Some Causal Models are Deeper than Others*

Tom Bylander
Laboratory for Artificial Intelligence Research
Department of Computer and Information Science
The Ohio State University

Abstract

The effort within AI to improve the robustness of expert systems has led to increasing interest in "deep" reasoning, which is representing and reasoning about the knowledge that underlies the compiled knowledge of expert systems. One view is that deep reasoning is the same as causal reasoning. Our aim in this paper is to show that this view is naive, specifically that certain kinds of causal models omit information that is crucial to understanding the causality within a physical situation. Our conclusion is that "deepness" is relative to the phenomena of interest, i.e. whether the representation describes the properties and relationships that mediate interactions among the phenomena and whether the reasoning processes take this information into account.

1. Introduction

Most expert systems depend upon *compiled* representations and reasoning processes. Their representations associate data with conclusions, and their reasoning processes use these associations, but they do not take into account the reasons why the data and conclusions are related. Without this extra knowledge, expert systems will be limited in what explanations they can provide and in reasoning about their own limitations.**

Within AI, there has been increasing interest in *deep* reasoning, i.e. representing and reasoning about these "reasons." A number of suggestions have been made that identify deep reasoning with *causal* reasoning. Hart suggests that deep reasoning involves commonsense ideas about causality as well as mathematical modeling (Hart, 1982). Michie suggests that the fundamental laws of the domain constitute deep reasoning (Michie, 1982). A number of programs could be said to perform deep reasoning based on these criteria. Instead of summarizing and comparing these programs, which would probably be confusing rather than enlightening given the plethora of domains and reasoning methods, my strategy is to take one program and compare an explanation of its domain by the program's builders with an explanation produced by the program. The goal of the comparison is to gain insight on the relationship between "causal reasoning" and "deep reasoning."

---

*This research is supported by Air Force Office of Scientific Research grant AFOSR-82-0255, and grant NIHR GOAOE 82048-02 from the National Institute of Handicapped Research.

**This is not a claim that expert systems cannot perform interesting problem solving. Chandrasekaran and Mittal (Chandrasekaran, 1983) have pointed out how an expert system, for a particular reasoning situation, can fully incorporate the appropriate deep knowledge. However, it would not incorporate the deep knowledge for those situations that were not considered in its design.

2. Two Causal Explanations

These two explanations are taken from a paper by Patil, Szolovits, and Schwartz, which describes a program called ABEL (Patil, 1981), one the first programs to perform interesting causal reasoning. The first explanation is by the authors; the second by the ABEL program. The reader is forewarned that these explanations, although they concern the same domain, do not involve exactly the same phenomena.

> *Explanation #1.* "... 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. 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)... 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)... 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 the extracellular compartment (hypovolemia). Because the concentration of K and HCO_3 in lower GI fluid is higher than that 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." (Patil, 1981 - p. 894)

> *Explanation #2.* "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 extracellular fluid 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 extracellular fluid K, which causes low serum K." (Patil, 1981 - p. 898)

Both of these explanations have a causal story to tell, but in different ways and in different terms. The crucial difference is that the first quote makes use of our ***physical*** understanding about how the world works. It evokes a physical representation of the body and appeals to our understanding of how physical phenomena behave. The second quote is a different type of a physical explanation. While the second quote causally relates physical states, it does not express any physical relationships that let us understand the causal assertions in terms of some physical principle. Assertions like "low serum pH causes K shift out of cells" im-

