-144 (1978).

[Szolovits78a]
Szolovits, P., The Lure of Numbers--How to Live With and Without Them in Medical Diagnosis, *Proceeding of the Colloquium in Computer-Assisted Decision Making using Clinical and Paraclinical (Laboratory) Data*, (1978).

[Szolovits79]
Szolovits, P., and Pauker, S. G., Computers and Clinical Decision Making: Whether, How, and For Whom? *Proceeding of the IEEE*, **67**:1224-1226 (1979).

[Szolovits81]
Szolovits, P., *Artificial Intelligence in Medicine*, Westview Press, forthcoming (1981).

[Szolovits81a]
Szolovits, P., Artificial Intelligence and Medicine, in P. Szolovits, ed., *Artificial Intelligence in Medicine*, Westview Press, forthcoming (1981).

[Szolovits81b]
Szolovits, P., Patil, R. S., and Schwartz, W. B., Reasoning Methods in AI programs in Medicine, in progress.

[Tversky71]
Tversky, A., and Kahneman, D., Belief in the Law of Small Numbers, *Psychol. Bull.* **76**:105-110 (1971).

[Tversky74]
Tversky, A., and Kahneman, D., Judgment under Uncertainty: Heuristics and Biases, *Science* **185**:1124-1131 (1974).

[Weed68]
Weed, L. L., Medical Records that Guide and Teach, *New Engl. J. Med.* **278**:593, (1968).

[Weiss74]
Weiss, S. M., A System for Model-Based Computer-Aided Diagnosis and Therapy, Ph.D. Thesis, Computers in Biomedicine, Department of Computer Science, Rutgers University, CBM-TR-27-Thesis (1974).

[Weiss76]
Weiss, S. M., Presentation at the Third Illinois Conference on Medical Information Systems, University of Illinois at Chicago Circle (1976).

[Weiss78]
Weiss, S. M., Kulikowski, C. A., Amarel, S., and Safir, A., A Model-Based Method for Computer-Aided Medical Decision-Making, *Artificial Intelligence* **11**:145-172 (1978).

[Winston77]
Winston, P. W., *Artificial Intelligence*, Addison-Wesley Publishing Company, Reading, Massachusetts, (1972).

[Yu79]
Yu, V. L., et. all, Evaluating the Performance of a Computer-Based Consultant, *Computer Programs in Biomedicine* 9:95-102 (1979)

[Zadeh65]
Zadeh, L. A., Fuzzy sets, *Information and Control* 8:338-353 (1965).

Appendix I - System Building Tool: XLMS

The ABEL system uses XLMS to represent and manage its knowledge base. The XLMS (eXperimental Llinguistic Memory System) was developed primarily by Lowell Hawkinson, William Martin, Peter Szolovits and the members of the Knowledge Based Systems and the Clinical Decision Making groups at MIT [Hawkinson80]. Although the representation of the ABEL system has been substantially influenced by the design philosophy and the details of the implementation of XLMS system, it is not necessary to have a complete understanding of the intricacies of XLMS to understand this document. This section is intended to elucidate only as much of XLMS as is required to comprehend this document. Furthermore, wherever possible, the XLMS notation is supplemented by its graphical representation to reduce the dependence of this document on the notation of the XLMS.

Perhaps the best way to think of XLMS is that it is an extension of LISP that allows one to use *unique* and *canonical* expressions and allows one to *label* these expressions uniquely. In LISP, atoms are used to name variables and functions. In XLMS, variables and procedures can be named by unique expressions (called concepts). Similar to LISP atoms, these concepts can have properties called *attachments*. They differ from the LISP in the sense that these concepts are structured objects and can have *superior* and *inferior* structures. In addition, these concepts have internal structure that can be taken apart and examined, while lisp atoms are indivisible.

I.1 XLMS Concepts

In XLMS, every concept is composed of three parts: *ilk, tie* and *cue*, and is written as:

```
[(<ilk>*<tie> <cue>)]
```

or, to take an actual example from the ABEL data base:

```
[(concentration*u ecf-na)]
```

The *ilk* of a concept is itself a concept. It describes the concept this concept is derived from. Thus the example concept described above, is a kind of "concentration". The *cue* of a concept is either a concept or a lisp symbol. It specializes the general concept described by the ilk, or in other words, indicates what it is that makes this concept different from others with the same ilk. The example represents the "concentration of ecf-na": a particular kind of concentration. The *tie* of a concept indicates the relationship between the ilk and the cue. In this case, the tie is "u" for *unique-role*. The *role* ties are used to indicate slots (attributes or properties) of a concept (furthermore, a unique-role indicates that there is only one role of the kind described by the concept). Thus this concept represents the "concentration" slot in the concept of "ecf-na". There are several other ties that are used in the system, some of them primitive to the XLMS system and some "user defined" for use in the ABEL system. These are listed in Table 1 together with examples of their use. Finally, any concept in the data-base can be (optionally) labeled using the notation

Table I.

