14

Causal Understanding of Patient Illness in Medical Diagnosis

Ramesh S. Patil, Peter Szolovits, and William B. Schwartz

In most medical AI programs, the use of notions such as causal relationships, temporal patterns, and aggregate disease categories has been limited. Yet studies of clinicians' behavior reveal that a diagnostic or therapeutic program must consider a case at various levels of detail to integrate overall understanding with detailed knowledge.

To explore these issues, Ramesh Patil, Peter Szolovits, and William Schwartz have applied the knowledge-based approach in a detailed study of consultation for electrolyte and acid-base disturbances. The resulting program, Patil's dissertation work, is known as ABEL (for Acid-Base and ELectrolyte program). ABEL and an earlier M.I.T./Tufts program known as the Digitalis Therapy Advisor (Gorry et al., 1978) were important departures from other systems in that they both viewed clinical problem solving as a process of constructing *an explanation of manifestations, what they have called a* patient-specific model. *In ABEL, this description includes data about the patient as well as the program's hypothetical interpretations of these data in a multilevel causal network. Proceeding from the lowest level, the concepts and relations gradually shift in content from pathophysiological to syndromic knowledge. The aggregate level of this description summarizes the patient data, providing a global perspective for efficient exploration of the diagnostic possibilities. The pathophysiological description provides the ability to handle complex clinical situations arising in illnesses*

From the *Proceedings of the Seventh International Joint Conference on Artificial Intelligence,* vol. 2, 1981, pp. 893–899. Used by permission of International Joint Conferences on Artificial Intelligence, Inc.; copies of the *Proceedings* are available from William Kaufmann, Inc., 95 First Street, Los Altos, CA 94022.

with multiple etiologies, to evaluate the physiological validity of diagnostic possibilities being explored, and to organize large amounts of seemingly unrelated facts into coherent causal descriptions.

The approach can be considered to be an outgrowth of PIP (Chapters 6 and 9), but using a complete causal model of disease. While CASNET (Chapter 7) simply propagates weights, ABEL symbolically manipulates—through operations such as aggregation and elaboration—causal concepts on multiple levels of detail. It is hierarchical, the kind of organization promoted in MDX (Chapter 13), but involves a principled abstraction of causes with complex links that can themselves be reasoned about; they are not just pointers connecting diseases.

The ABEL research is evolving into a study of reasoning strategies for using the principled representation of medical knowledge (Patil and Szolovits, 1982). This is clearly the state of the art in medical knowledge representation, with strong implications for producing robust consultation programs. The empirical psychological methodology—studying expert problem solving in detail to derive better representations—has been strongly promoted by the group at M.I.T. and Tufts and is an idea we see in much of the research reported in this volume (Chapters 10, 12, 13, 15, and 16).

14.1 Introduction

We have studied difficulties arising in the operations of the "first generation" of AI programs in medicine and have undertaken the development of knowledge representation structures to support needed improvements. The description of a patient in existing programs such as INTERNIST-I (Pople et al., 1975), PIP (see Chapter 6), and MYCIN (Shortliffe, 1976) starts from a single list of findings about the patient. Using a data base of associations between diseases and findings (or rules establishing those connections), these programs form an interpretation of the patient's condition that is essentially a list of possible diseases, ranked by a calculated estimate of likelihood or degree of belief in each.

Researchers (Patil, 1979; Pople, 1977; Smith, 1978) have recognized the need to use notions such as causal relationships, temporal patterns, and aggregate disease categories in the description of a program's diagnostic understanding, but the mechanisms provided to do this have been too weak. For example, although *causality* appears as a term in descriptions in PIP and INTERNIST-I, in both cases its use is limited to guiding the propagation of likelihood measures. These programs fail to capture the human notion that explanation should rest on a chain of cause-effect deduction. Although the CASNET/Glaucoma (see Chapter 7) program uses a network of causally related states and defines diseases as paths in this

network, its primary reasoning mechanism is nevertheless the local propagation of probability weights.

Similarly, it has been realized that a diagnostic or therapeutic program must consider a case at various levels of detail in order to integrate its overall understanding with its detailed knowledge. This insight also has not prevailed in the actual mechanisms provided in existing programs.

To explore the issues outlined here, we have undertaken a study of the medical problem of providing expert consultation in cases of electrolyte and acid-base disturbances. We have partly completed implementation of a program, ABEL, that is the diagnostic component of our overall effort. In this paper we concentrate on ABEL's mechanism for describing a patient. Called the *patient-specific model* (PSM) (Gorry et al., 1978), this description includes data about the patient as well as the program's hypothetical interpretations of these data in causal hierarchical networks. We describe the representations of medical knowledge and the processing strategies needed to enable ABEL to construct a PSM from the initial data presented to the program about a patient. The same representations and procedures will also be useful to revise the PSM during the process of diagnosis, but we will concentrate here on the logically prior operation of building the PSM.

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 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 interpretation of abnormal findings. For our program to reason at a sophisticated level of competence, it will need to share such a range of representations. The PSM is, therefore, a multilevel causal model, each level of which attempts to give a coherent account of the patient's case. This model also serves as the basis for an English-generation facility that provides explanations of the program's understanding.

The PSM is created by instantiating portions of ABEL's general medical knowledge and filling in 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. *Aggregation* and *elaboration* make connections across the levels of detail in the PSM by filling in the structure above and below, respectively, a selected part of the network. 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 that 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 etiologic 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.

14.2 Hierarchical Representation of Medical Knowledge

Based on our observation that a physician's knowledge is expressed at various levels of detail, we have developed a hierarchical multilevel representation scheme to describe medical knowledge and procedures to instantiate this knowledge to describe a particular patient's illness. The lowest level of description consists of pathophysiological knowledge about diseases, which is successively aggregated 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 diagnostic possibilities. The physiological knowledge provides us with the capability of handling complex clinical situations arising in patients with multiple disturbances, evaluating the physiological validity of the diagnostic possibilities being explored, organizing a large number of seemingly unrelated facts, and formulating therapy recommendations and prognosis. Finally, since the causal-physiological reasoning tends to be categorical and the syndromic reasoning probabilistic, the hierarchical description allows us to blend together the use of categorical and probabilistic reasoning (see Chapter 9).