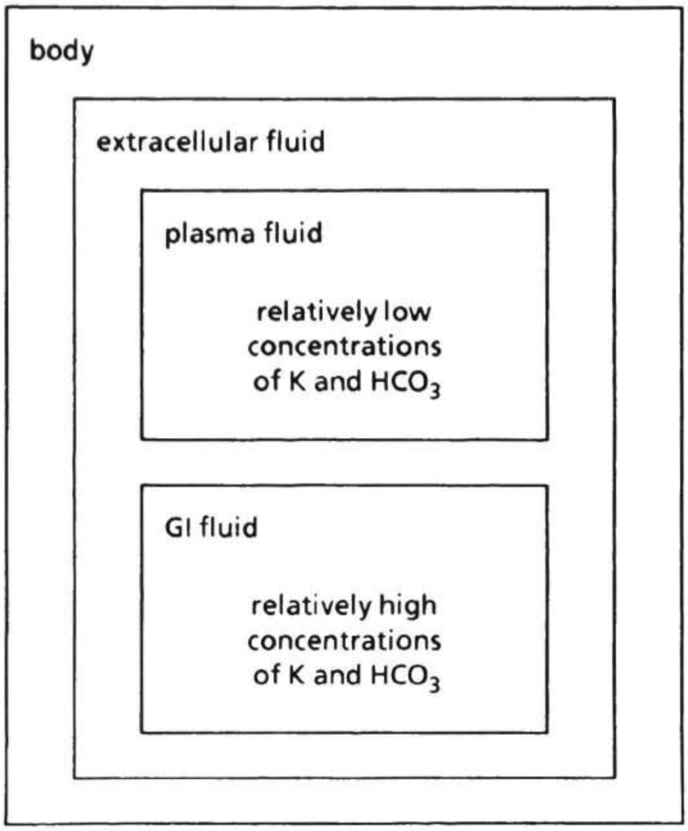

Figure 1: Representation of Patil, Szolovits, and Schwartz's Explanation

plicitly depend on the structure of the human body and how certain parts of the body behave. *With respect to physical phenomena, the first explanation is deep and the second explanation is compiled.*

3. An Analysis of the First Explanation

The first quote builds up the representation displayed in figure 1. (Na and Cl have been omitted for the purposes of this discussion.) The body can be thought of as having a container of extracellular fluid. The extracellular fluid compartment can be decomposed into a plasma fluid compartment and lower GI fluid compartment. Lower GI fluid has certain concentrations of HCO_3 and K, which happen to be greater than in plasma fluid. When the amount of lower GI fluid decreases (as happens in diarrhea), a corresponding amount of HCO_3 and K also decrease. It can be inferred that the total concentration of HCO_3 and K in extracellular fluid also decreases.

This representation lists the parts of the situation: fluid compartments, fluids, HCO_3, and K. It incorporates structural relationships between the parts, e.g., container, composed-of, and concentration, as well as behavioral information about them, e.g., fluid is something that can be contained, and can move. Also a fluid can be composed of other things, including HCO_3 and K in this case. The physical principle that this explanation appeals to is that when a certain amount of fluid moves, the fluid also takes what it is composed of along with it. With a little bit of qualitative (or quantitative) analysis about concentrations, it is not hard to determine how certain concentrations will increase or decrease depending on how fluid moves.

In general, reasoning about physical situations faces two problems: (1) changes in physical structure can change the overall behavior and properties of a situation, and (2) changes in a part's behavior can change the overall behavior and properties of a situation. So to perform deep reasoning about physical phenomena, representations need to express the structure and behavior of physical situations and their constituents, and reasoning processes need to be able to take this information into account. Much of the work in naive physics is aimed at reasoning about physical information such as behavioral properties of components, connections between components, and containment of substances (Hayes, 1985, deKleer, 1984, Forbus, 1984, Bylander, 1985). There has also been research on reasoning about how shape affects behavior (Forbus, 1983, Stanfill, 1983, Shoham, 1985).

4. An Analysis of the Second Explanation

The second quote is a description of the causal network illustrated in figure 2. The physical relationships that supports the causal network is not present in this explanation. For example, one part of the causal network is that loss of GI fluid contributes to low concentration of K in the extracellular fluid. However, this representation does not have structural and behavioral information such as "Extracellular fluid can be decomposed into plasma fluid and GI fluid."