Tie	Name	Example Use	English Form	Purpose
*f	function	[(ball*f red)]	(the) red ball	functional restriction
*r	role-in	[(color*r ball)]	(the) color of (the) ball	slot filling
*u	unique-role	[(weight*u ball)]	(the) weight of (the ball)	slot filling
*i	individual	[(ball*i 1)]	ball	instance
*s	species	[(bird*s robin)]	robin	mutually exclusive decomposition

---

```
[<label> = <concept>]
```

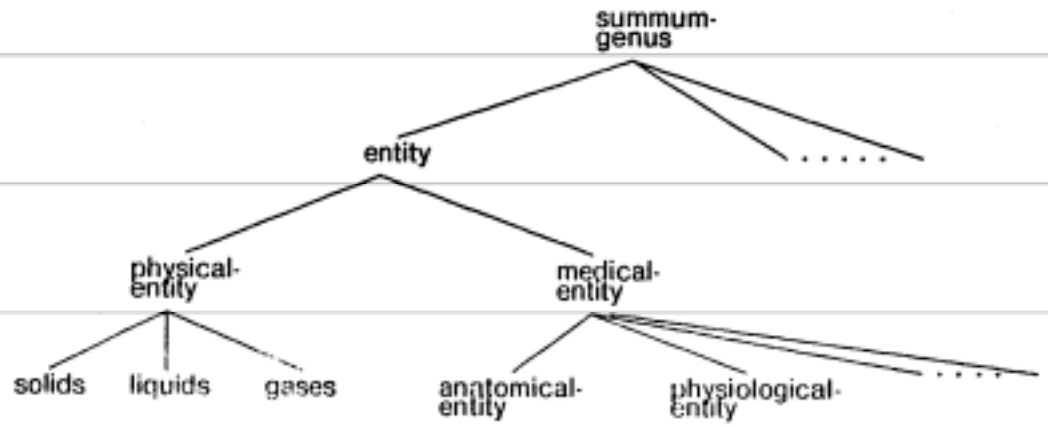
or, to name the concept defined above

```
[serum-na = (concentration*u ecf-na)]
```

As was indicated above, concepts in XLMS are organized into an *AKO* hierarchy (see figure 57). The root concept is [summum-genus] and all concepts are defined as specialization of this concept.

In LISP, a symbol may have a property-list, which can be used to to attach properties (lists and other atomic symbols) to a symbol. Similarly, in XLMS, we have *attachment* which can be used to associate concepts relating to a concept with that concept. The notation for attachment is:

Fig. 57. The XLMS hierarchy



```
[<concept> #<attachment-relation>
          <attached-concept>  .... <attached-concept>]
```

or, for example:

```
[(sex*u patient) #Value male]
```

The attachment-relation specifies how the concept and the attached concept (called the *attachment*) are related. The example above states that the slot "sex of the patient" has the "value" of "male". An alternate way would have been to create a specific concept to describe the same relation. For example, the relation described above could be alternately specified as

```
[((sex*u patient)*f male)]
```

which states that the "sex of the patient" is functionally restricted to being "male". The built in functions of XLMS tend to make it easier to work with the concept hierarchy than with attachments. Typically, primary characterizations of a concept are placed in the kind hierarchy while secondary ones are indicated by attachments [Martin79]. The commonly used attachments in the ABEL system are: #v (value), #f (function) #c (characterization), #m (meta-characterization) and #s (standard-error). Some additional attachment relations such as # meta-link are also used.

Program fragments in the ABEL system are described using sequences of XLMS concepts. Sequences are described in XLMS notation by a list of concepts separated by commas:

```
[<concept>,<concept>, .... <concept>]
```

The reader may have noticed that the XLMS notation is delimited by square brackets. These brackets identify the concept as a piece of XLMS notation and delimit the scope of its attachments (if any). Any expression delimited by square brackets is called a *complex*. The first concept to appear after a left bracket is called the head of complex. If an xlms-complex is contained within some piece of XLMS notation, the XLMS reader makes any attachments or builds any structure indicated by the complex, and then replaces the complex by the head of the complex.

