
Heuristics and Mathematical Discovery: The Case of Bayesian Networks

DONALD GILLIES

1 Introduction

I will begin this paper by discussing some ideas to be found in two recent books on the philosophy of mathematics. These are (i) Carlo Cellucci's *Filosofia e matematica*, published by Laterza in 2002, and (ii) David Corfield's *Towards a Philosophy of Real Mathematics*, published by Oxford University Press in 2003. I will start with Cellucci's book.

In his book, Cellucci is highly critical of the traditional or foundational approach to the philosophy of mathematics, based on the attempt to justify mathematics. Instead he advocates what he calls the heuristic approach to the philosophy of mathematics. As he says (2002, p. viii):

According to the dominant point of view the principal problem in the philosophy of mathematics is that of the justification of mathematics. . . . In this book I maintain instead that the principal problem of reflection on mathematics is that of mathematical discovery. This problem includes the problem of justification . . .

I partly agree and partly disagree with this. It is certainly true that traditional philosophy of mathematics focussed exclusively on the problem of the justification of mathematics and neglected the problem of mathematical discovery. So I definitely think that philosophers of mathematics should now take up the problem of mathematical discovery and that interesting results are to be expected from investigating it. On the other hand, I do not think that the problem of discovery includes that of justification. So I hold that the problem of justification should remain on the agenda of philosophers of

mathematics, as a problem partly related to, but partly separate from that of discovery. For the purpose of this paper, however, I want to emphasize my agreement with Cellucci and to adopt his heuristic approach.

Now the obvious objection to the claim that philosophers should study the problem of mathematical discovery is that discoveries in mathematics depend on psychological factors such as insights of genius, the subjective intuitions of creative mathematicians and so on; and that, consequently, mathematical discovery cannot be given a systematic philosophical treatment. Cellucci strongly challenges this point of view in the following passage [2002, p. xvii]:

According to the dominant point of view mathematical discovery is an irrational process, which is not based on logic but rather on intuition. . . . In this book I maintain instead that mathematics is a rational activity at every moment, including the most important, discovery. Since antiquity many have recognised not only that mathematical discovery is a rational process, but also that a method exists for it, namely the analytic method. This method gave a great heuristic power to the ancient mathematicians for the solution of geometrical problems, and has had a decisive role in the new developments of mathematics and physics at the beginning of the modern era. In it logic plays an essential role in the discovery of hypotheses, though this is not logic understood in the restricted fashion . . . but in a wider fashion which includes also and above all non-deductive inferences.

Cellucci does not merely advocate a heuristic approach to the philosophy of mathematics, but actually makes a start with developing it, particularly in Chapters 30 to 38 of his book. Here he lists and illustrates quite a number of principles which he regards as fruitful for mathematical discovery. This investigation of Cellucci's does indeed call into question the claim that mathematical discovery is exclusively a matter of subjective intuitions and the like. There is however a point which can be regarded as doubtful. Cellucci makes clear in the passage just quoted that he believes that the principles underlying mathematical discovery are logical in character, so that there is, in effect, a logic of mathematical discovery. However, another point of view would be that there are indeed principles underlying mathematical discovery but that these principles are heuristics, or guides to discovery,

which are not logical in character. It is not an easy matter to decide between these two points of view, since it is not clear what we should regard as constituting logic. If there is to be a logic of mathematical discovery, then logic will certainly, as Cellucci stresses, have to be extended to include non-deductive inferences. Yet how far can we extend logic beyond its core of deductive inferences while still retaining something that is recognisably logic? Is there an inductive logic for example? And if so, what is its character? More generally what are the boundaries of logic? In the last section of this paper (Section 6) I will come back to this question and discuss some of the interesting ideas of Ladislav Kvasz on this subject. However for the moment, I will take the goal to be that of elucidating some of the heuristic principles involved in mathematical discovery, and leave aside the question of whether these principles should be regarded as logical in character.

Let me now turn to Corfield's new book. This contains a mass of interesting material ranging from automated theorem proving, through Bayesianism applied to mathematics, to a consideration of groupoids and higher-dimensional algebra. However, for the purposes of this present paper, I want to consider only one general methodological point which Corfield makes towards the beginning of his book. He points out that the mathematics considered by philosophers of mathematics tends to be almost exclusively the foundational mathematics of the period 1880–1930, and that, in particular, the mathematics of the last 70 years is largely ignored except perhaps, in some cases, for a consideration of further developments of foundationalist mathematics. As Corfield himself says [2003, p. 5]:

By far the larger part of activity in what goes by the name *philosophy of mathematics* is dead to what mathematicians think and have thought, aside from an unbalanced interest in the 'foundational' ideas of the 1880–1930 period, . . .

Corfield calls this attitude 'the foundationalist filter'. This filter removes from the attention of philosophers of mathematics any mathematics which is not foundationalist. Corfield thinks that philosophers of mathematics should remove this filter and consider mathematics which is not foundationalist. This could be some of the mathematics of the past, but Corfield recommends very strongly that philosophers of mathematics should take an interest in the non-foundationalist mathematics of the last seventy years which he thinks that they have hitherto largely ignored. As he says [2003, pp. 7–8]:

Straight away, from simple inductive considerations, it should strike us as implausible that mathematicians dealing with number, function and space have produced nothing of philosophical significance in the past seventy years in view of their record over the previous three centuries.

Corfield attempts in his book to redress the balance by considering from the philosophical point of view many developments in mathematics during the last seventy years.

That concludes my discussion of some of the ideas in the new books by Cellucci and Corfield. I will now explain how they have led to the plan for the present paper. Essentially I have taken from Cellucci the idea of studying the heuristics of mathematical discovery, and I will try to add to his treatment by considering an example of mathematical discovery different from the ones which he considers. Following the recommendations of Corfield, I have taken this example from the field of non-foundational mathematics in the last seventy years. The example in fact comes from my own favourite branch of mathematics: probability theory. Probability theory is usually considered by philosophers of science rather than philosophers of mathematics, and there are obvious reasons for this. Probability is closely connected to induction whose analysis, or in some cases denial, is a central issue in philosophy of science. Probabilities also appear in many scientific theories, notably quantum mechanics. But despite its interest for philosophers of science, probability theory is after all a branch of mathematics and an important one. So there may be some value in considering some of the general problems of the philosophy of mathematics in relation to probability theory.