Why is this additional information important? If the program only has causal networks such as in figure 2, the omitted physical information becomes a large set of assumptions that are implicitly encoded into the causal network. The result is that the robustness of the causal network depends on the likelihood that these physical assumptions are true.

For example, suppose that GI fluid in a particular person had a lower concentration of K than plasma fluid, then the causal network would be wrong. Since the causal network does not express where GI fluid sits in the body's structure and that GI fluid normally has a greater concentration of K than plasma fluid, the possibility that this information is wrong cannot be hypothesized and cannot be reasoned about. These are the same characteristics of compiled reasoning that typical expert systems have. Causal networks represent more information about associations between data and conclusions, but because they do not represent physical relationships, causal networks and their reasoning processes are also compiled.*

*Each causal link in ABEL has a "slot" for stating its assumptions. It is unclear what kind of information was being represented by the assumptions, and what reasoning processes could be performed on them. It is conceivable that a causal network could point to the information that supports it, but this additional information would be something different than causal networks.

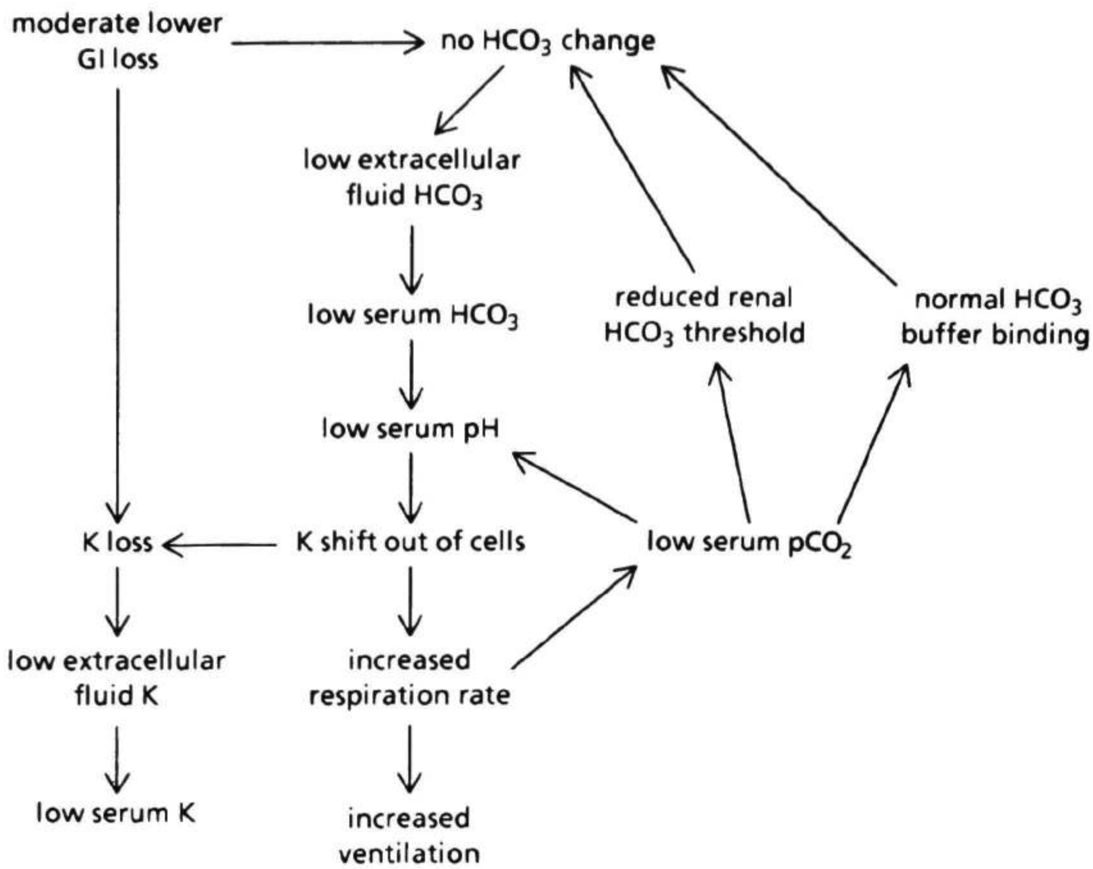

Figure 2: Representation of ABEL's Explanation

5. Some Misconceptions about Deep Reasoning

It might be claimed that representations like figure 1 are no better off than those like figure 2 because the information in figure 1 is a very qualitative representation, while figure 2 could relate physical states in more detail. This leads to the misconception that reasoning at a greater level of detail is "deeper" reasoning. This simply misses the point. Any representation worth considering can describe things at various levels of detail, but without representing physical relationships, certain kinds of reasoning processes can never be applied, no matter the level of detail.

Another misconception is that quantitative reasoning, such as solving or simulating differential equations, is deeper than qualitative reasoning. This is a misconception about the role of quantitative reasoning in reasoning about the world. A quantitative model is used when a situation can be mapped into it, and the results of applying the quantitative process can be interpreted in terms of the situation. To do this, there needs to be an understanding of what the situation is like, when the mapping is applicable, how to apply the mapping, and how to interpret the results. Each of these steps involve representation and reasoning (presumably qualitative) over and above the quantitative model. Quantitative reasoning supplements other reasoning processes; it does not substitute for them.