Finally, the colon anaphora provide a convenient shorthand for specifying the slots (roles) of a concept. If a concept appears in XLMS notation with a colon (or several colons) immediately following it, then the XLMS reader replaces that concept with a new concept whose ilk is the concept with that notation and whose tie is *r. If the colons are immediately followed by a number (e.g., 1), then the XLMS reader replaces this concept with the first instance of the new concept. If colons are immediately followed by a "u", the XLMS reader replaces this concept with a new concept whose ilk is the concept with the concept with that notation, whose tie is *u and whose cue is the head of the conplex n levels in from the top level complex, where n is the number of colons in the notation. For example:

```
[table                ==>  [table
  [top:u #c ...]]              [(top*u table) #c ...]]

[table                ==>  [table
  [leg: #c ...]]               [(leg*r table) #c ...]]

[table                ==>  [table
  [leg:1 #c ...]]              [((leg*r table)*i 1) #c ...]]
```

A similar anaphora mechanism, but counting from the inside, is provided with the use of uparrow ("↑").

I.2 The XLMS Interpreter

A simple XLMS interpreter LINT (Little INTerpreter) was implemented by the author to execute the mapping relations associated with links and component summation/decomposition relations associated with primitive links. The evaluation of functions and handling of arguments in the interpreter is similar to LISP. For example, a function "`(compute-ph serum-hco3 serum-pco2)`" in LISP is equivalent to "[(compute-ph*c serum-hco3,serum-pco2)]" in LINT. They differ in the way the variables are evaluated. In LINT a variable (indicated by a role tie) is evaluated by first binding the concept containing the role (slot associated with the variable) to its instantiation in the initiating pattern (i.e., a specific link or constituent summation) or the selected PSM, and then accessing the value associated with the slot in the instance, or by inheriting the appropriate default value associated with the slot. For example, evaluation of the above LINT function in context of PSM-1 of example 1 in chapter 6 would result in binding "serum-hco3" to "serum-hco3-1" with the value of 15.0 and similarly "serum-pco2" to "serum-pco2-1" with the value of 30.0.

Finally, we note that the program fragments in xlms are expressed as concepts, they are naturally organized into XLMS hierarchy of concepts allowing the program to inherit the function definition for specific tasks from more general definitions. For example, the function to compute the concentration of serum-Na from the total quantity of Na in the extracellular compartment can be computed using the more general function for computing the concentration of a serum electrolyte from its total store in the extracellular compartment.

Appendix II - Explanation

The English explanation generator in the ABEL system is implemented using the methodology developed by Swartout [Swartout80] as a part of an interactive system which explains and justifies portions of an expert system for prescription of Digitalis. In this chapter we will review the methodology for generating English from causal paths and XLMS concepts developed by Swartout and discuss techniques for organizing the flow of explanation in translating descriptions contained in a PSM.

II.1 Phrase Generator

The phrase generator generates English phrases from XLMS concepts. An example of an XLMS concept and the phrase generated for it is:

```
[(severity*u diarrhea)] ==>  (the) severity of diarrhea
```

In XLMS, the tie of a concept indicates the relationship between the ilk and the cue of the concept. Thus, *r indicates that the ilk is a role in the cue, and *f indicates that the cue is a modifier of the ilk. Because the tie indicates the relationship between the ilk and cue, it also determines the primary English form of a concept. Therefore, the phrase generator is organized around the types of the tie. Examples of primary English phrases associated with concepts with different ties is shown in table 1 in appendix I.

The phrase generator contains a set of translation rules; one rule for each type of tie. However, for a labeled concept, the phrase generator prefers the label of the concept over its translation except when the use of the label is explicitly forbidden. This can be done by meta-characterizing the labeled-concept or any of its superiors with [do-not-use-label].

The translation of a causal link into English is initiated when a concept with the tie *e is encountered. However, this translation is also dependent upon whether the link is being traversed forward from source to destination or vice versa. For example,

```
[((caused-by*b diarrhea)*e metabolic-acidosis)]
```

would be translated while being traversed forward as

```
diarrhea causes metabolic acidosis
```

and, while being traversed backwards as

```
metabolic acidosis is caused by diarrhea
```

These low level primitives for translating individual XLMS expressions are used by the higher level of explanation generator which traverses the causal net.

II.2 Higher level explanations

With the ability to translate a state and a link we next focus on describing causal relations occurring in a causal net. First, let us focus on describing simple chains of causal links, such as:

```
[((caused-by*b A)*e B)], [((caused-by*b B)*e C)], [((caused-by*b C)*e D)]
```

which is translated into

```
A causes B, which causes C.  C causes D.
```

Note that this is somewhat of a compromise. It prevents the monotony of having three sentences with identical structures; "A causes B. B causes C. C causes D.". However, the number of components in any given sentence are restricted to two, therefore, in situations where use of three causal relations in a single sentence is desired, the explanation generated by this method comes out rough.