14.2.1 Multilevel Description of States

Medical knowledge about different diseases and their pathophysiology is understood in varying degrees of detail. While it may be easier for a program to reason succinctly with medical knowledge artificially represented at a uniform level of detail, we must be able to reason with medical knowledge at different levels of detail to exploit all the medical information available. Although this does not pose any difficulty in medical domains where the pathophysiology of diseases is not well developed, in a domain such as electrolyte and acid-base disturbances where, 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, we are constantly faced with this problem.

Second, the information about a patient parallels the physician's medical knowledge about diseases and therefore also comes at different levels of detail. For example, "serum creatinine concentration of 1.5" is at a distinctly different level than "high serum creatinine,"[1] and "lower gastrointestinal loss" is at a different level than "diarrhea." We need some mechanism by which we can interrelate these concepts. Finally, in order to be effective in diagnostic problem solving and communicating with clinicians, we ought to have the ability to portray the diagnostic problem in a small and compact space. Yet to be efficacious, we must maintain the ability to take every possible detail into consideration. We have solved this problem by representing the medical knowledge in five distinct levels of detail from a deep pathophysiological level to a more aggregate level of clinical knowledge about disease associations.

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 physiological state and each link represents some relation (causal, associational, etc.) between different states. A state (interchangeably used with node) in the system, such as "diarrhea," is represented as a node in the causal network. Each node is associated with a set of attributes describing its temporal characteristics, severity or value, and other relevant attributes. A state is called a *primitive node* if it does not contain internal structure and is called a *composite node* 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 in this causal network is designated as the *focus node,* and the causal network is called the *elaboration structure* of the composite node. The focus node identifies the essential part of the causal structure of the node above it. Indeed, the collection of focus 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; 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. Nodes that do not play a role as the focal definition of any node at a higher level are called *nonaggregable nodes.* They represent a detailed aspect of the causal model that 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 diarrhea, which is the excessive loss of lower gastrointestinal fluid (lower GI loss). The composition of the lower gastrointestinal fluid and plasma fluid are as follows:

---

[1]For a muscular patient whose previously known value of creatinine is 1.3 we can assume this to be normal, but for a patient with a previously known value of 1.0 this is definitely high and could imply a loss of about one-third of the kidney function.

	Lower GI fluid	*Plasma fluid*
Na	100–110	138–145 mEq/L
K	30–40	4–5 mEq/L
Cl	60–90	100–110 mEq/L
HCO_3	30–60	24–28 mEq/L

In comparison with plasma fluid, the lower GI fluid is rich in bicarbonate (HCO_3) and potassium (K) and is deficient in sodium (Na) and chloride (Cl). This information is represented in the knowledge base by decomposing lower GI loss into its constituents (and associating appropriate quantitative information with the decomposition). The loss of lower GI fluid would result in the loss of corresponding quantities of its constituents (in proportion to the total quantity of fluid loss) as shown in Figure 14-1.

Therefore, an excessive loss of lower GI fluid without proper replacement of fluid and electrolytes would result in a net reduction in the total quantity of fluid in extracellular compartments (hypovolemia). Because the concentration of K and HCO_3 in lower GI fluid is greater than it is in plasma fluid, there is a corresponding reduction in the concentration of K (hypokalemia) and HCO_3 (hypobicarbonatemia) in the extracellular fluid. Finally, as the concentration of Cl and Na in the lower GI fluid is lower than that in plasma fluid, there is an increase in the concentration of Cl (hyperchloremia) and Na (hypernatremia) in the extracellular fluid. This is represented at the next level of description as shown in Figure 14-2.

Figure 14-3 shows the aggregation of this information along with some additional causes and consequences of lower GI fluid loss at the next more aggregate level of detail.

The lower GI fluid loss at this level is a nonaggregable state and therefore does not have an aggregation at the next level above. Figure 14-4 shows the description of the aggregate effects of diarrhea (one of the causes of lower GI loss).

The summarization of the description of lower GI fluid loss and diarrhea shown in Figure 14-4 is achieved through the use of link aggregation and elaboration, described in the next subsection.

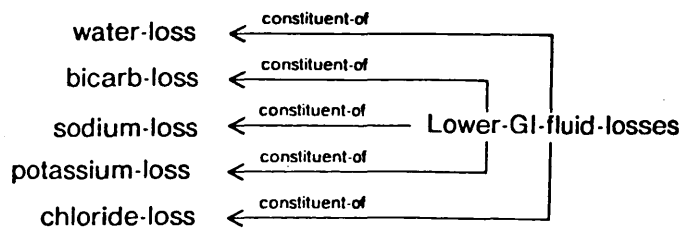

FIGURE 14-1 Effects of lower GI fluid losses on lower GI fluid constituents.

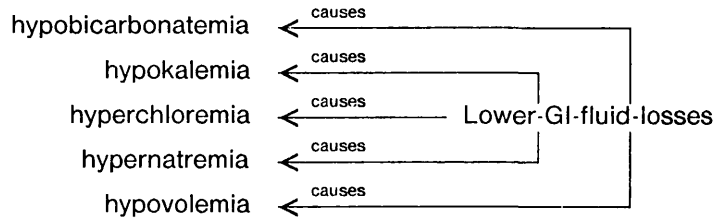

FIGURE 14-2 Effects of lower GI fluid losses at the next level of description.

