Qualitative Simulation as Causal Explanation

Benjamin J. Kuipers

AI TR88-24        April 1988
Qualitative Simulation as Causal Explanation*

Benjamin Kuipers

Department of Computer Sciences
University of Texas at Austin
Austin, Texas 78712
January 24, 1986

"An individual fact is said to be explained, by pointing out its cause, that is, by stating the law or laws of causation, of which its production is an instance. ... a law or uniformity in nature is said to be explained, when another law or laws are pointed out, of which that law itself is but a case, and from which it could be deduced."

John Stuart Mill, A System of Logic, Book III, Chapter 12, Section 1 (1856).

"To give a causal explanation of an event means to deduce a statement which describes it, using as premises of the deduction one or more universal laws, together with certain singular statements, the initial conditions."


1 Introduction

How can a medical diagnosis program show that a diagnostic hypothesis explains a set of observed findings? Following the model of scientific explanation described above, it must demonstrate how

*This research was supported in part by the National Library of Medicine through NIH Grants R01 LM 04135, LM 04374, and LM 04515, and by the National Science Foundation through grants MCS-8303640, DCR-8417934, and DCR-8512779.
the observed findings are consequences of a particular set of physiological mechanisms (the laws) responding to the state of the patient (the context). A correctly functioning physiological mechanism provides a set of specialized "laws" governing the way the body responds to certain events. If the mechanism becomes defective, the way the body responds to its environment is modified, essentially "changing the laws." A diagnostic hypothesis may propose as the primary cause of disease either a modified law, e.g. a regulatory mechanism losing its sensitivity to the quantity being regulated; or a modified context, e.g. excessive salt intake.

In general, different manifestations of a disease may be explained by several mechanisms affected in different ways. In a kidney disorder called the nephrotic syndrome, a defective membrane in the kidney leads to protein loss into the urine which the filter was supposed to retain in the blood. The laws of this particular mechanism were modified. This has the effect, through two normally-operating mechanisms, of first shifting the fluid balance between the plasma and the tissues, and second, shifting the balance between ingested, retained, and excreted salt and water. These balance mechanisms follow the correct laws, but in an abnormal context, leading to an abnormal state. Thus we can say that the resulting swelling of the tissues is explained by the diagnosis of nephrotic syndrome, through three mechanisms, with the primary cause being loss of protein from the plasma.

Existing medical diagnosis programs do not reason this way. Early approaches, such as MYCIN [Shortliffe, 1976], Internist-1 [Miller, Pople, Myers, 1982], and the Present Illness Program [Pauker, et al, 1975], are based on accumulating conditional probabilities of both the disease given the findings and the findings given the disease [Szolovits and Pauker, 1978]. Table 1 shows fragments of knowledge from MYCIN and from Internist-I. In MYCIN, the certainty factor (.8) is, roughly speaking, a scaled conditional probability of the diagnosis given the observations. It is composed with the confidence factors of the premises to determine the confidence of the conclusion. In Internist-I, each observable feature of a diagnostic hypothesis has two scores: the evoking strength which measures the likelihood of the hypothesis given the observation, and the frequency weight which measures the likelihood of the observation given the hypothesis. The relations between each finding and the diagnostic hypothesis are treated as independent of each other, to make the scoring algorithm tractible. In such systems, a diagnostic hypothesis is accepted once its probability, compared with its alternatives, is sufficiently high. The diagnosis, in turn, is held to account for a finding if the probability of the finding given the diagnosis is sufficiently high. This method of
“accounting for” a finding is quite different from the deductive method of scientific explanation.

**** Table 1 near here. ****

It is also important, in medical diagnosis and explanation, to recognize that the human body operates by elaborate and robust mechanisms. Even in serious illness, the fact that the patient is alive means that most physiological mechanisms are functioning largely correctly, although perhaps in an abnormal context. Most disease states correspond to relatively small perturbations of the healthy mechanisms. Thus, adequate explanation of disease states must depend on the ability to understand and explain the healthy physiological mechanisms.

An adequate explanation of observed findings by a diagnostic hypothesis must include:

1. a description of the correctly functioning physiological mechanisms,
2. a description of the change to the mechanism or context asserted as primary cause by the diagnostic hypothesis,
3. a demonstration that the perturbed mechanism can predict the relevant manifestations observed in the patient.