Let us consider another situation

```
[((caused-by*b A)*e B)]
[((caused-by*b C)*e B)]
```

If this situation occurs in the general medical knowledge, then it implies that A and C are two possible causes of B.[37] This is translated as

```
A may cause B.  C may also cause B.
```

However, if this situation occurs in the PSM, then A and C combine to cause B. This can be stated as

```
A and C jointly cause B.
```

However, if the we are discussing the relation between A and B then this is stated with the help of an adjunct clause, e.g.,

```
A along with C causes B.
```

Conversely, when we have

```
[(caused-by*b A)*e B)]
[(caused-by*b A)*e C)]
```

It is translated as

```
A causes B and C.
```

A high level English generator is based on this translation facility. Its primary goal is to organize the medical knowledge or the patient specific knowledge into a sequence of objects which are then translated using the translation facility.

---

37. In the medical knowledge base a causal link is interpreted as indicating a possible causal relation.

II.3 Organizing causal Explanation

The operation of the higher level explanation generator can be described in three steps: (1) describing a node (state), (2) describing a relation between two nodes, and (3) describing a causal network.

The description for a node is generated in the following way. The translation of the concept associated with the node generates a noun phrase (NP). Each attribute (slot) of this node can be then be described as an adjective which modifies the noun phrase associated with the node. For example;

```
[(diarrhea*i 1)
      [severity:u #v 0.7 #c severe]
      [duration:u #v 60.o #c chronic]]

==> severe chronic diarrhea
```

In addition the explanation generator distinguishes between the first time a concept is described from all subsequent references to the concept. At the first time every attribute of the concept that has been specified in the instance is described whereas on subsequent references only those attributes are mentioned which are necessary to distinguish this instance from other references to other instances of the same concept. For example, if during the discussion we had also made reference to "severe acute diarrhea" then in a later reference to the [(diarrhea*i 1)] the program will distinguish this from the others by specifying "chronic diarrhea".

A description of the relation between given two nodes in a causal network is generated as follows. (1) Identify all loop free paths from the first node to the second. This generates a partial order graph (acyclic graph). (2) Impose a linear order on the partial order graph of step 1. (3) Translate this linear order of concepts using the translator described above. However, we must note that there may be a causal path between the two nodes in each direction — that is, the two nodes might be part of a feedback cycle. This is handled by repeating step 1 with each of the two nodes as the starting nodes. If both of the partial order graphs are null, we know that the two nodes are unrelated, if one of them is null and the other non-null, then the relation between them is one directional, and if both of them are non-null, then we know that there is a feedback relation between the two nodes. Luckily, the above algorithm has already decomposed the feedback relation between the forward and the feedback components. Thus we can divide the explanation of the relation into two parts: from first node to second and from second to first. An example of the relation between acidemia and hypocapnia for hypothesis 1 of example 1 in chapter 6 is shown in figure 58.

The English description for a given causal network is organized in three parts: (1) a one line introduction describing the primary causes and the important electrolyte and acid-base states in the causal network being explained, (2) a detailed description of the causal network being

Fig. 58. Feedback relation betweeen acidemia and hypocapnia

The forward path: `The acidemia causes hypocapnia.`

The feedback path: `The hypocapnia along with metabolic-acidosis causes hypobicarbonatemia. Hypocapnia and hypobicarbonatemia jointly cause acidemia.`

---

explained, (3) a one or two line summary of the causal network which focuses on the nodes in the causal network that cannot be adequately explained by the network. As the first step in organizing the explanation the program divides the nodes in the given causal network into one of the following groups: (1) ultimate etiologies, (2) acid-base nodes (3) fully unaccounted nodes (4) partially unaccounted nodes, and (5) other nodes.

The one line introduction to the causal network is generated by describing the age, sex and weight of the patient and all the ultimate etiologies and the acid-base nodes that have not been accounted for by any of the ultimate etiologies.

The generation of the detailed description of the causal network closely follows the procedure used for describing the relation between two states. The program takes each of the ultimate etiologies and fully unaccounted nodes and identifies all loop free paths from these nodes to all other nodes reachable from them. As discussed before, these paths impose a partial order on the causal network. The program then imposes a linear order on the partial order graph generated in the previous step. Finally it translates this linear order sequence of nodes and links into English as discussed before.

Finally, the program summarizes the description of the causal network by listing all the fully and partially unaccounted nodes. That is, it summarizes the inadequacies in the causal explanation and points out the nodes that need further diagnostic exploration.

In this appendix we have briefly discussed the techniques used by ABEL in organizing explanations of ABEL's medical and patient specific knowledge and have reviewed the English translation methodology developed by Swartout [Swartout80]. The primitive explanation capabilities of ABEL, in spite of their inadequacies, have already proved to be valuable to the developers of the program. Substantial further developments in improving the quality of the English generated, identifying the level at which the explanation should be provided, and in tailoring the explanations to users' needs using modeling of user's understanding of the program and the domain of medical expertise are needed.