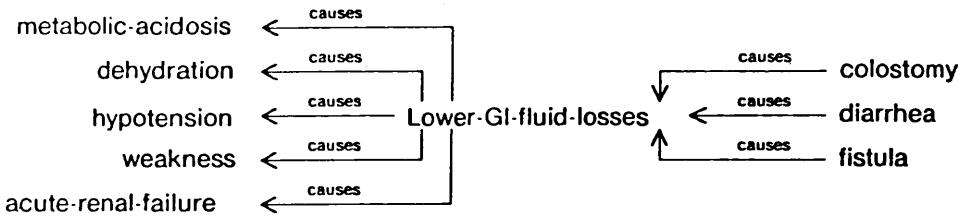

FIGURE 14-3 Aggregation of information in Figure 14-2 with some additional causes and consequences of lower GI fluid loss.

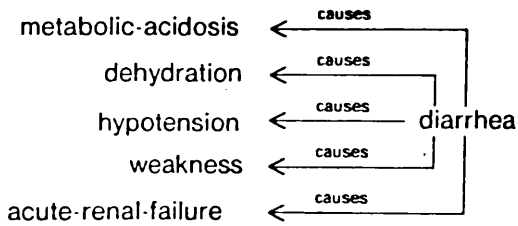

FIGURE 14-4 Summarization of the description of lower GI fluid loss and diarrhea.

14.2.2 Multilevel Description of Causal Links

A causal link specifies the cause-effect relation between the cause (the antecedent) and the effect (the consequent) states. In past programs (e.g., PIP, INTERNIST), causal links were described by specifying the type of causality (may-be-caused-by, complication-of, etc.) and a number or a set of numbers representing in some form the likelihood (conditional proba-

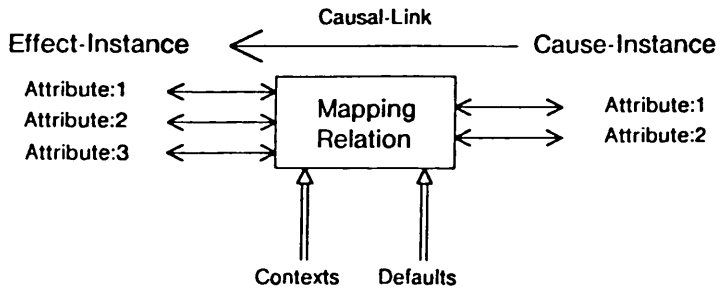

FIGURE 14-5 A causal link in the system.

bility), importance, etc., of observing the effect given the cause or *vice versa.* We now believe that this simple representation of the relation between states is inadequate. The form of presentation of an effect and the conditional probability of observing it depend on various aspects of the cause, such as severity, duration, etc., as well as other factors in the context in which the link is invoked[2] (such as the patient's age, sex, and weight, and the current hypothesis about the patient). Therefore, a causal link in the system (an object denoting the causal relation between a cause-effect pair) specifies a multivariate relation between various aspects of the cause and effect and also specifies the context and assumptions that constrain the causal relation, as shown in Figure 14-5.

One 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 link. This is the reason why links are objects in their own right rather than simple pointers between nodes.

To reason with a causal network representation effectively, a program must make conclusions about a node or link depending only on information that is locally available from the neighborhood of the mechanism in question. If nonlocal effects are to be invoked in causal explanations, they must be explicitly identified (e.g., as part of the context of the causal link), or else they corrupt our ability to reason with any portion of the network. If at some level of detail two distant phenomena interact, we must aggregate the description of the causal network to a level where the two phenomena are adjacent to one another. Further, because the causal relations

[2]For example, a severe diarrhea causes severe hypokalemia, and a mild diarrhea causes mild hypokalemia.

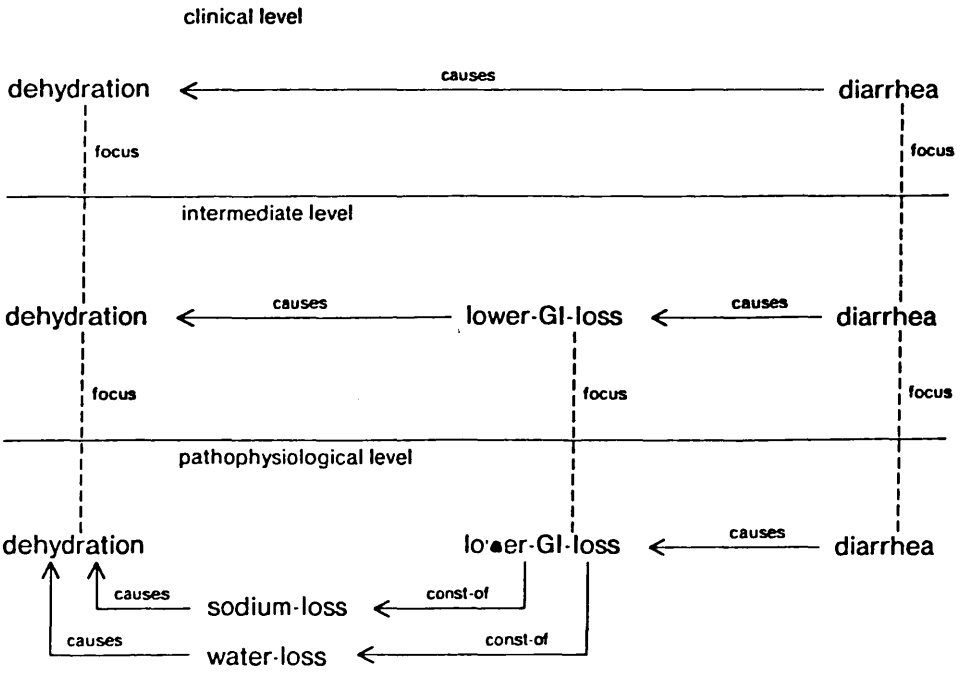

FIGURE 14-6 Causal relation between diarrhea and dehydration.