Such a demonstration argues that the hypothesized primary cause provides a causal explanation for the observations.

We might hope that biomedical science will produce precise and accurate models of the human body in complete detail, to allow us to predict the effect of a disorder in a particular patient. Unfortunately, the knowledge of the individual expert physician, and even that of the scientific community itself, is typically incomplete. We lack a complete understanding of the structure and behavior of important physiological mechanisms. Similarly, the physician lacks a complete description of the internal state of the patient. Parameters that are important to the description of a mechanism may be unobservable in the clinical setting, or even in the laboratory. Nonetheless, physicians do appear to reason with causal explanations and mental models of the mechanisms of the body [Kuipers and Kassirer, 1984].

The reasoning methods physicians use appear to be qualitative. For example, they may conclude that “blood pressure is very high but not immediately life-threatening,” or that “salt retention causes fluid retention.” Physiological mechanisms are described in terms of continuously varying
quantities, e.g. blood pressure, and their values at particular points in time. The values are described qualitatively rather than quantitatively. The functional relations between parameters are often known to be monotonically increasing or decreasing, but without further specification. [Kuipers and Kassirer, 1984.] It is perhaps surprising that it is possible to construct a knowledge representation with these properties that is capable of correct qualitative predictions about the behavior of important physiological mechanisms.

In the remainder of this paper, we will present a knowledge representation capable of expressing a qualitative description of the general laws of a mechanism and the context within which it operates. We will also demonstrate, through an example of a kidney disorder,

- how we can infer the responses of the healthy mechanism in normal and abnormal contexts,
- how a simple perturbation of the mechanism description yields a model of the disease state, and
- how we can model the response of the disease mechanism to therapy.

We conclude by presenting a new type of abstraction relation that may help link simple equilibrium mechanisms into a hierarchical description of a complex mechanism such as human physiology.
2 Qualitative Simulation Overview

Many physical mechanisms are effectively described by differential equations. However, to formulate a differential equation requires specifying all of the local constraints between parameters as particular mathematical functions, at least down to certain coefficients which may be left symbolic. If some of these constraints are known only qualitatively, we need a descriptive language which is an abstraction of differential equations, but which nonetheless can be solved or simulated to infer the behavior of the system. The QSIM representation and simulation algorithm [Kuipers, 1984, 1985, in press] has these properties. Starting with a mechanism model consisting of qualitative constraints that abstract the relationships in a differential equation, QSIM predicts the possible qualitative behaviors of the mechanism (figure 1).

**** Figure 1 near here. ****

Qualitative simulation of a mechanism starts with a set of constraints modeling the structure of the mechanism and its initial state, and produces a graph consisting of the possible future states of the system. The possible behaviors of the mechanism are the paths through the graph starting at the initial state.

The constraint model consists of a set of symbols representing the physical parameters of the system (continuously differentiable real-valued functions of time), and a set of constraints on how those parameters may be related to each other. The constraints are two- or three-place relations on physical parameters. Some specify familiar mathematical relationships such as $\text{DERIV}(\text{vel}, \text{acc})$, $\text{ADD}(\text{net}, \text{out}, \text{in})$, $\text{MULT}(\text{mass}, \text{acc}, \text{force})$, $\text{MINUS}(\text{fwd}, \text{rev})$. Others assert qualitatively that there is a functional relationship between two physical parameters, but only specify that the relationship is monotonically increasing or decreasing: $M^+(\text{price}, \text{power})$ and $M^-(\text{mph}, \text{mpg})$. It is the $M^+$ and $M^-$ constraints which make this representation qualitative, and thus able to represent states of incomplete knowledge. The constraints are designed to permit a large class of differential equations to be mapped straight-forwardly into constraint models. Appendix A shows the Lisp form of a constraint model. A graphical representation of the model is used in the figures in section 3.

Each physical parameter is a continuously differentiable real-valued function of time with only finitely many critical points. Its value at any given point in time is specified qualitatively, in terms
of its order relations with a totally ordered set of landmark values. The landmark values for a function include zero, \( \pm \infty \), and all of the known critical values of the function. Landmark values may be described either numerically (e.g. zero) or symbolically; their ordinal relationships are their essential properties. A qualitative value is either equal to a landmark value, or refers to the open interval between adjacent landmarks. As the qualitative simulation proceeds, it can discover new critical points and thus add new landmark values to the sequence. The qualitative state of a parameter consists of its ordinal relations with the landmark values and its direction of change.

