Causal understanding in reasoning about the world

B. Chandrasekaran

Laboratory for AI Research
The Ohio State University
Columbus, OH 43210, USA

Abstract. In this paper I survey over a decade of work on how we understand how things work. Much of this work has been conducted in the context of reasoning about functions of devices. I briefly overview a representational framework called Functional Representation, and indicate how a device representation in this language can be used for simulation, diagnosis and design. I make remarks about the generality of this approach in terms of causal understanding.

1. The Problem: Understanding how things work

For a number of years several associates and I have been investigating what it means to understand "how things work." Understanding how things work — from "how clouds make rain," through "how cancer is formed," to "how nuclear plants generate power" — is a ubiquitous cognitive activity. Given some description of physical configuration, i.e., some description of components and how they are connected, and given some knowledge of the domain of interest — the underlying laws, the kinds of behaviors the components are capable of, the kinds of causality pathways that the connections provide, etc. — we can often predict aspects of the behavior of the physical configuration as a whole, and conversely, if we are given a behavior of the configuration, we can often construct an account of how that behavior comes about. Everyday reasoning as well as specialized scientific and engineering reasoning is full of examples where we need to put together a causal explanatory account of how some behavior of interest is produced. Such an account would explain the observed behavior by appealing to our understanding of the components, how they are connected, and the underlying scientific laws. This account would constitute our understanding of how the behavior was produced. Normally we expect a certain range of problem solving and inferential capabilities of someone who claims to understand how something works. We would expect that person to answer questions about changes in behavior as the structure of the configuration changes, to be able to identify the component responsible if the behavior of the configuration is not as expected, and so on.

The first research issue is the logical structure of this kind of causal explanation and the vocabulary of explanatory terms used in it. Once we have
identified the terms used in the explanation and how the explanation is organized, we have the basis for a knowledge representation language. For a given configuration and behaviors of interest, using this language, we can represent how the configuration produces the behavior. We can process this representation with appropriate inference machinery to answer the kinds of questions that we would expect a human understander to be able to answer. Developing a formal representation language to capture one's understanding of how something works, and techniques for using such representations for problem solving activities such as simulation, diagnosis and design have also been major concerns of our research.

An important underlying idea is that causal understanding of the world doesn't come simply in the form of "facts" about the world -- propositions or causal rules -- but in the form of causal packages which are organized from specific perspectives and which point to other causal packages. The work reviewed is based on the idea that these packages are basic units of comprehension and prediction. The Functional Representation (FR) theory is a proposal about the logical structure, representation and use of such causal packages. My associates and I have reported extensively on this research. The citations [1-3], [9-10], and [12-28] in the bibliography are an illustrative, but not a complete, listing. A review article [4] summarizes the research in some detail. For a detailed discussion of related work in the literature, the reader is referred to [4] as well. In this paper, I simply outline the approach and some of the issues, to provide a sense of the problem and our framework for it.

2. Functional Representation

Let us look at the examples that I started the paper with: "how clouds make rain," "How cancer is formed," and "how nuclear plants generate power." In this list, only the third is a device, that is, an artifact that is constructed such that it achieves an intended function. Clouds were not designed so that rain is produced (at least not in the scientific account of the world), nor is cancer an intended function of the human physiological system. Nevertheless, all of them have a structure of interconnection of physical components, and the explanation of the behavior of interest follows similar principles: we explain the behavior as arising out of the causal properties of the components, the physical laws and the interconnections. Much of the theory of causal explanation and understanding is thus applicable not only to designed artifacts, but to any configuration of causally efficacious components linked together in ways that they can affect each other. However, in much of what follows, we will talk of devices, in order to focus the discussion.

The Functional Representation framework is a proposal about the top-down representation for goal-directed, flexible reasoning that bridges abstraction levels. It was originally proposed in [19] for the causal processes that culminate in the achievement of device functions. (Some devices achieve their functions by means of causal processes, while the function of others is explained directly from their structure. We discuss this distinction later, but for now consider only devices