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 aggregate the diagnostic space to a level where each link represents an aggregate causal phenomenon covering larger distances and thus minimizing the possibility of error in the deduction. This ability to move from one level of description to another is provided by the multilevel description proposed here.

Links can be categorized, as nodes are, into two types: the *primitive links* and the *composite links*. To illustrate the concept of elaborating causal links to form a causal pathway, let us consider the causal relation between diarrhea and dehydration shown in Figure 14-6. The causal mechanism of diarrheal dehydration can be elaborated as follows: diarrhea causes lower GI fluid loss, which causes dehydration. Expressed at the next level of detail, the lower GI fluid loss can be described as consisting of the loss of water and sodium along with other electrolytes. The water loss in the

presence of the reduced total quantity of extracellular sodium results in lower extracellular volume, which at the higher level of description is described as dehydration.

14.3 Reasoning About Components

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. This problem does not place serious limitations on programs dealing with single problems such as the therapy of glaucoma or the diagnosis of bacteremia. But, in the case of electrolyte and acid-base disturbances, where a large fraction of cases involve multiple diagnoses, the ability to evaluate the joint influence of multiple diseases and the ability to decompose their influences on observable findings is particularly important.

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. Without the ability to decompose the hypokalemia, we would have to attribute it completely to the diarrhea or completely to something else. In either case[3] we fail because the total state of hypokalemia is inconsistent with any of its possible single causes. Thus any single cause hypothesized by the program (e.g., vomiting) will not be severe enough to account for the observed hypokalemia by itself. As argued above, we need the ability to hypothesize that only a part of the hypokalemia is accounted for by diarrhea. We introduce the notion that any primitive node in the causal hierarchy[4] may have *components,* which are other primitive nodes that together make up the given node.

In our system this is achieved by a pair of operators: *component summation* and its dual, *component decomposition.* Using our example, these operators allow us to attribute only a part of hypokalemia to the diarrhea and to compute that part of hypokalemia that is not caused by diarrhea (called the *unaccounted component* of the hypokalemia). 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

[3]All of the previous programs would allow the entire hypokalemia to be accounted for by diarrhea. In particular, PIP, after allowing the hypokalemia to be accounted for by diarrhea, will not allow hypokalemia to lend any support to the hypothesis of vomiting. INTERNIST-1, 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.

[4]Recall that *primitive* means that it is not the aggregation of a further defined causal structure.

primitive node a multivariate relation that constrains attributes of the node and its constituents. Component summation combines attributes of the components to generate the attributes of the joint node; component decomposition identifies unaccounted components 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.

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. 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 components by component links.

Because components are defined only for primitive nodes, the instantiation of composite nodes that involve component summation must be in terms of the summation of components in the node's elaboration structure. If the node is composite, we elaborate the constituent nodes around their focal nodes until we reach the primitive nodes associated with them at a level of greater detail. Then we combine these primitive nodes and aggregate their effects back. For example, if we know that a patient has two disturbances, diarrhea and shock, causing metabolic acidosis (Figure 14-7)[5], we evaluate their contribution to metabolic acidosis and then focally elaborate the two components until the metabolic acidosis is described in terms of the quantity of serum bicarbonate lost.[6] We then aggregate the joint effects to derive the actual severity of metabolic acidosis.

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 14-8 shows the primary change in serum pH caused by low serum bicarbonate and the response of the respiratory system in defense against 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.

---

[5]This is a hypothetical example; in the program this component summation will take place at the pathophysiological level.

[6]The quantities of serum bicarbonate lost may be summed by simply adding the loss due to each cause to evaluate their combined effect.

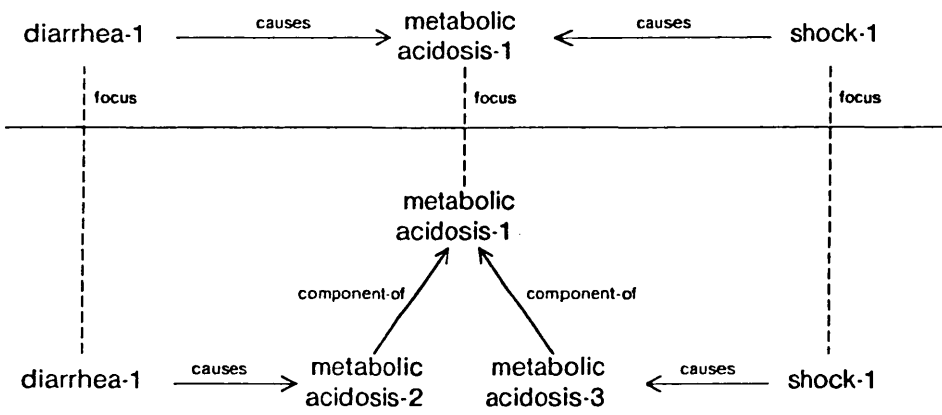

FIGURE 14-7 Metabolic acidosis caused by diarrhea and shock.

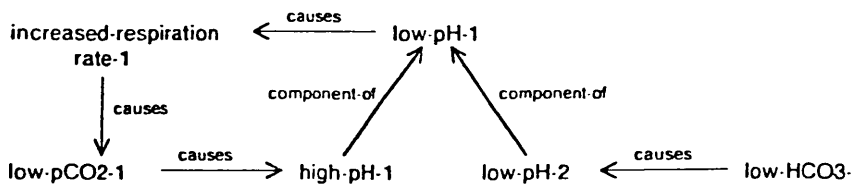

FIGURE 14-8 Primary change in serum pH caused by low serum bicarbonate causes response by respiratory system.

The decomposition of an effect with multiple causes into its causal components also provides us with valuable information in evaluating prognosis and in formulating therapeutic interventions.