Time, within one possible behavior, is represented as a totally ordered set of symbolic distinguished time-points. The current time is either at or between distinguished time-points. All of the time-points are generated as a result of the qualitative simulation process.

**** Figure 2 near here. ****

The predicted behaviors are best displayed graphically as qualitative plots such as Figure 2. The vertical axis represents the set of landmark values for the parameter; the only meaningful vertical positions are at, or midway between, landmark values. For example, a point plotted midway between the landmarks 0 and \(-\infty\) indicates that the value of the parameter lies in the open interval \((-\infty,0)\); i.e. it is negative. The horizontal axis represents the time-sequence of predicted behavior. To reduce visual clutter on the plot, time points are not labelled. Distinguished time-points are indicated by tick-marks on the axis; qualitative descriptions representing the state of the system in the open interval between distinguished time-points are plotted in the middle of the interval. The symbol plotted at each point may be up-arrow, down-arrow, or circle, indicating the direction of change. In the examples below, certain known states (e.g. NORMAL) are plotted to the left of the time sequence, for reference.

At a distinguished time-point, if several physical parameters linked by a single constraint are equal to landmark values, they are said to have corresponding values which can be represented and used by the qualitative simulation. Corresponding values provide additional qualitative constraints on the behavior of structural relationships otherwise described only as \( M^+ \) or \( M^- \). The case of corresponding values \((0,0)\) is sufficiently common to justify a special notation, \( M^+_0 \) and \( M^-_0 \).

The initial state of the system is defined by set of qualitative values for the physical parameters. The qualitative simulation proceeds by determining all possible changes in qualitative value
permitted to each parameter, then checking progressively larger combinations of qualitative transitions and filtering out the inconsistent ones. If more than one qualitative change is possible, the current state has multiple successors, and the simulation produces a tree. The predictions generated by the QSIM algorithm are guaranteed to include all actual behaviors. With second order, oscillatory mechanisms, there is the possibility of impossible behaviors being predicted because of the qualitative nature of the simulation process [Kuipers, in press].

The precise definition of this representation and the QSIM algorithm for qualitative simulation are given in detail in [Kuipers, 1985, in press]. Next we will describe the physiological mechanism behind water balance, and show how this qualitative simulation representation can produce explanatory models of its behavior.
Example: Water Balance and its Disorders

3.1 Anti-Diuretic Hormone Controls Water Balance

The balance between water intake and excretion is maintained by the kidney and the hypothalamus, primarily by sensing and responding to changes in the concentration of sodium in the blood. Since the total amount of sodium in the body is reasonably stable over short time periods, this amounts to regulating the total volume of water in the body.

The kidney, in general, works by filtering the plasma (the fluid component of the blood), along with all but the cells and the largest molecules dissolved in it, by the nephrons. Each nephron consists of a capillary filter called the glomerulus, attached to a long tubule which leads, eventually, to an external collecting system and to the bladder. Thus, anything which is filtered into the tubular system will be excreted unless it is reabsorbed through the walls of the tubules back into the bloodstream. (Certain substances are also actively secreted into the tubules for elimination, but that is not important here.) For water, and critical electrolytes such as sodium, the vast majority of what is filtered is reabsorbed again. The overall balance between intake and excretion is maintained by mechanisms that exercise a precise and delicate control over the balance between filtering and reabsorption.

From the point of view of the computer scientist, these mechanisms are strikingly clever and elegant. An important quality to consider in the design of a computer program is robustness: the ability of the program to handle inputs that had not been anticipated by the designer. A particularly elegant aspect of the design of the kidney is that it can excrete substances never before encountered in the biological or evolutionary history of the organism. Filtering is non-selective: every molecule below a certain size passes with the plasma into the nephron. Reabsorption is selective: there are a variety of active and passive mechanisms for controlling the levels of certain important substances. Everything else is simply excreted.