Since I started studying probability theory in the 1960s, the most important development in the field has been, in my opinion, the discovery of Bayesian networks, which took place in the 1980s — fortunately well within the Corfield limit of seventy years. Many mathematical discoveries are of proofs of theorems, but some discoveries are of new mathematical concepts which give rise to new theories involving many theorems and having many uses in different areas. The most famous discovery of this type is perhaps the discovery of the group concept. The discovery of the concept of Bayesian network has this character. It has resulted in the development of an entirely new branch of probability theory which is now expounded in textbooks like Neapolitan 1990. None of the contents of Neapolitan 1990

would have appeared in a textbook of probability theory written before the 1980s. We have something here that is really new and that has also been applied with great success in a wide variety of different areas. We are thus dealing with a discovery of considerable importance and an analysis of the heuristics which led to this discovery may be not without some interest. In the next Section 2, I will give a brief historical account of how the discovery of Bayesian networks was made. This should also serve as an introduction to the concept for those who have not met it as yet. Then in Sections 3,4 and 5, I will state and analyse three heuristics which seem to me to have been involved in the discovery.

2 The Development of Artificial Intelligence and the Discovery of Bayesian Networks

One route which led to the discovery of Bayesian networks began with investigations into artificial intelligence (AI). This is the route which I will describe in what follows. The full story however is more complicated. There was another largely independent route which began with investigations into decision theory and which led to concepts not dissimilar from Bayesian networks. Another strand in the story is constituted by attempts to find economical ways of storing probability distributions in computers. The developments which I will describe, however, were largely self-contained and are suitable for analysis from the point of view of the heuristics involved. I will therefore leave the full account as the task for a more detailed history.

Research in AI began in the 1950s and many important ideas were developed by the pioneers. Then in the 1970s a breakthrough was produced by the creation of expert systems. The lead here was taken by the Stanford heuristic programming group, particularly Buchanan, Feigenbaum, and Shortliffe. What they discovered was that the key to success was to extract from an expert the knowledge he or she used to carry out a specialised task, and then code this knowledge into the computer. In this way they were able to produce 'expert systems' which performed specific tasks at the level of human experts. One of the most important of these early expert systems (MYCIN) was concerned with the diagnosis of blood infections. This system will now be briefly described, and it will then be shown that its implementation led to the problem of how to handle uncertainty in AI.

MYCIN was developed in the 1970s by Edward Shortliffe and his colleagues in collaboration with the infectious diseases group at the Stanford

medical school. The medical knowledge in the area was codified into rules of the form: IF such and such is observed, THEN likely conclusion is such and such. MYCIN's knowledge base comprised over 400 such rules which were obtained from medical experts. An example of such a rule will be given in a moment, but first it would be as well to present some evidence of MYCIN's success.

To test MYCIN's effectiveness a comparison was made in 1979 of its performance with that of nine human doctors. The program's final conclusions on ten real cases were compared with those of the human doctors, including the actual therapy administered. Eight other experts were then asked to rate the ten therapy recommendations and award a mark, without knowing which, if any, came from a computer. They were requested to give 1 for a therapy which they regarded as acceptable and 0 for an unacceptable therapy. Since there were eight experts and ten cases, the maximum possible mark was 80. The results were as follows [Jackson, 1986, p. 106]:

MYCIN	52	Actual therapy	46
Faculty-1	50	Faculty-4	44
Faculty-2	48	Resident	36
Inf dis fellow	48	Faculty-5	34
Faculty-3	46	Student	24

So MYCIN came first in the exam, though the difference between it and the top human experts was not significant.

Let us now examine one of MYCIN's rules. The following rule is given by Shortliffe and Buchanan [1975, p. 357]:

If: (1) the stain of the organism is gram positive (S_1), and
 (2) the morphology of the organism is coccus (S_2), and
 (3) the growth conformation of the organism is chains (S_3)
 Then: there is suggestive evidence (0.7) that the identity of the
 organism is streptococcus (H_1)

In symbols this could be written: If S_1 & S_2 & S_3 , then there is suggestive evidence p that H_1 , where $p = 0.7$. Here S_1, S_2, S_3 are the observations/symptoms, which support hypothesis H_1 to a particular degree. These rules were obtained from the medical experts. The numbers they contain such as 0.7 were also obtained from the experts. The expert was in

fact asked: “On a scale of 1 to 10, how much certainty do you affix to this conclusion?” The answer was then divided by 10.

At first sight it looks as if the figure 0.7 in the rule from MYCIN is an ordinary probability, but this is not the case, as Shortliffe and Buchanan make clear in the following passage [1975, p. 358]:

... this rule at first seems to say $P(H_1|S_1 \& S_2 \& S_3) = 0.7, \dots$
 Questioning of the expert gradually reveals, however, that despite the apparent similarity to a statement regarding a conditional probability, the number 0.7 differs significantly from a probability. The expert may well agree that $P(H_1|S_1 \& S_2 \& S_3) = 0.7$, but he becomes uneasy when he attempts to follow the logical conclusion that therefore $P(\text{not}.H_1|S_1 \& S_2 \& S_3) = 0.3$. The three observations are evidence (to degree 0.7) *in favor* of the conclusion that the organism is a streptococcus and should not be construed as evidence (to degree 0.3) *against* streptococcus.

Shortliffe and Buchanan used this observation to motivate the introduction of a *non-probabilistic* model of evidential strength. Their measure of evidential strength was called a *certainty factor*, and certainty factors neither obeyed the standard axioms of probability theory, the Kolmogorov axioms, nor combined like probabilities.

Certainty factors were criticized by those who favoured a probabilistic approach, cf. Adams [1976] and Heckerman [1986], and in fact the next expert system we will consider (PROSPECTOR) did move more in the direction of standard probability.

PROSPECTOR, an expert system for mineral exploration, was developed in the second half of the 1970s at the Stanford Research Institute. A good general account of the system is given by Gaschnig in his 1982. PROSPECTOR’s most important innovation was to represent knowledge by an *inference network* (or *net*). This is motivated by Duda *et al.* in their [1976, p. 1076] as follows:

A collection of rules about some specific subject area invariably uses the same pieces of evidence to imply several different hypotheses. It also frequently happens that several alternative pieces of evidence imply the same hypothesis. Furthermore, there are often chains of evidences and hypotheses. For these

reasons it is natural to represent a collection of rules as a graph structure or *inference net*.