14.4 The Patient-Specific Model

Diagnosis is the process of actively seeking information and identifying the disease process(es) causing the patient's illness. In other words, diagnosis involves ascertaining the facts and their implications. The effectiveness of the information-gathering process depends on the analysis of the available facts. From our experience with the existing diagnostic systems (Pople et al., 1975) (see also Chapter 6) we are convinced that a relatively simple

representation of physician's analysis of patient's illness (i.e., a list of disease hypotheses) is incapable of providing the desired level of expertise. The patient description must unify all known facts about the patient, their interpretations, their suspected interrelationships, and disease hypotheses in order to explain these findings. Finally, we observe that at any point in diagnostic reasoning practiced by human experts, there are only a few significantly different explanations for the patient's illness under consideration.

In the program, each such explanation is represented by a patient-specific model (PSM). Note that within each PSM all the diseases, findings, etc., are mutually complementary, while the alternate PSM's are mutually exclusive and competing. In this section we describe procedures for building and extending a patient-specific model based on the known findings and the program's medical knowledge. These operations are *initial formulation* to create an initial patient description from the presenting complaints and laboratory results, *aggregation* to summarize the description at a given level of detail to the next more aggregate level, *elaboration* to elaborate the description at a given level of aggregation to the next more detailed level, and *projection* to hypothesize associated findings and diseases suggested by states in the PSM.

14.4.1 Initial Formulation

From observing the clinical behavior of physicians, we have noticed that when presented with the chief complaints and other voluntarily provided information in a case, the physicians set up a tentative diagnosis. This diagnosis serves as a specific framework that can be used in soliciting information and for organizing the incoming information. Similarly, the program, when provided with the initial findings and a set of serum electrolyte values, constructs a small set of PSM's as its initial possible diagnoses, using the following steps. First, it analyses the electrolytes and formulates all possible single or multiple acid-base disturbances that are consistent with the electrolyte values provided and selects from them a small set that 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 performed by elaborating all known clinical information to the pathophysiological level, where its relationships to the laboratory data are determined by projecting the unique causes and definite consequences of every node. Then the program summarizes these pathophysiological descriptions to the clinical level by repeated application of aggregation operations. This process results in the initial description of the patient at every level of detail. It is this description that is later modified by the diagnostic process as new information becomes available. Note that each of the mechanisms, aggregation, elaboration, and projection, are used in the initial formulation of the PSM.

14.4.2 Aggregation

The aggregation process allows us to summarize the description of the patient's illness at any given level to the next more aggregate level. The summarization of a causal network can be achieved by recognizing that a central node and its surrounding causal relationships may be expressed at a more aggregate level by a single node (called *focal aggregation*) and by summarizing a chain of relations between nodes by a single causal relation between the initial cause and the final effect nodes (called *causal aggregation*).

Focal Aggregation

In aggregating a causal network, we must first identify the nodes in the network that form anchor points (i.e., landmarks, points of special significance) around which the causal phenomenon can be summarized. Consider a partially completed PSM in which some nodes at a detailed level of aggregation have been instantiated. Any of these nodes is an anchor point if (1) in the medical knowledge base such a node is the focus of some node at the next more aggregate level in the network and (2) at least one such higher-level node already exists or can be instantiated within the PSM. If it exists and the constraints on the focal link are satisfied, then the focal link connecting the two is instantiated. If it does not exist, then both it and the focal link are instantiated. Finally, if more than one possible description of the node is consistent with the causal structure above, we defer the aggregation process until we can obtain some additional information to resolve this ambiguity.

Causal Aggregation

Once we have determined the focal aggregations for nodes at a given level of aggregation, we need to describe the causal relations among these aggregate nodes. 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 (called *predecessor paths*) have a node with a focal aggregation, then the focal aggregation of the node either is an ultimate etiology or is totally unaccounted for and does not need to be causally aggregated. 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 instantiating a causal link between the focal aggregation of the node and the first focal aggregation in each path. Finally, if only some of the predecessor paths have nodes with focal aggregations, then the focal aggregation of this node is partially ac-

counted for. The causal aggregation is achieved by decomposing the node into two components: (1) the component due to paths that have focal aggregation (called the *accounted component*), and (2) the component due to paths that do not have focal aggregation (called the *unaccounted component*). Then the focal aggregation of the node is decomposed based on the decomposition at the present level, and the two cases are treated as described above.

14.4.3 Elaboration

Elaboration is the dual of the aggregation operation described above and is used to elaborate the description of a causal network at a given level of aggregation to the next more detailed level. This is achieved by elaborating each link in the causal network by first describing the cause and effect of the link at the next more detailed level (called *focal elaboration*) and then instantiating the causal pathway between these detailed nodes (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 to reflect the better understanding of the causal phenomenon at higher levels of aggregation.

Focal Elaboration

Focal elaboration is the inverse of focal aggregation. To focally elaborate a composite node, the program computes the possible profile of the focal concept associated with the given node. If a node at the next lower level of aggregation matches this profile and is consistent with the node above, the program instantiates the focal link connecting the two. If not, it instantiates the focal node and the focal link connecting the two.

Causal Elaboration

Causal elaboration is the dual of causal aggregation. A composite causal link can be elaborated if the cause and the effect nodes of the link have focal elaborations. To elaborate a composite link, the program matches the causal path associated with the link starting at the focal nodes of the cause and the effect of the link with existing paths in the PSM. If some part of this pathway is not present in the PSM, the program recursively calls itself on each link in the pathway (starting from the focus node of the source) that is absent in the PSM. If the link being recursively elaborated is a primitive link and if its effect node is not present in the PSM, the effect node and the link are instantiated. Otherwise, if the effect node is present,

it matches the attributes of the cause and the effect nodes. If they are compatible, it instantiates the link. Otherwise, if the effect node is an observed node,[7] the program decomposes the effect node and instantiates the link connecting the cause and the component of the effect node contributed by it. Otherwise, if the effect node is accounted for by some other cause, it instantiates the combined effect by summing the components of the two causes. Finally, it aggregates the effect node to revise the description at the next more aggregate level.