Water balance is maintained by controlling the passive diffusion of water from the relatively dilute solution within the collecting duct into a much more concentrated environment surrounding it. (How that concentrated environment is created and maintained is another story, and a fascinating

---

1Notice to the reader. The medical content of this example reflects the understanding of the author, a computer scientist, with the aid of collaborating physicians and medical and physiology references. The purpose of the example is to illustrate the capability of a representation and inference technique to express a particular type of knowledge.
A hormone, called anti-diuretic hormone or ADH, regulates the permeability of the collecting duct walls to this outward diffusion of water. The more ADH, the more permeable are the walls to diffusion, and the less water is excreted. Hence, ADH is anti-diuretic, and leads to the retention of water by increasing its reabsorption. The hypothalamus is sensitive to the concentration of sodium in the plasma, and responds to high concentration by "telling" the pituitary gland to secrete more ADH, which in turn "tells" the kidney to retain water and dilutes the plasma.

The mechanism can be illustrated by the consequences of drinking a large glass of water. Ingestion of water dilutes the sodium in the plasma and the low plasma sodium concentration is sensed by certain collections of cells in the hypothalamus. In turn, these cells reduce both the secretion of ADH and its release into the plasma. The reduction in ADH reaching the kidney makes the collecting duct wall less permeable to water, so water is lost into the urine. The cycle is now complete: water is ingested, and water is excreted. The direction of these physiological events is reversed when water ingestion is withheld and an individual becomes "dehydrated."

With respect to salt, the mechanisms by which the body handles salt are less well defined. What is known is that changes in the volume of body fluids have a dramatic effect on salt excretion: increased volume enhances salt excretion and decreased volume reduces salt excretion. The intermediate physiologic mechanism is thought to reside, at least in part, in a hormone secreted by the wall of the atrium of the heart, in response to changes in pressure: high salt intake leads to high volume; high volume yields high pressure, which in turn yields more atrial hormone and the hormone promotes salt excretion. Again the cycle is complete, and overall sodium balance is maintained.

Things can go wrong, of course. If a person's salt intake is very high, the pituitary secretes more ADH and more water is retained. Increased blood volume causes increased blood pressure and cardiovascular load, which may be unhealthy in some instances. In this case, the water balance mechanism is functioning perfectly, but responding to an abnormal environment (high sodium concentration) to produce a consequence (high blood volume) that can be harmful.

In the Syndrome of Inappropriate ADH Secretion (SIADH), the body is flooded with excess ADH, for example by a hormone-secreting tumor. Although the sodium concentration sensor and other parts of the mechanism may be operating correctly, sustained high levels of ADH increase water reabsorption and the water balance mechanism loses its sensitivity to changing concentrations.
of plasma sodium. The result is substantial water retention, and low plasma sodium concentration, even with normal water intake. At the same time, the normal volume response remains operative, and the expansion of volume resulting from water retention produces a salt loss. The salt loss contributes further to the fall in plasma sodium concentration. The usual therapeutic response is restriction of water intake, an action that permits water loss and restoration of normal body volume and plasma sodium concentration.

3.2 A QSIM Constraint Model of the Water Balance Mechanism

The QSIM constraint model of the water balance mechanism is presented graphically in figure 3. The terms in figure 3 represent the parameters of the constraint model; the boxes and connecting lines represent the constraints themselves. The graphical representation can be helpful in that information flows only along the links shown.

The multiplication constraint at the top of the figure represents the definition of sodium concentration in terms of the amounts of sodium and water in the plasma compartment (P). Two parallel paths of $M^+$ constraints leading down from there represent the filtered and reabsorbed flows of water, eventually combining to yield a net rate of water excretion. This is balanced against water intake to yield a net rate of change for overall water volume. The sodium concentration in the urine is also included as the ratio of the excretion rates for sodium and water, since it is closely related to a frequently measured parameter, even though it plays no essential role in the equilibrium.

**** Figure 3 near here. ****

This equilibrium mechanism operates in an environment where the total amount of plasma sodium, the rate of water intake, and the rate of sodium excretion (i.e. the parameters underlined in the diagram) are assumed to be constant, though not necessarily at their normal values. The values of these constant parameters constitute the environment within which the mechanism operates. The $M^+$ constraints, for example that between $c(Na, P)$ and $c(ADH, P)$ represent embedded processes whose detailed structure or behavior are not specified, and which operate at a time-scale much faster than the current process. (This concept is explored further in section 4.)