A part of PROSPECTOR's inference network is shown in Figure 1.

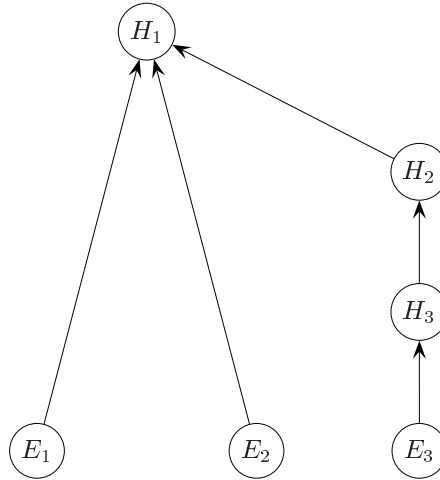


Figure 1.

- H_1 = There are massive sulfide deposits.
- H_2 = There are clay minerals.
- H_3 = There is a reduction process.
- E_1 = Barite is overlying sulfide.
- E_2 = Galena, sphalerite, or chalcopyrite fill cracks in rhyolite or dacite.
- E_3 = There are bleached rocks.

Evidence E_1 is taken as supporting hypothesis H_1 , and this is indicated by the arrow joining them in the inference network. Similarly E_2 supports hypothesis H_1 , while E_3 supports H_3 which supports H_2 which supports H_1 . Note how these rather complicated relations are simply and elegantly represented by the arrows of the network. Each inference arrow has a strength associated with it, and this obtained from the expert as in the case of MYCIN.

PROSPECTOR, however, differs from MYCIN in using subjective Bayesianism rather than certainty factors. This subjective Bayesianism is not entirely pure, since it is combined with fuzzy logic formulae, which were also used in MYCIN. This use of fuzzy logic tended to disappear in further developments.

In PROSPECTOR, Bayesianism is formulated using odds rather than probabilities. The odds on a hypothesis H [$O(H)$] are defined as follows:

$$O(H) = P(H)/P(\neg H)$$

Writing down Bayes theorem first for H and then for $\neg H$, we get

$$P(H|E) = P(E|H)P(H)/P(E)$$

$$P(\neg H|E) = P(E|\neg H)P(\neg H)/P(E)$$

So dividing gives

$$(1) \quad O(H|E) = \lambda(E)O(H)$$

where $\lambda(E)$ is the likelihood ratio $P(E|H)/P(E|\neg H)$. (1) is the odds and likelihood form of Bayes theorem, and it is used in PROSPECTOR to change the prior odds on H to the posterior odds given evidence E .

Let us now consider the problems which arise if we have several different pieces of evidence E_1, E_2, \dots, E_n say. We might in practice have to update using any subset of these pieces of evidence E_i, E_j, \dots, E_k say, where (i, j, \dots, k) is any subset of $(1, 2, \dots, n)$. If we use (1), this would involve having values of $\lambda(E_i \& E_j \& \dots \& E_k)$ for all subsets of $(1, 2, \dots, n)$. When we remember that, on this approach the values of λ are obtained from the domain experts, we can see that obtaining the requisite values of λ is scarcely possible. Clearly some simplifying assumptions are necessary to produce a workable system, and the designers of PROSPECTOR therefore made the following two conditional independence assumptions:

$$(2) \quad P(E_1, \dots, E_n|H) = P(E_1|H) \dots P(E_n|H)$$

$$(3) \quad P(E_1, \dots, E_n|\neg H) = P(E_1|\neg H) \dots P(E_n|\neg H)$$

Given these assumptions, the whole problem of updating with many pieces of evidence becomes simple, and, in fact,

$$O(H|E_1 \& \dots \& E_n) = \lambda_1 \lambda_2 \dots \lambda_n O(H) \text{ where } \lambda_i = \lambda(E_i)$$

The only remaining problem was whether the conditional independence assumptions (2) and (3) are plausible. The search for a justification of these assumptions led, as we shall see, to the modification of the concept of inference network, and the emergence of the concept of *Bayesian network*.

The concept of Bayesian network was introduced and developed by Pearl in a series of papers: Pearl [1982; 1985a; 1985b; 1986], Kim and Pearl [1983], and a book: Pearl [1988]. An important extension of the theory was carried out by Lauritzen and Spiegelhalter [1988], while Neapolitan's 1990 book gave a clear account of these new ideas and helped to promote the use of Bayesian networks in the *AI* community.

The actual term *Bayesian (or Bayes) network* was introduced in Pearl's [1985b] where it is defined as follows (p. 330):

Bayes Networks are directed acyclic graphs in which the nodes represent propositions (or variables), the arcs signify the existence of direct causal influences between the linked propositions, and the strengths of these influences are quantified by conditional probabilities.

This verbal account is illustrated by a diagram which is reproduced, with different lettering, in Figure 2.

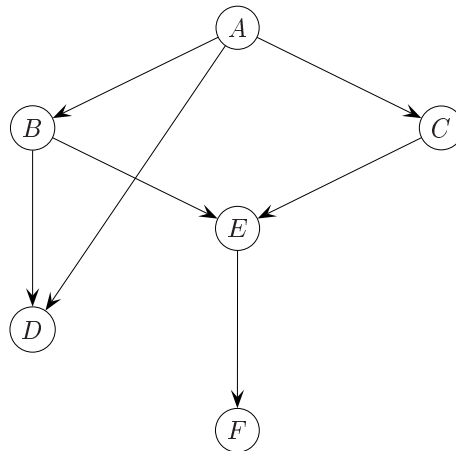


Figure 2.

If we compare the network of Figure 2 with that of Figure 1, two differences should be noted immediately. First of all the arrows in the inference network of Figure 1 represent a relation of support holding between e.g. E_3 and H_3 , while the arrows in the Bayesian network of Figure 2 represent causal influences, so that, e.g. the arrow joining A to B means that A causes B . Secondly, corresponding to the first difference, we can say that, in a certain sense, the arrows of a Bayesian network run in the opposite direction to those of an inference network. Pearl puts this point as follows [1986, pp. 253–4]:

... in many expert systems (e.g. MYCIN), ... rules point from evidence to hypothesis (e.g. if symptom, then disease), thus denoting a flow of mental inference. By contrast, the arrows in Bayes' networks point from causes to effects or from conditions to consequence, thus denoting a flow of constraints in the physical world.