14.4.4 Projection

The projection operation is used to hypothesize and explain the associated findings and diseases suggested by the states in the 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 aggregation. 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.

We envision using the projection operation in the diagnostic problem solver for exploring diagnostic possibilities, for evaluating their physiological validity, and in generating expectations about the consequences of hypothesized diagnoses.

14.5 An Example

Let us consider a 40-year-old 70-kg patient who has been suffering from moderately severe diarrhea for the last two days and, as a result, has developed moderately severe metabolic acidosis and hypokalemia. The laboratory analysis of the patient's blood sample (serum analysis) is Na, 140; K, 3.0; Cl, 115; HCO_3, 15; pCO_2, 30; and pH, 7.32.

14.5.1 Initial Formulation

To exercise the program, let us provide it initially with only the laboratory data. Based on these data, the program generates all possible acid-base disturbances that can account for the laboratory data, as follows:

[7]Or if the effect node is a causal predecessor of some observed node that completely accounts for it.

1. metabolic acidosis
2. chronic respiratory alkalosis + acute respiratory acidosis
3. metabolic acidosis + chronic respiratory alkalosis + acute respiratory acidosis
4. metabolic alkalosis + chronic respiratory alkalosis + acute respiratory acidosis

Based on the complexity,[8] likelihood, and severity of each component, the list of possible disturbances is pruned and rank ordered.[9] The rank-ordered list of likely disturbances is

1. metabolic acidosis (severity: 0.4)
2. chronic respiratory alkalosis (severity: 0.68) + acute respiratory acidosis (severity: 0.32)

The program now creates a PSM[10] for each possible acid-base disturbance and asserts in it instantiations of the laboratory data (at the pathophysiological level) and the appropriate acid-base disturbances (at the clinical level). In the rest of the example we will focus on the first acid-base disturbance, metabolic acidosis. The program focally elaborates the metabolic acidosis through the intermediate levels until it reaches the pathophysiological level and thus identifies the amount of HCO_3 loss corresponding to the severity of the metabolic acidosis. Based on this information and the laboratory data, it instantiates the feedback loop corresponding to the acid-base homeostatic mechanism. Next, it projects back[11] each node whose cause can be uniquely determined and projects forward the definite consequences of each node in the PSM. We now have the explanation at the pathophysiological level of the electrolytes consistent with the diagnosis of metabolic acidosis as shown in Figure 14-9.

14.5.2 Aggregation

After the pathophysiological description is completed, this description is aggregated through the intermediate levels to the clinical level of detail. To illustrate this operation, let us consider the low-serum-K-1 node at the

---

[8]Triple disturbances are quite rare and are generally not considered during initial formulation unless there is compelling evidence for their presence.

[9]The rank ordering of the diseases is based on Occam's Razor—simpler hypotheses are preferred.

[10]For ease of explanation, the example described here uses a three-level PSM instead of the five-level PSM used in the program.

[11]Note here that as we are at the pathophysiological level, each link being projected is primitive. Thus projecting back at this level can be restated as instantiating the cause and the link connecting the cause and the effect node.