SIADH can be modeled by a modified constraint model (figure 4) in which the $M^+$ constraint between $c(Na, P)$ and $c(ADH, P)$ is deleted, and $c(ADH, P)$ is asserted to be constant and higher than normal.
In the next sections, we will see how these constraint models can be simulated to predict the behavior resulting from:

- the normal water balance mechanism responding to increased plasma sodium (normal water intake);
- the normal water balance mechanism responding to increased water intake (normal plasma sodium);
- the defective, SIADH mechanism responding to normal water intake and sodium levels by accumulating increased water volume and decreasing plasma sodium concentration;
- the SIADH mechanism, starting from its abnormal equilibrium, responding to restricted water intake.

3.3 Normal Response to Environmental Changes

The normal behavior of a homeostatic mechanism such as the water balance mechanism is to respond to environmental changes to maintain balance. The constraint model in Figure 3, in spite of the qualitative $M^+$ constraints representing incompletely known relationships among various parameters, is able to predict the qualitative responses of the mechanism to environmental changes. Changes to the environment are modeled by changing the values of the fixed, "context" parameters and the initial state description.

Suppose first that we wish to determine the response of the mechanism to a sudden increase in water volume. This might be the result of a "spike" in the rate of water intake: a very large drink of water modeled by a discontinuous increase in $amt(water, P)$, with the parameter $net\ flow(water, intake \rightarrow P)$ left constant at its normal value. The other environmental parameters, $amt(Na, P)$ and $net\ flow(Na, P \rightarrow U)$, remain fixed and equal to their normal values. As shown in figure 5, after propagating the initial assertions to determine values for all parameters in the mechanism, QSIM predicts that low plasma sodium concentration causes decreased ADH levels to reduce reabsorption, so that excretion increases until volume returns to its normal value, along with all the other parameters.
A second normal response occurs when the rate of water intake increases and remains constant at a high value. The other environmental parameters, $\text{amt}(Na, P)$ and $\text{net flow}(Na, P \rightarrow U)$, are again constant and equal to their normal values. Initially, we assume that plasma water volume, $\text{amt}(water, P)$, is normal. Figure 6 shows how volume begins to increase in response to increased intake. At the same time, filtered water increases while reabsorption decreases, so that the increase in volume stops at a new steady state. As long as water intake remains high, however, volume cannot return to normal, so a new steady state is reached, with increased volume, increased urine output, and decreased plasma sodium concentration.

The same methods can be used to predict the results of several simultaneous changes to the external environment. Of course, since the qualitative description does not include magnitudes, if the two changes have opposite effects, such as an increase in $\text{amt}(Na, P)$ along with a decrease in $\text{net flow}(water, ingest \rightarrow P)$, the prediction will branch according to which effect dominates.

### 3.4 Abnormal Response to Normal Environment

The purpose of this causal modeling approach is to explain disease states in terms of perturbations of the structure of the physiological mechanism and its resulting behavior. Suppose a person is initially in a normal water balance, and suddenly develops SIADH, as described by the constraint model in figure 4.\(^2\)

For this qualitative simulation, all of the context parameters are set to their normal values and held constant, except that $c(ADH, P)$, no longer controlled by $c(Na, P)$, is set abnormally high and constant. As shown in figure 7, the water balance mechanism reabsorbs water excessively, increasing volume until the amount filtered balances the amount reabsorbed, and the system regains its balance with increased water volume and decreased plasma sodium concentration.

\(^2\)Realistically, of course, these conditions develop gradually. We are currently attempting to extend the representation to handle the "fading away" of an $M^+$ relationship, such as the one between $c(Na, P)$ and $c(ADH, P)$.  

Note that although certain parameters, such as water volume and plasma sodium concentration, make this situation look very similar to increased water consumption (figure 6), other parameters such as rate of water excretion and urine sodium concentration distinguish between them. One advantage of relatively complex physiological models is that they predict configurations of changes to an entire set of physiological parameters. We plan to analyze this information for distinguishing characteristics usable in diagnosis.

3.5 In Abnormal State, Response to Therapy

The ultimate treatment for SIADH is to remove the cause of the problem, for example by surgical removal of a hormone-secreting tumor. In the meantime, however, the usual therapeutic intervention is restriction of water intake. We can model the effect of restricted water intake on the SIADH mechanism by initializing it in the final equilibrium state of figure 7, named STABLE in figure 8, then setting water intake, \( \text{net flow}(\text{water, ingest} \rightarrow P) \) to a value less than normal and constant.