This reversal of arrows from inference networks to Bayesian networks is illustrated in Figure 3, which shows one pair of nodes taken from the portion of PROSPECTOR's inference network shown in Figure 1.

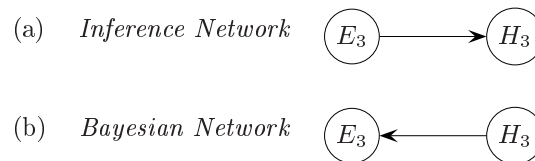


Figure 3. Reversal of Arrows

Here E_3 = There are bleached rocks, while H_3 = There is a reduction process. From the point of view of an inference network (a), we regard the evidence of bleached rocks as supporting the hypothesis that there is a reduction process, while, from the point of view of a Bayesian network (b), we regard there being a reduction process as a cause of there being bleached rocks. In his 1993, Pearl gives an account of his discovery of Bayesian networks, and says that one factor that led him to the idea was his consideration of the concept of influence diagrams introduced by Howard and Matheson (1984). Pearl decided to limit the influences specifically to

causal influences. Now Howard and Matheson were working on decision theory. So this is one point where the investigations of decision theory may have had an input into the investigations in artificial intelligence.

I will now make a few further points about Bayesian networks. If, in such a network, an arrow runs from node A to node B , then A is said to be a *parent* of B , and B a *child* of A . Children of A , children of children of A , and so on are known as *descendants* of A . If a node has no parents, it is called a *root*, so that in Figure 2, A is a root. In a Bayesian network, it is possible for a child to have several parents. Thus in Figure 2, E has parents B and C . If, however, every child has at most one parent, the network is called a tree. As in the earlier case of PROSPECTOR's inference networks, in order to make computation feasible, some conditional independence assumptions have to be made. For a Bayesian network, these are that a node is conditionally independent given its parents of the rest of the network except its descendants. I will call the conditional independence assumptions defining a Bayesian network *the generalised Markov condition*.

The nodes of a Bayesian network are random variables. Suppose we specify for each node the conditional probability distribution of that node given its parents, then it follows from the generalised Markov condition that these conditional probability distributions suffice to determine the joint distribution of all the variables of the network. This is an important result since it shows that Bayesian networks enable us to store joint distributions in a very concise way.

After introducing the concept of Bayesian network, Pearl developed algorithms which allow Bayesian updating to take place in such networks. If one of the variables which represents an observation is set to a particular value, the changes brought about by this new information in all the probabilities throughout the tree can be computed in an efficient manner. Pearl began in his 1982 by developing an updating algorithm for a simple form of network, namely a tree. He then extended his algorithm to more complicated networks. Kim and Pearl [1983] generalised from trees to Bayesian networks which are singly connected, i.e. there exists only one (undirected) path between any pair of nodes. Pearl in his 1986 tackled the further extension to Bayesian networks which are multiply connected. This problem was also investigated by Lauritzen and Spiegelhalter who in their 1988 solved it using the idea of reducing a multiply connected network to a tree of cliques. Their algorithm has been generally adopted by the *AI* community.

Let us now turn from these powerful mathematical developments to the consideration of a conceptual point. How exactly are causes and probabilities connected in Bayesian networks? In his original definition which he gave above, Pearl mentions both causes and probabilities. The arrows signify causal influences, while the nodes have associated with them probability distributions conditional on their parents. Pearl's idea about the link between causes and probabilities seems to have been that, if in a network the parents of every node represented the direct causes of that node, then the relevant conditional independence assumptions (the generalised Markov condition) would automatically be satisfied. As he says [1993, p. 52]:

Causal utterances such as “ X is a direct cause of Y ” were given a probabilistic interpretation as distinctive patterns of conditional independence relationships that can be verified empirically.

A suggested link between causality and conditional independence in fact goes back to Reichenbach [1956]. Reichenbach considers two events B and C say which are correlated. For example, in a travelling troupe of actors, B = the leading lady has a stomach upset, and C = the leading man has a stomach upset. We can explain such correlations, according to Reichenbach, by finding a common cause, namely that the leading lady and the leading man always have dinner together. The common stomach upsets occur when the food in the local restaurant has gone off. Denote ‘dining together’ by A . We then have the causal graph shown in Figure 4.

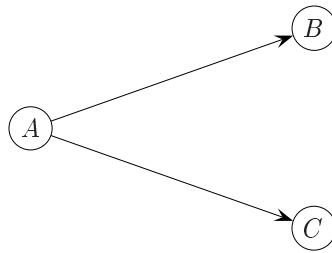


Figure 4.

Reichenbach then claimed that, conditional on A , B and C were no longer correlated but independent, i.e. $P(B\&C|A) = P(B|A)P(C|A)$. He also expressed this idea by saying that a common cause A screens one of its

effects B off from the other C . Reichenbach's causal fork is just a simple case of a Bayesian network. We can indeed apply his term 'screening off' to Bayesian networks by saying that in such networks, the parents of a node screen it off from all the other nodes in the network except its descendants.

We are now in a position to summarise the ingenious way in which Bayesian networks solved the problem of handling uncertainty in expert systems. In most of the domains considered, e.g. medical diagnosis, a domain expert is very familiar with the various causal factors operating. It should therefore be an easy matter to get him or her to provide a causal network. By the addition of probabilities this can be turned into a Bayesian network. In earlier systems such as MYCIN or PROSPECTOR, conditional independence assumptions were made for the purely *ad hoc* and pragmatic reason of allowing the updating to become possible. For Bayesian networks, however, the causal information obtained from the expert provides a justification for making a set of conditional independence assumptions (the generalised Markov condition) in the manner first suggested by Reichenbach. Moreover as Pearl, Lauritzen and Spiegelhalter have shown, the generalised Markov condition is sufficient to allow Bayesian updating to become computationally feasible. Everything fits together in a most satisfying manner. There is only one weak link in the chain. It turns out that it is possible to have a *bona fide* causal network in which the generalised Markov condition is not satisfied. I have discussed this last point with examples in Gillies [2002], but I will not pursue the development of the theory of Bayesian networks further here. I have given enough of the history of their discovery to enable us to examine in the next three sections the heuristic principles involved.