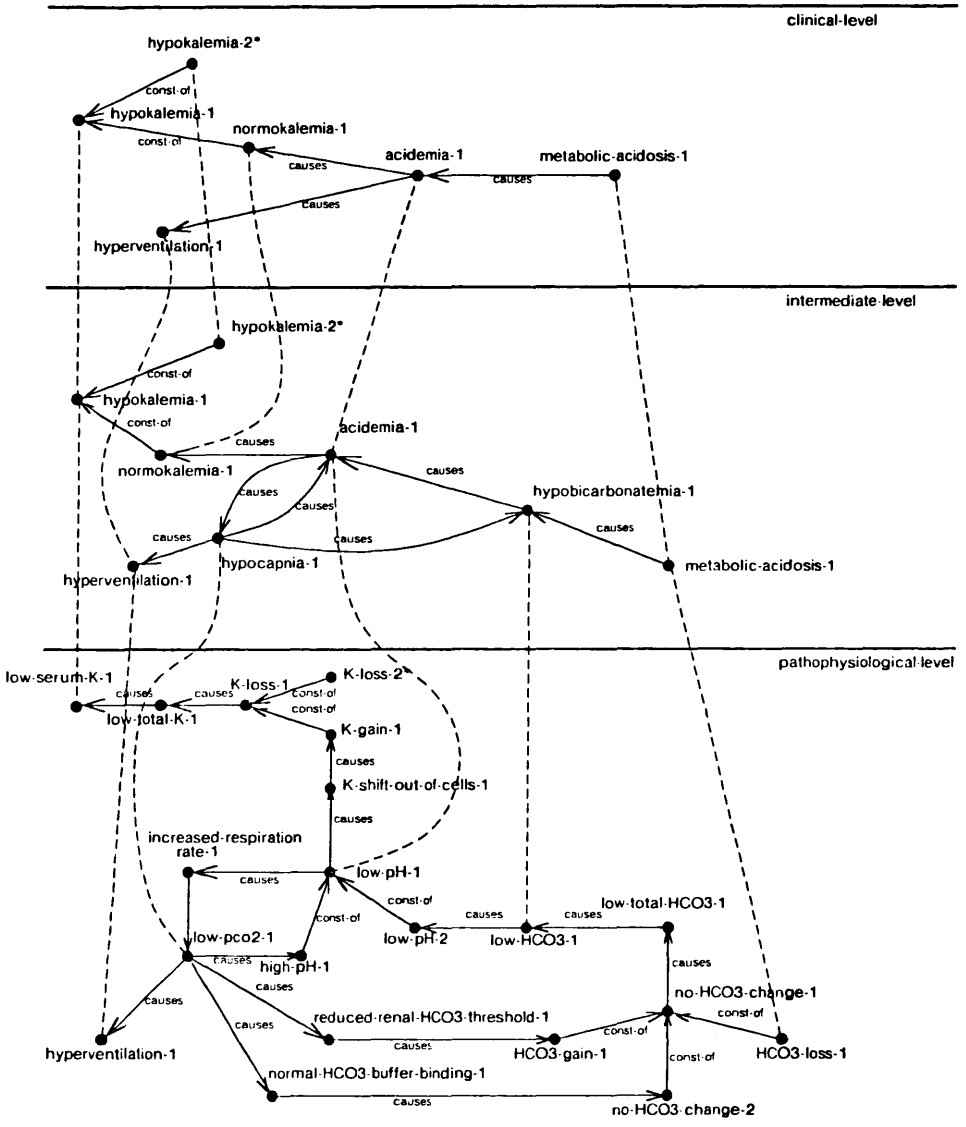

FIGURE 14-9 Initially formulated PSM.

pathophysiological level. Focally aggregating this node, we instantiate hypokalemia-1 as shown in Figure 14-9. To determine the causal aggregation of this node at the next level of detail, we must focally aggregate the first aggregable node in each path leading back, in this case low-pH-1. Focally aggregating low-pH-1, we instantiate acidemia-1. Next, we compute the component of low-serum-K that can be accounted for by low-pH-1 and the

component that remains to be accounted for because of the unaccounted K-loss-2. Then we compute the mapping of these 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 K-loss-2). We then connect the normokalemia-1 to acidemia-1 and mark the hypokalemia-2 as unaccounted (indicated in the figure by an asterisk). Next, in order to causally aggregate low-pH-1, we focally aggregate low-pCO_2-1 and low-HCO_3-1 into hypocapnia-1 and hypobicarbonatemia-1, respectively. As each path leading back from low-pH-1 terminates in a node with focal aggregation, the focal aggregation of low-pH-1 (acidemia-1) is a fully accounted node. Therefore, we connect acidemia-1 to hypocapnia-1 and hypobicarbonatemia-1. This process is repeated for each aggregable node at the current level, and then the whole process is repeated at the next level until we reach the clinical level of aggregation.

14.5.3 Projection

To illustrate the projection operation, let us assume that the diagnostic component has hypothesized that the unaccounted component of hypokalemia at the clinical level (hypokalemia-2) is caused by diarrhea and wishes to determine if this is so and how this assumption fits with the current PSM. The result of this operation is shown in Figure 14-10.

To project the link between hypokalemia and diarrhea, the program evaluates the link to determine the attribute profile of the diarrhea consistent with hypokalemia-2, from which it determines the profile of diarrhea at the next more detailed level. It then attempts to match the causal path associated with the link (hypokalemia ← lower-GI-loss ← diarrhea) at the next level. As none of the links in this pathway are present and as this causal pathway is consistent with the description at the next level, the program recursively calls itself on each link in the path. Considering the first link (that is, hypokalemia ← lower-GI-loss), it finds the causal path associated with this link at the next level of detail (low-serum-K ← low-total-K ← K-loss ← lower-GI-loss). Matching this path with the description in the PSM, it finds that all but one link (K-loss ← lower-GI-loss) is already present. Since this link is primitive, the program evaluates the profile of the lower-GI-loss consistent with the unaccounted component of K-loss and instantiates it and the causal link connecting lower-GI-loss-1 to K-loss-2. To reflect this addition at the higher levels of detail, the program aggregates the low-serum-K-1 (the effect node in the path). As the low-serum-K-1 is now a fully accounted node, the component structure associated with its focal aggregation (hypokalemia-1) is deleted, and the causal links associated with the accounted component of hypokalemia-1 and an additional link from lower-GI-loss-1 are connected to it. This process is repeated until we establish the relation between the diarrhea and hypokalemia at the clinical level.