**** Figure 8 near here. ****

QSIM correctly predicts that plasma water volume decreases and plasma sodium concentration increases. The qualitative simulation branches three ways, however, according to whether the new equilibrium point of \( \text{amt}(\text{water}, P) \) is greater than, equal to, or less than, the normal water volume. We can see that these three alternatives correspond to the severity of the water restriction, but QSIM currently has no way of expressing that distinction. In all three cases, \( \text{net flow}(\text{water, ingest} \rightarrow P) \) is simply described as being between zero and the normal value. The next section discusses this problem in the context of the broader problem of hierarchical description of complex mechanisms, and proposes a solution.
4 Hierarchical Descriptions of Complex Mechanisms

4.1 Time-Scale As An Abstraction Relation

The complexity of an enormous mechanism like human physiology can only be managed with a hierarchical description consisting of many simple views, each capturing one level of detail, linked by abstraction relations. An abstraction relation allows a description at one level of detail to refer to other descriptions without taking into account all the distinctions made at other levels of detail.

One approach to hierarchical description of mechanisms uses the part-whole relation as the abstraction relation. A mechanism is viewed as composed of components parts connected by interaction paths, e.g., electronic circuits [deKleer 1979, 1984]. A component is treated as a “black box,” obeying certain behavioral specifications. When examined in greater detail, each component is a mechanism with parts of its own, whose behavior implements the specifications.

In medical physiology, the causal structure and behavior of a mechanism is not as directly derivable from physical anatomy as it is in electronic circuits. Physiological mechanisms consist of large numbers of interacting equilibrium mechanisms, continually adjusting to changing circumstances. The proper unit to treat as a simple view of a physiological phenomenon is the equilibrium mechanism.

In this section we discuss a different abstraction relation among mechanisms operating at significantly different time-scales. A mechanism that operates over days can treat one that reaches equilibrium in seconds as being instantaneous. Conversely, a fast mechanism can treat parameters controlled by a slow one as constants.

Starting from an equilibrium state, if we perturb one of the fixed, context parameters of a constraint model to a new value where it remains fixed, after some delay the mechanism settles into a new equilibrium state with new values for the responding parameters. Under two assumptions, we can abstract the relation between the perturbed context parameter $X$ and a particular responding parameter $Y$ into a functional constraint: $M^+(X,Y)$ or $M^-(X,Y)$.

1. All other context parameters are held fixed.

2. The new $M^+$ constraint is considered only at a time-scale where the equilibrium mechanism can be treated as responding instantaneously to a perturbation.
For example, the water balance mechanism operates in a matter of minutes to respond to increases in water intake by increasing water excretion. At the same time, of course, total plasma water volume increases. At a time-scale of hours to days, then, we may treat the water balance mechanism as instantaneous and summarize this relationship by a single $M^+$ constraint between $\text{net flow}(\text{water}, \text{ingest} \rightarrow P)$ and $\text{amt}(\text{water}, P)$.

If water intake is held constant, but there are long-term variations in sodium intake, we may abstract the same mechanism a different way to obtain the constraint $M^+(\text{amt}(\text{Na}, P), \text{amt}(\text{water}, P))$. This is a key constraint in the sodium balance mechanism, which operates at a time-scale of hours [Kuipers and Kassirer, 1985].

Notice that a relationship which is temporally asymmetrical at the detailed level is abstracted into a symmetrical constraint by shrinking its temporal extent to zero. Within the water balance mechanism, a change to $\text{net flow}(\text{water}, \text{ingest} \rightarrow P)$ eventually produces the result that $\text{amt}(\text{water}, P)$ is changed in the same direction. Abstracted, the two parameters are treated as changing simultaneously, their values related by an unspecified monotonic function described qualitatively by an $M^+$ constraint. By implicitly recognizing the underlying causal mechanism, we normally read $M^+(X,Y)$ as meaning that “a change in X causes a change in Y,” or “a change in Y must have been caused by a change in X.” Thus, although the QSIM simulation algorithm treats an $M^+$ constraint as symmetrical and non-causal at a single level of detail, a trace of its derivation from some faster underlying process can preserve the asymmetry of the causal relationship.