3 Heuristics Involved: (a) the Use of Philosophical Ideas

The first of the heuristics which I think was involved in the discovery of Bayesian networks was the use of philosophical ideas as a guide to the development of new mathematical concepts. The process which led to the discovery of Bayesian networks was begun by Shortliffe and Buchanan's attempt to construct a formal model for evidential support which could be implemented in their expert system: MYCIN. Shortliffe and Buchanan's key 1975 paper: 'A model of inexact reasoning in medicine' contains 33 references and no less than 14 of these (or over 42%) are to works in the philosophy of science concerned with the confirmation of scientific hypothe-

ses by evidence and related questions concerned with induction and the interpretation of probability. These 14 references are: Barker [1957], Carnap [1950], De Finetti [1972], Harré [1970], Helmer and Rescher [1960], Hempel [1965], Keynes [1921], Popper [1959], Ramsey [1931], Salmon [1966; 1973], Savage [1954], and Swinburne [1970; 1973]. In fact Buchanan and Shortliffe referred to nearly all the philosophers of science who were famous for their works on probability, induction and confirmation.

The main debate within philosophy of science about the confirmation of scientific hypotheses was at the time between the Bayesians and the anti-Bayesians. The Bayesians were divided in turn between the logical Bayesians such as Carnap and the subjective Bayesians such as De Finetti, Ramsey, and Savage. The leading anti-Bayesian was Popper. As we have seen, Shortliffe and Buchanan in constructing their formal model adopted an anti-Bayesian position. They were then immediately attacked by the Bayesians, and it was the members of the subjective Bayesian school, particularly Pearl, who succeeded in developing the successful theory of Bayesian networks. Recently Pearl has introduced some qualifications into his support for Bayesianism. His 2001 paper is significantly entitled: ‘Bayesianism and Causality, or Why I am only a Half-Bayesian’, but at the very beginning of the paper he reveals that he had no such doubts about the correctness of subjective Bayesianism when he introduced the concept of Bayesian network. This is what he says (2001, p. 19):

I turned Bayesian in 1971, as soon as I began reading Savage’s monograph *The Foundations of Statistical Inference* [Savage, 1962]. The arguments were unassailable: (i) It is plain silly to ignore what we know, (ii) It is natural and useful to cast what we know in the language of probabilities, and (iii) If our subjective probabilities are erroneous, their impact will get washed out in due time, as the number of observations increases.

In other words, Pearl adopted a particular philosophical position (subjective Bayesianism) and this acted as a heuristic guide to his mathematical work.

Pearl may also have been influenced by Reichenbach’s philosophical views on causality, for, as we saw earlier, Reichenbach’s notion of a causal fork anticipates the concept of Bayesian network in a simple case. However, the textual evidence here is not decisive. In his 1988, Pearl refers to Reichenbach’s 1949 book on probability, but this book did not contain a discussion

of causal forks which are introduced by Reichenbach in his 1956. In his 1988, Pearl refers to another philosophical work on causality, namely Suppes 1970 monograph: *A probabilistic theory of causation*. However, this work of Suppes does not mention Reichenbach's notion of causal fork.

I think this establishes beyond doubt that philosophical ideas were used as a heuristic guide in the discovery of the mathematical theory of Bayesian networks. But is this an unusual and exceptional case, or does philosophy quite often act as a heuristic in mathematical discovery? The idea that philosophy could be a heuristic guide in the natural sciences is in fact now quite familiar. It was introduced by Popper in 1934 as part of his critique of the Vienna Circle. While the Vienna Circle held that metaphysics was meaningless, Popper argued that metaphysics was not only often meaningful but could be helpful to science. Popper cited the example of atomism which began as a metaphysical theory and long remained one, but which was eventually turned into a scientific theory. Before Popper, Duhem had given many interesting examples of metaphysical ideas acting as heuristics for the development of science. More details of the work of Duhem and Popper on metaphysics in relation to the development of the natural sciences is to be found in Gillies [1993, Chapter 9, Sections 1–3, pp. 189–201].

Although the idea of philosophy acting as heuristic guide is familiar in the case of the natural sciences, there has been surprisingly little discussion of philosophy as a heuristic guide for mathematics. If we examine the history of mathematics, however, we can find many examples of philosophical ideas acting as heuristic guides to mathematical discoveries, though, at the same time, there are also many mathematical discoveries in which philosophy played no role. An obvious example of the influence of philosophy on mathematics is provided by the development of mathematical logic. Frege's revolution in the subject arose from his attempt to support the philosophical view that arithmetic was reducible to logic (see Gillies [1992] for details). The mathematical theory of probability too was strongly influenced by philosophy at earlier periods. The mathematical work of Thomas Bayes was designed to promote Bayesianism which, in turn, was devised in order to answer Hume's sceptical doubts about induction, as I have argued in Gillies [1987]. Another example from probability theory is provided by von Mises who in his development of his frequency theory of probability gave a philosophical analysis and definition of randomness. This definition appeared to have some flaws, and attempts to resolve this difficulty led to important

mathematical results by Wald and Church. Details are to be found in Gillies [2000, pp. 105–9].

These examples of the influence of philosophy on the development of mathematics taken from the history of mathematical logic and mathematical probability are not dissimilar from Popper's leading example of the influence of metaphysical ideas on the development of the natural sciences, namely: atomism. Before a precise experimentally testable theory of atomism could be developed it was necessary that atomism as a general view of the world should be elaborated in a less precise, metaphysical fashion. Now logic and probability form an integral part of philosophy because of their importance for epistemology. Some preliminary philosophical analysis of logic and probability was surely necessary to provide a jumping off point for a more precise mathematical theory of these concepts. This explains why in these cases, philosophical ideas were able to act as a guide to mathematical development.

Logic and probability, so I have argued, are part of the subject matter of both philosophy and mathematics. The same is true of the concept of infinity. This is the subject of philosophical disquisitions as well as of Cantor's theory of the transfinite. Indeed Cantor in developing his theory of the transfinite, made an intensive study of philosophical and also theological ideas about the infinite. Details of this are to be found in Dauben's 1979 life of Cantor, which is significantly entitled: *Georg Cantor. His Mathematics and Philosophy of the Infinite*.