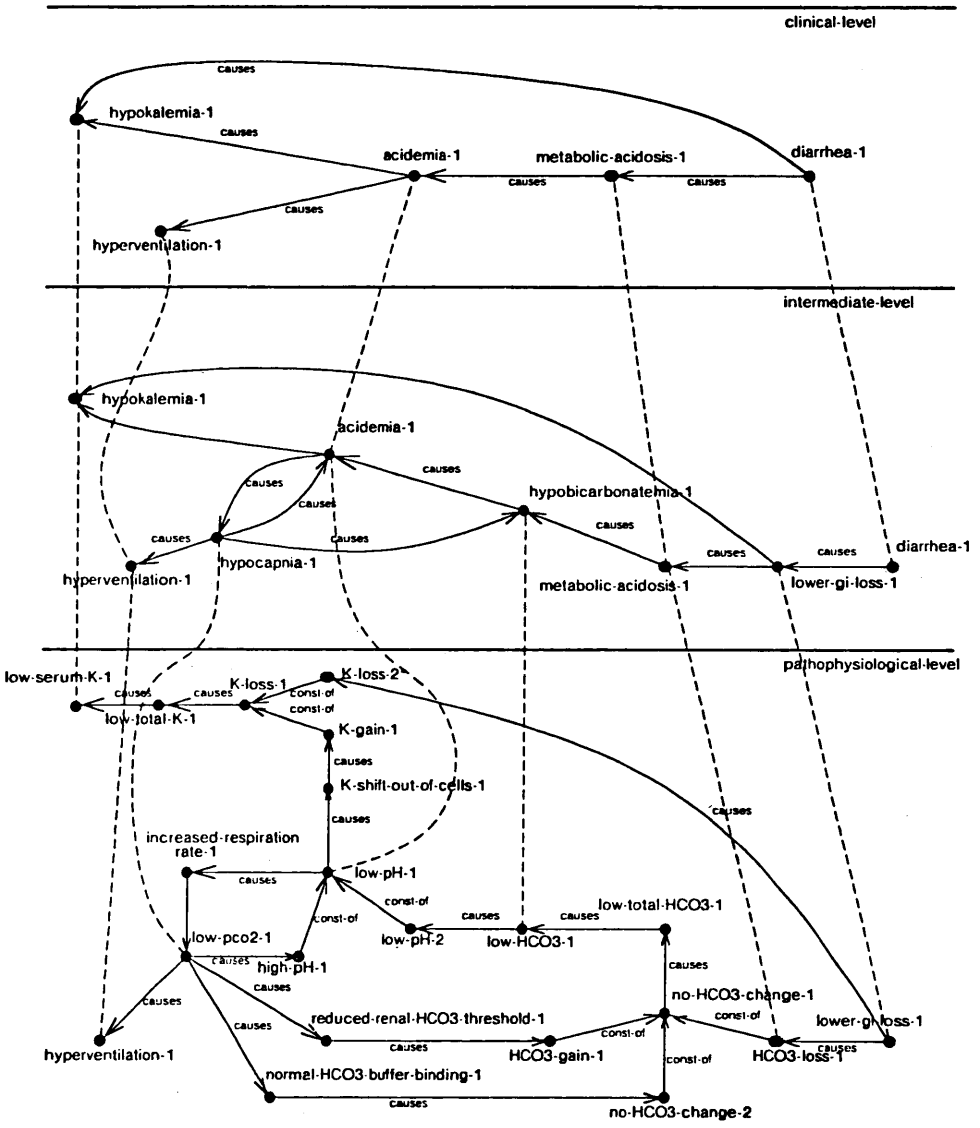

FIGURE 14-10 PSM extended to include diarrhea.

14.5.4 Elaboration

The process of elaboration is similar to that of projection described above and differs from it in two major ways: (1) the causal link and the associated nodes already exist in the PSM at the higher level of aggregation, and (2) we have already determined that the causal link being elaborated is valid.

Therefore, if a causal pathway associated with the link at some level of detail is not consistent with the description in the PSM, the program modifies the PSM appropriately to accommodate the pathway. In the example being described, the second (and more interesting) case does not arise. To demonstrate the elaboration process, let us establish the relation between diarrhea-1 and metabolic-acidosis-1 at the clinical level. The result of elaborating this link is shown in Figure 14-10.

14.6 English Explanation

To illustrate the program's understanding of the patient's illness at various levels of detail, an English generator was implemented to translate the PSM at any given level into its English description.[12] The descriptions are given at three levels of detail in Figure 14-11.

14.7 Conclusion

We have begun a complex and challenging task: to reason about difficult medical problems with a representation that is capable of capturing the subtlety and richness of knowledge and hypotheses used by expert physicians. We have thus far succeeded in creating a representation and a set of structure-building operators that are able to create a patient description based on causal models, multiple levels of detail in description, and the explicit use of components of quantities and states. The various viewpoints on the patient represented by different cuts through this complex description are kept consistent by the operators. We believe that this approach displays a level of understanding not achieved before in medical reasoning programs or other programs that need to describe an organization of hypotheses or mechanisms at different levels of detail.

In continuing to develop our diagnostic and therapeutic programs, we believe that the organizational framework provided by the PSM and its associated operators gives us a suitable machinery for exploring the choice of reasoning strategies and recording our programs' changing conceptions of a case. The rich network of interconnections in the PSM constrains a diagnostic reasoner to generate only a relatively small number of coherent explanations, thereby reducing the space of possibilities to be investigated

[12]The generator makes use of the methodology and some of the code of a generator built by William Swartout as part of an interactive system that explains and justifies portions of expert programs (Swartout, 1981).

Clinical Level

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

Na: 140.0	HCO_3: 15.0	Agap: 13.0
K: 3.0	pCO_2: 30.0	
Cl: 115.0	pH: 7.32	

The diarrhea causes moderate metabolic acidosis, which causes mild acidemia. The acidemia and diarrhea cause mild hypokalemia, and acidemia causes hyperventilation. All findings have been accounted for.

Intermediate Level

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

The diarrhea causes moderate lower GI loss, which causes moderate metabolic acidosis. The metabolic acidosis along with moderate hypocapnia causes moderate hypobicarbonatemia. The hypobicarbonatemia along with hypocapnia causes mild acidemia. The acidemia and lower GI loss cause mild hypokalemia, and acidemia causes hypocapnia. The acidemia also causes hyperventilation. All findings have been accounted for.

Pathophysiological Level

This is a 40-year-old 70.0-kg male patient with moderate lower GI loss. His electrolytes are: . . . .

Moderate lower GI loss, reduced renal HCO_3 threshold, and normal HCO_3 buffer binding jointly cause no HCO_3 change. The no HCO_3 change causes low ecf HCO_3, which causes low serum HCO_3. The low serum HCO_3 and low serum pCO_2 jointly cause low serum pH. The low serum pH causes K shift out of cells and causes increased respiration rate. The increased respiration rate causes low serum pCO_2, which causes normal HCO_3 buffer binding. The low serum pCO_2 also causes reduced renal HCO_3 threshold and increased respiration rate causes increased ventilation. The lower GI loss and K shift out of cells jointly cause K loss. The K loss causes low ecf K, which causes low serum K. All findings have been accounted for.

FIGURE 14-11 English explanation at different levels of detail.

in seeking a diagnosis. In particular, enforcing the requirements of causal consistency (at each appropriate level of detail) on any tenable explanation provides a means of pruning the diagnostic space and permits us to try a "hypothesize and debug" reasoning strategy. The multilevel interconnections of the PSM also help us merge decisions and considerations we have described as categorical and probabilistic. Although much work clearly remains before developments such as those described here form the fabric of truly successful medical consulting systems, we have proposed here a useful new representational basis for such work.

ACKNOWLEDGMENTS

This research was supported (in part) by a National Institutes of Health grant (No. 1 PO1 LM 03374-02) from the National Library of Medicine.