4.2 New Landmarks To Express New Distinctions

The ability to abstract behaviors of an equilibrium mechanism to a monotonic constraint can clarify ambiguous situations such as the effect of water restriction therapy on SIADH (figure 8).

The mechanism begins in the stable state resulting from SIADH with normal water intake, including abnormally high volume. Upon restricting water intake, volume decreases, branching three ways according to whether the final volume is greater than, less than, or equal to the normal volume. Intuitively, we know that this corresponds to the severity of the water restriction. But how to allow the symbolic representation to determine this?

The water balance mechanism, even with SIADH, remains an equilibrium mechanism, and all context parameters have fixed values except for water intake, so we can abstract the effect of this
mechanism to an $M^+$ relation between $\text{net flow}(\text{water}, \text{ingest} \rightarrow P)$ and $\text{amt}(\text{water}, P)$. At this more abstracted level, the $M^+$ constraint expresses the monotonic relation between severity of water restriction and the resulting plasma volume.

We can go one step further. Since there is a landmark value for normal plasma volume within the range of this functional relationship, the strict monotonicity of the relation requires that there must be a unique corresponding value of water intake. We may thus create a new landmark value for $\text{net flow}(\text{water}, \text{ingest} \rightarrow P)$ to represent the water intake that results in precisely normal plasma volume. Of course, there is not enough information to determine the quantitative value of the landmark, but we can discover and name this new qualitative distinction in the space of input values, thereby clarifying the reason for the branching prediction.

In the QSIM algorithm, new landmark values are created only when the simulation detects new critical points of a parameter. This method utilizes the abstraction relation to infer new landmarks, mapping branching behavioral predictions from a mechanism onto new distinctions in the space of input parameters.

These distinctions are particularly useful when mapping predictions onto actual clinical observations. For example, if we know that a relatively severe water restriction still resulted in increased water volume and decreased plasma sodium concentration, we may conclude that normal values are only reachable by even more severe water restriction, which might be contraindicated. This type of conclusion, though obvious to humans, is a type of commonsense causal reasoning about mechanisms that is not present in current medical diagnosis programs.
5 Current Status and Future Directions

The QSIM algorithm has been implemented on the Symbolics 3600 and tested on numerous examples from medical physiology and commonsense physics. The examples in section 3 have all been run; figures 5 through 8 are screen images of QSIM predictions. The abstraction relations discussed in section 4 have not yet been added to the implementation.

We are currently in the midst of designing, implementing, and evaluating an expert system called RENAL which will use the interaction between the first-generation, feature-matching approach to diagnosis and a causal reasoning system to do medical diagnosis in a domain of kidney diseases. Observations of physicians in action make it clear that the hypothesis-driven approach is much more common than recourse to causal reasoning from first principles [Elstein, et al., 1978; Kassirer and Gorry, 1978]. Furthermore, it is not feasible to use physiological knowledge alone to reason backwards from observations to causes. Accordingly, we have selected the problem-solving architecture shown in figure 9, combining a first-generation hypothesis-driven program with a causal model.

**** Figure 9 near here. ****

- Plausible hypotheses are generated by a simple, hypothesis-driven diagnostic program called the Evoker which takes clinical findings and evokes a set of hypothesized diseases.

- Each disease hypothesis is associated with descriptions of the physiological mechanisms and primary causes from which the hypothesis would explain the observed findings.

- The mechanism description is then simulated from the specified initial state to produce a qualitative description of the possible behaviors of that physiological system.

- The results of the simulation are matched against clinical observations (possibly requiring further questions about the patient). If a hypothesis is confirmed, qualitative simulation produces a rich description of both externally observable and internal parameters, and their evolution over time starting with the primary cause.

This problem-solving method may be viewed as an instance of either the Generate and Test, or the Hypothesize and Match weak method, depending on how the components are grouped [Newell, 1973].
Thus, RENAL will evaluate diagnostic hypotheses according to their ability to explain the observed findings, in the sense of scientific explanation. General laws, expressed as constraint models of physiological mechanisms, operating in the context of specific facts about the patient, will be asked to predict the observations starting from hypothesized primary causes. Only by operating at a qualitative level of description can this approach be feasible, because of necessarily incomplete knowledge of biomedical science and of the state of the individual patient. The QSIM representation makes this type of qualitative reasoning feasible and mathematically tractible.