As my final example of philosophical ideas as a heuristic for mathematical discovery, I want to consider a case which is rather different from those of logic, probability, and the infinite. This is Riemann's discovery of non-Euclidean geometry. I have argued that logic, probability and infinity are all subjects of both philosophy and mathematics and that a preliminary qualitative philosophical analysis of these notions was needed before more precise mathematical theories could be developed. Geometry, however, is not *per se* part of philosophy. However, since the time of Plato, geometry has been of great significance for Western philosophy as a prime example of excellent, indeed certain, knowledge, and therefore as a most important example for epistemology. In a famous passage from his 5th Meditation, Descartes says [1641, p. 181]:

...I clearly see that existence can no more be separated from the essence of God than can its having its three angles equal to

two right angles be separated from the essence of a [rectilinear] triangle, . . .

Now the proposition that the three angles of a rectilinear triangle are equal to two right angles is equivalent to Euclid's 5th postulate. So Descartes is claiming that the truth of Euclidean geometry is as certain as the existence of God. Of course by this he means that the truth of Euclidean geometry is completely certain. Later on Kant claimed that Euclidean geometry was synthetic *a priori*, implying that its truth was known with certainty independently of experience.

These well-known philosophical doctrines affirming the certain truth of Euclidean geometry certainly constituted an obstacle to the discovery of non-Euclidean geometry. Riemann presented his new ideas on non-Euclidean geometry in his famous lecture: 'Über die Hypothesen, welche der Geometrie zu Grunde liegen' (On the Hypotheses which lie at the Foundations of Geometry) delivered as a qualifying lecture (*Habilitationsvorlesung*) for the title of *Privatdozent* to the faculty at Göttingen on 10 June 1854. Riemann regarded it as necessary to begin his lecture with some philosophical analysis. This in effect constitutes an empiricist account of geometry which criticizes implicitly the Kantian view of Euclidean geometry as synthetic *a priori*. Riemann says that he has made use of some philosophical investigations of Herbart, an empiricist philosopher, and he remarks rather modestly [1854, p. 412]:

. . . I think myself the more entitled to ask considerate judgment inasmuch as I have had little practise in such matters of a philosophical nature, where the difficulty lies more in the concepts than in the construction . . .

In fact Riemann had studied theology before turning to mathematics and was by no means unfamiliar with philosophy. Here is a passage from his preliminary philosophical discussion [1854, p. 412]:

. . . the propositions of geometry are not derivable from general concepts of quantity . . . those properties by which space is distinguished from other conceivable triply extended magnitudes can be gathered only from experience. There arises from this the problem of searching out the simplest facts by which the metric relations of space can be determined, a problem which in nature

of things is not quite definite; for several systems of simple facts can be stated which would suffice for determining the metric relations of space; the most important for present purposes is that laid down for foundations by Euclid. These facts are, like all facts, not necessary but of a merely empirical certainty; they are hypotheses; one may therefore inquire into their probability, which is truly very great within the bounds of observation, and thereafter decide concerning the admissibility of protracting them outside the limits of observation, not only toward the immeasurably large, but also toward the immeasurably small.

The title of Riemann's lecture is itself an implicit criticism of Kant, since Riemann's point is that *hypotheses* (which may be empirically confirmed or disconfirmed) and *not a priori truths* lie at the foundation of geometry. This point is made more explicit in the passage just quoted, since Riemann claims that Euclidean assumptions are 'not necessary but of a merely empirical certainty', and that since 'they are hypotheses', 'one may therefore inquire into their probability'. Riemann regards this probability as very high for what falls within the bounds of observation, but still regards it as possible that Euclidean assumptions might break down 'toward the immeasurably large' or 'toward the immeasurably small'.

More details about Riemann's discovery of non-Euclidean geometry and his empiricism in the philosophy of geometry are to be found in Gillies, [1999, pp. 174–78]. For the purpose of the present paper, however, we can observe that Riemann's empiricist philosophy of geometry, which he developed with the help of Herbart's writings, played a very important role in his discovery of non-Euclidean geometry. It formed the basis of his criticism of the doctrine of Kant and other philosophers who held that Euclidean geometry was known with certainty *a priori*, and so opened up the way to introduce new forms of geometry which contradicted the Euclidean axioms.

In the present section I have given quite a number of examples of mathematical discoveries where philosophical ideas played an important heuristic role. However it should be stressed in conclusion that this is not a general law of mathematical development and there have been many mathematical discoveries in which philosophy played little or not part. An obvious example of such a discovery is the discovery of the concept of group in algebra. This arose from mathematical research into the problem of finding solutions

to polynomial equations in terms of radicals. Lagrange found a connection between this problem and permutations of the roots of the equation, and collections of such permutations constituted the first examples of the later concept of abstract group. Here we have a discovery emerging from internal mathematical investigations which did not have a connection with external philosophical questions.

4 Heuristics Involved: (b) New Practical Problems

The study of new practical problems often leads to mathematical discoveries. The discovery of Bayesian networks is a perfect example of this. As we have seen the discovery arose out of the problem of implementing expert systems for medicine, geological exploration and other areas. These expert systems involved handling uncertainty in a way which was rather different from previous applications of the probability calculus. The solution of this problem involved the development of new techniques involving a new mathematical concept.

Once again the pattern here exhibited in the discovery of Bayesian networks is to be found in many other discoveries in the history of mathematics. The mathematical theory of probability itself originated from the problem of calculating fair odds in gambling games. This was a very practical problem at the time, since gambling houses of that period offered odds which were empirically based. A mathematician who could calculate the correct odds stood a good chance of making money. New practical problems about the kinematics and mechanics of moving bodies such as cannonballs, planets or comets stimulated the development of calculus in the 17th century. In the previous section we saw how a philosophical research programme (the attempt to establish logicism in the philosophy of mathematics) led to the development of mathematical logic. However mathematical logic, though it originated in philosophy, was to find practical applications in the field of computer science. The new practical applications led to developments in mathematical logic itself, and, in particular, to the discovery of a quite new type of logic — non-monotonic logic. Some details about the discovery of non-monotonic logic are to be found in Gillies, [1996, pp. 72–75].

