Benedikt **Löwe**, Wolfgang **Malzkorn**, Thoralf **Räsch** (*eds.*)
Foundations of the Formal Sciences II
Applications of Mathematical Logic in Philosophy and Linguistics
Papers of the conference held in Bonn, November 10-13, 2000

Bayesian Networks in Philosophy

Luc Bovens*

Department of Philosophy
University of Colorado at Boulder
CB 232
Boulder CO 80309
U.S.A.

E-mail: `bovens@spot.colorado.edu`

Stephan Hartmann

Fachbereich Philosophie
Universität Konstanz
78457 Konstanz
Germany

E-mail: `stephan.hartmann@uni-konstanz.edu`

Abstract. We provide a short introduction to the theory of Bayesian Networks and construct a Bayesian Network model of confirmation with an unreliable instrument. We indicate how this model can be extended to investigate the variety-of-evidence thesis, the Duhem-Quine thesis and calibration in philosophy of science and to give precise accounts of notions such as reliability and coherence in epistemology.

Received: April 6th, 2001;
In revised version: June 23rd, 2001;
Accepted by the editors: June 18th, 2001.
2000 *Mathematics Subject Classification.* **60A99 03A05** 68T99.

* The research was supported by a grant of the National Science Foundation, Science and Technology Studies (SES 00-80580), and grants of the TransCoop Program and of the Feodor Lynen Program of the Alexander von Humboldt Foundation. Stephan Hartmann also thanks Jim Lennox and the Center for Philosophy of Science at the University of Pittsburgh for their hospitality.

1 Introduction

There is a long philosophical tradition of addressing questions in philosophy of science and epistemology by means of the tools of Bayesian probability theory (*cf.* [Ear92,HowUrb89]). In the late '70s, an axiomatic approach to conditional independence was developed within a Bayesian framework. This approach in conjunction with developments in graph theory are the two pillars of the theory of Bayesian Networks, which is a theory of probabilistic reasoning in artificial intelligence. The theory has been very successful over the last two decades and has found a wide array of applications ranging from medical diagnosis to safety systems for hazardous industries.

Aside from some excellent work in the theory of causation (*cf.* [Pea00,SpiGlySch$_1$01]), philosophers have been sadly absent in reaping the fruits from these new developments in artificial intelligence. This is unfortunate, since there are some long-standing questions in philosophy of science and epistemology in which the route to progress has been blocked by a type of complexity that is precisely the type of complexity that Bayesian Networks are designed to deal with: questions in which there are multiple variables in play and the conditional independences between these variables can be clearly identified. Integrating Bayesian Networks into philosophical research leads to theoretical advances on long-standing questions in philosophy and has a potential for practical applications.

In the remainder of this contribution we will give a short introduction into the theory of Bayesian Networks (Section 2). We will then study one of the applications of Bayesian Networks in philosophy in more detail (Section 3) and finally discuss further possible applications and open problems (Section 4).

2 Bayesian Networks in Artificial Intelligence

Bayesian Networks are a powerful tool to deal with probability distributions over a large class of variables if certain (conditional) independence relations between these variables are known. A probability distribution over n binary propositional variables contains 2^n entries. The number of entries will grow exponentially with the number of variables. A Bayesian Network organizes these variables into a *Directed Acyclical*

Graph (DAG), which encodes a range of (conditional) independences. A DAG is a set of nodes and a set of arrows between the nodes under the constraint that one does not run into a cycle by following the direction of the arrows. Each node represents a propositional variable. Consider a node at the tail of an arrow and a node at the head of an arrow. We say that the node at the tail is the parent node of the node at the head and that the node at the head is the child node of the node at the tail. There is a certain heuristic that governs the construction of the graph: there is an arrow between two nodes iff the variable in the parent node has a direct influence on the variable in the child node. From DAG to Bayesian Network, one more step is required. A Bayesian Network contains a probability distribution for the variable in each root node (*i.e.*, in each unparented node), and a probability distribution for the variable in each child node, conditional on any combination of values of the variables in their parent nodes. When implemented on a computer, a Bayesian Network performs complex probabilistic calculations with one keystroke (*cf.* [Cow+99,Nea90,Pea88]).

3 An Example: Confirmation with an Unreliable Instrument

In philosophy of science, and more specifically in confirmation theory, there is a common idealization that the evidence in favor of a hypothesis is gathered by fully reliable instruments. What happens if we relax this idealization and permit that the evidence may have come from less than fully reliable (LTFR) instruments, as is common in scientific experimentation? Bayesian Networks proof to be useful to study situations like this. Consider a very simple scenario. Let there be a hypothesis, a (test) consequence of the hypothesis, a LTFR instrument and a report from the LTFR instrument to the effect that the consequence holds or not. To model this scenario, we need four propositional variables (written in italic script) and their values (written in roman script):

1. HYP can take on two values: HYP, *i.e.*, the hypothesis is true and $\overline{\mathrm{HYP}}$, *i.e.*, the hypothesis is false;
2. CON can take on two values: CON, *i.e.*, the consequence holds and $\overline{\mathrm{CON}}$, *i.e.*, the consequence does not hold;