By providing access to explanatory mechanism models of the underlying physiology, we hope that this problem-solving framework will allow us to explore a number of deep problems in diagnosis such as the following (cf. [Patil, 1981; Pople, 1982]). How can we reason about multiple simultaneous diseases? How can we tell which observed findings are fully accounted for by a particular hypothesis, and which remain to be explained? Using this approach we hope to build medical diagnosis programs that not only know about individual diseases, but that also have real knowledge of "how the body works."
6 Acknowledgments

Much of the work reported here was done while the author was at the Clinical Decision Making Group at the MIT Laboratory for Computer Science. I am grateful to all the members of that group for a stimulating environment, and especially Bill Long, Ramesh Patil, and Peter Szolovits. My medical collaborator, Dr. J. P. Kassirer, has been very helpful over a long period, and is responsible for the accurate parts of the medical knowledge reported here. I am responsible for the rest. Robert Causey has been very helpful with philosophical questions.

7 References


A The ADH+WATER Constraint Model

The following Lisp form is the actual input form of the constraint model for the water balance mechanism presented graphically in figure 3.

(define-structure ADH+WATER
  (parameters AFP CNH FFWPU ANP CNP CADH RFUP NFPU NFWIP NFWOP CNU NFNPU)
    ; Each parameter X is defined to have landmarks (minf O X* inf), where
    ; X* is the normal value of X.
  (landmarks (ANP (minf O ANP%% ANP* ANP% inf))
    (NFWIP (minf O NFWIP%% NFWIP* NFWIP% inf))
    (NFNPU (minf O NFNPU* inf)))
  (invariants (CNH ((O inf) nil))
    (CADH ((O inf) nil))
    (RFUP ((O inf) nil))
    (NFPU ((O inf) nil))
    (ANP ((O inf) std)) ; The initialization sets
    (NFWIP ((O inf) std)) ; these to appropriate
    (NFNPU ((O inf) std))) ; constant values.
  (constraints (M+ AFP CNH)
    (M+ CNH FFWPU)
    (mult AFP CNP ANP)
    (add RFUP NFPU FFWPU)
    (M+ CNP CADH)
    (M+ CADH RFUP)
    (d/dt AFP NFWOP)
    (add NFPU NFWOP NFWIP)
    (mult NFPU CNU NFNPU))
  (states
    ; The structure must know about its normal equilibrium state in order
    ; to learn corresponding values of the different parameters.)
(normal '((NFWDP (O std)) (NFWIP (NFWIP* std)) (NFPU (NFPU* std))
(RFUP (RFUP* std)) (CADH (CADH* std)) (CNP (CNP* std))
(ANP (ANP* std)) (FFWPU (FFWPU* std)) (CNH (CNH* std))
(AFP (AFP* std)) (CNU (CNU* std)) (NFPNU (NFPNU* std))))

B Tables
MYCIN rule

IF:  
(1) The identity of the organism is not known with certainty, and
(2) The stain of the organism is gramneg, and
(3) The morphology of the organism is rod, and
(4) The aerobicity of the organism is aerobic

THEN: There is strongly suggestive evidence (.8) that the class of the organism is enterobacteriaceae.

Internist-I manifestation list (partial)

Alcoholic Hepatitis

factor VII proconvertin decreased ... 1 2
LDH blood increased ... 1 3
magnesium blood decreased ... 2 2
prothrombin time increased ... 2 3
SGPT 200 to 600 ... 1 2
SGPT 40 to 199 ... 2 3
SGPT gtr than 600 ... 1 1
liver biopsy bile plugging ... 1 2
liver biopsy fatty metamorphosis ... 2 4
...

Table 1: In MYCIN and Internist-I, diseases are described by observable features.
C Figures
Figure 1: Qualitative simulation and differential equations are both abstractions of actual behavior.

Figure 2: A "qualitative plot" of the height of a ball, rising then falling.
Figure 3: The constraint model of the water balance mechanism
Figure 4: The constraint model of water balance with SIADH
Figure 5: The water balance mechanism responding to a volume "spike."

Figure 6: The water balance mechanism responding to increased intake.
Figure 7: The water balance mechanism with SIADH, beginning in a normal state.
Figure 8: The water balance mechanism, with SIADH and water restriction therapy, produces three behaviors.
Figure 9: The Evoker and QSIM in a hypothesize-and-match cycle.