Although the investigation of new practical applications often leads to the discovery of new mathematical concepts, sometimes this is not the case because the existing body of mathematics is sufficient for handling the new application. An example of this is provided by Schrödinger's work in quan-

tum mechanics. Schrödinger's equation was a very important discovery in physics, but the equation turned out to be of a type which was familiar to mathematicians, and which could be solved by existing techniques. So, although Schrödinger was investigating some very new, indeed one might almost say, weirdly new phenomena, he was not led to formulating any new mathematical concepts.

Let me conclude this section by comparing the heuristic of using philosophical ideas with that of studying new practical problems. At first sight they seem to be quite distinct and rather opposed approaches. Philosophy, one might think, may be suitable for the abstract pure mathematician like Cantor who is far removed from any practical problem in the real world. Such a person would, it might be thought, be very different from the down to earth researcher working on practical problems. Of course this point of view is correct in Cantor's case, but one finds in many other cases, including our principal example of the discovery of Bayesian networks, that the study of practical problems and philosophical considerations, far from being opposed, actually go hand in hand. The reason for this is that philosophy need not be remote from the real world, but can be closely related to practical action, and, conversely, it may often be difficult to act in practice without some philosophical orientation.

5 Heuristics Involved: (c) Domain Interaction

The third heuristic which I will consider is what I will call: *domain interaction*. This occurs when two separate domains are brought together and partially unified. This process can often result in new discoveries and the growth of knowledge. Domain interaction has been studied by Emily Grosholz, who has emphasized its role in the development of mathematics. In this section, therefore, I will reverse the order used in the two preceding sections. I will first give a general account of Grosholz's ideas on domain interaction, including examples of where it has led to mathematical discovery. I will then show that domain interaction was an important heuristic principle involved in the discovery of Bayesian networks. In fact the example of Bayesian networks provides a striking vindication of Grosholz's ideas on this subject.

In a series of publications [1981; 1985; 1991; 1992], Grosholz has studied a number of cases in which knowledge (particularly mathematical knowledge) has advanced through the interaction of separate domains. In 1981,

she considers Logic and Arithmetic, in 1985 Logic and Topology, while in her 1992 she argues that Leibniz invented and developed the calculus by bringing together geometry, algebra, number theory, and mechanics. Her 1991 book shows that to a remarkable extent all Descartes' intellectual work can be seen as bringing together different domains. As she says [1991, pp. 2–3]:

... Cartesian domains ... can be understood as a novel amalgamation of formerly distinct or at least very incompletely unified domains: the *Geometry* brings together geometry and algebra, the *Principles* geometry and physics, the *Treatise of Man* physics and medical physiology, and the *Meditations* mechanical philosophy and scholastic theology.

This is an interesting passage since it shows that the heuristic of domain interaction is not limited to mathematics, but applies to other subjects as well. However the passage also gives one of the most famous examples of domain interaction in mathematics, namely the bringing together of geometry and algebra to create analytic geometry. Although Grosholz approves of Descartes's method of bringing together separate domains, she nonetheless criticizes the way in which he carries out this process. In her view the interaction of different domains is most fruitful, if, while interacting, they nonetheless retain some degree of autonomy. An attempt to reduce one domain to the other will generally inhibit fruitful developments. As she says [1991, p. 3]:

... the unification of domains contributes to the growth of knowledge when and because it exploits partially shared structure between domains that none the less retain their autonomy and distinctness. Revelation is impaired when domains are held too far apart, or assimilated too closely. But Descartes's way of constructing knowledge can produce both these unfortunate outcomes ...

According to Grosholz, Leibniz was more successful than Descartes in handling domain interaction (see [Grosholz, 1992]).

Another important concept which Grosholz uses in this connection is the concept of hybrid. As she says [2000, p. 82]:

Moreover, my examination of the growth of mathematical knowledge sheds important light on mathematical hybrids, objects which exist in the overlap of domains and provoke discovery in unexpected ways.

An important feature of such hybrids is that they exhibit a kind of instability or inconsistency. As Grosholz says [2000, p. 88]:

... the two domains as it were overlap, or are superimposed. At this overlap, objects are constituted which must simultaneously exhibit features of both domains; if the domains are truly heterogeneous, one must expect a kind of submerged heterogeneity in these objects. And in fact such hybrids often exhibit an instability or inconsistency that is however held in place or made tractable by the rational relatedness provided by the abstract structure that holds the domains together.

This instability or inconsistency is not seen by Grosholz as a defect, but rather as a potential stimulus to further growth and development.

Let me now show that these ideas apply very well to the example of Bayesian networks. In fact, Bayesian networks involved two instances of domain interaction. To begin with, Bayesian networks put together the domains of probability theory and graph theory which had previously been largely separate. Secondly, however, Bayesian networks put together the domains of probability theory and causality. In fact there had earlier been the beginning of an attempt in the philosophy of science community to connect these domains. Suppes [1970] *A probabilistic theory of causation* is a leading example of this trend. However the development of the concept of Bayesian network was a notable advance in linking the two domains. In a Bayesian network, an arrow joining two nodes A and B usually indicates that there is a causal connection between A and B . Furthermore each node in a Bayesian network has a conditional probability distribution associated with it. Thus causality and probability are brought together.

However this hybrid of causality and probability is by no means unproblematic. Pearl originally hoped that the causal connections between the nodes of a network would justify adopting the generalised Markov condition for the probability distributions. However it emerged that there can be genuine causal graphs for which the generalised Markov condition does not

hold. More details about this are to be found in Gillies [2002]. So the relations between causality and probability in a Bayesian network turn out to be highly problematic. Bayesian networks thus fit very well the descriptions which Grosholz gives of other mathematical hybrids. As she says [2000, p. 88]: ‘...one must expect a kind of submerged heterogeneity in these objects. And in fact such hybrids often exhibit an instability or inconsistency ...’

6 Heuristics of Mathematical Discovery versus Logic of Mathematical Discovery

Having given my example of a recent mathematical discovery and attempted to analyse the heuristics which were involved, I now want to raise the general question of whether such heuristics constitute a kind of generalised logic so that one could speak of a logic of mathematical discovery, or whether heuristic principles are not logical in character. This question is by no means an easy one. The core of logic is obviously standard deductive logic. However, it has often been suggested that logic could be extended to include not just deductive inferences but ampliative inferences of various kinds. For example, many philosophers of science have supported the idea of an inductive logic. Might heuristic principles constitute an extension of logic of which inductive logic is just a part?