6. The General Nature of Deep Reasoning

On the basis of these examples, I propose the following definition of "deep":

> A representation is "deep" with respect to a class of phenomena iff the representation describes the properties and relationships by which the phenomena interact.
>
> A reasoning strategy is "deep" with respect to a class of phenomena iff the strategy reasons based on how the phenomena interact.

Relative to a certain class of phenomena, deep representations describe the properties and relationships that leads to interaction among these phenomena, and deep reasoning processes operate on this information. Because physical phenomena interact on the basis of physical structure and behavior, there need to be representational primitives whose meaning are structural and behavioral, and reasoning processes that can take this information into account.

References

T. Bylander and B. Chandrasekaran. (1985). Understanding Behavior Using Consolidation. *Proc. Ninth International Joint Conference on Artificial Intelligence.* Los Angeles.

B. Chandrasekaran and S. Mittal. (1983). Deep versus Compiled Knowledge Approaches to Diagnostic Problem-Solving. *International Journal of Man-Machine Studies, 19*, 425-436.

J. de Kleer and J. S. Brown. (1984). A Qualitative Physics Based on Confluences. *Artificial Intelligence, 24*, 7-83.

K. D. Forbus. (1983). Qualitative Reasoning about Space and Motion. In D. Gentner and A. Stevens (Eds.), *Mental Models.* Hillsdale, New Jersey: Lawrence Erlbaum.

K. D. Forbus. (1984). Qualitative Process Theory. *Artificial Intelligence, 24*, 85-168.

P. E. Hart. (1982). Directions for AI in the Eighties. *SIGART, 79*, 11-15.

P. J. Hayes. (1985). The Second Naive Physics Manifesto. In J. Hobbs and R. Moore (Eds.), *Formal Theories of the Commonsense World.* Norwood, New Jersey: Ablex.

D. Michie. (1982). High-road and Low-road Programs. *AI Magazine, 3*(1), 21-22.

R. S. Patil, P. Szolovits, and W. B. Schwartz. (1981). Causal Understanding of Patient Illness in Medical Diagnosis. *Proc. Seventh International Joint Conference on Artificial Intelligence.* Vancouver.

Y. Shoham. (1985). Naive Kinematics: One Aspect of Shape. *Proc. Ninth International Joint Conference on Artificial Intelligence.* Los Angeles.

C. Stanfill. (1983). The Decomposition of a Large Domain: Reasoning about Machines. *Proc. National Conference on Artifical Intelligence.* Washington, D.C..