I will begin my examination of this problem by considering an interesting related discussion by Ladislav Kvasz in his 2002. Kvasz here deals not with the relation between heuristics and logic, but with the obviously closely connected question of the relation between dialectics and logic. Kvasz in his paper criticizes dialecticians, but under that heading he includes not just the paradigm dialecticians (Hegel and the Marxists), but also Popper and Lakatos whom he regards as also dialecticians, notwithstanding their striking disagreements with Hegel and the Marxists. What is common to all these thinkers according to Kvasz is that they regard dialectics as a branch of logic. As Kvasz himself says [2002, p. 211]:

Usually, the dialecticians believe that the pattern of the development of knowledge is of a logical nature (Hegel’s idea of dialectical logic, Popper’s logic of scientific discovery, or Lakatos’ logic of mathematical discovery), which creates a tension between the development of knowledge and formal logic.

This ‘confusion of dialectic with logic’ [Kvasz, 2002, p. 211] is responsible, according to Kvasz, for grave failings in Hegel on the one hand and in Popper and Lakatos on the other. However, these failings are different in the two cases. The problem with Hegel and the Hegelians is that they regard their dialectical logic as being in competition with and superior to ordinary deductive logic. Hegelians therefore reject ordinary deductive logic which Kvasz thinks is a mistake. Popper and Lakatos did not give up ordinary deductive logic, but their attempt to reconcile it with the development of knowledge led to them confining their analyses to cases in which the conceptual changes in the growth of knowledge are relatively small. This is how Kvasz puts this argument [2002, p. 229]:

Both solutions to the dialectician’s conflict between logic and evolution of knowledge are unsatisfactory. Philosophers who follow Hegel, in the attempt to replace classical logic by some new dialectical one, were unable to offer anything comparable to the successive formal logic, and thus their research programme degenerated. On the other hand, dialecticians like Popper or Lakatos, who were not prepared to sacrifice logic, and thought that logical consistency is crucial to rational discourse, were forced to give up evolution. The fact that Lakatos was unable to reconstruct any deeper conceptual change in history of mathematics or physics is not accidental. As a dialectician, he conceived evolution to be in conflict with logic, but as Popper’s disciple he was not prepared to give up logic. Thus he omitted some of the most interesting moments in the history of mathematics. If he had tried to reconstruct them, he would have been forced to violate logic. Therefore he reconstructed only those changes, in which relatively small conceptual changes occur. . . .

The one extreme is *dialectical logic* (of Hegel and Marxism), which for the sake of evolution sacrifices logic. The other extreme is *logical dialectic* (of Popper or Lakatos), which for the sake of logic sacrifices evolution.

Kvasz argues for this general position by giving an analysis of some changes in the development of mathematics which he regards as being too large to be compatible with formal logic. These changes all involve a change in the form of the language used. Following Wittgenstein in the *Tractatus*,

Kvasz regards any language (L say) as having a form which is not expressible in the language. We can however incorporate the form of the language L into L thereby creating a new language L' say. A simple example of this process occurred in the transition between the language of perspective used by Renaissance painters and the language of projective geometry created by Desargues. As Kvasz says [2002, p. 221]:

... the centre of projection represents, in an abstract form, the eye of the painter from Dürer's drawing. For Desargues, . . . , the point of view is explicitly incorporated into language.

In fact Kvasz analyses a whole series of examples of changes in mathematics which follow this pattern in his papers [1998] and [2000].

We can now see clearly why formal logic is inadequate to deal with such changes. Any system of formal logic presupposes a language L in which it is formulated, and this language is held constant when the deductions are being made. If therefore we make a fundamental change in the character of the language, altering it from L to L' , this change cannot be captured using formal logic. On the other hand we can apply formal logic without any problems either within L or within L' so that there is no need to abandon formal logic altogether as the Hegelians deem to be necessary. Formal logic has only to be given up temporarily in the course of a large change involving a considerable alteration in the form of the language used. This then is a brief summary of Kvasz's position. Let us now see if we can apply it to our problem about heuristics and logic.

It is clear that the discovery of Bayesian networks involved the creation of a new language formed through the synthesis of the languages of earlier probability theory and graph theory. The language of Bayesian networks with its network diagrams has an iconic character which is not to be found in earlier probability theory. As this is a major change in language, then we can use Kvasz's argument to conclude that the discovery of Bayesian networks is a transition which cannot be explicated logically so that the heuristics involved are not logical in character. Indeed we can generalise to say that many applications of the domain interaction heuristic take us outside logic. Cartesian geometry, for example, has its own specific language which differs both from the language of classical Euclidean geometry and from that of algebra unrelated to geometry. Similarly calculus introduced new symbolisms such as dy/dx or \ddot{y} which made the language of calculus radically different from preceding mathematical languages.

Changes of this magnitude cannot, according to Kvasz's argument, be explicated using logic. However, Kvasz's analysis also indicates that some changes might be logically explicated. These would be smaller changes. An example might be the discovery of the proof of a mathematical conjecture where both the conjecture and the subsequent proof are formulated within a well-defined mathematical system, which is not changed in the process of discovery. There is no reason why the heuristics of discoveries of this sort might not be explicated in a way that could be described as logical.

These conclusions are supported by another approach to the problem. This approach relies on the connection between logic and mechanisation. If mathematical proofs are translated into formal logic then the validity of each step can be checked mechanically by means of a computer. The discovery of the proof, however, can be left entirely in the hands of human mathematicians. The development of automated theorem proving, and of non-monotonic logic programming languages such as PROLOG has carried the mechanisation process one stage further by mechanising the construction of proofs. In this respect, then, it goes beyond Fregean formal logic.

I have suggested (in [Gillies, 1996, p. 85]) a way of characterising this new kind of logic which has been introduced by investigations into artificial intelligence. The formula proposed is

$$\text{Logic} = \text{Inference} + \text{Control}$$

When we employ Logic, we start with a set of assumptions from which we want to derive some conclusions. To carry out these derivations we need a set of rules of inference (the Inference component). If the derivation is carried out by a trained mathematician, then he or she will rely on intuition to decide which rule of inference to use at a particular stage in order to carry out the derivation. If, however, we are trying to program a computer to carry out the derivation, then we will have to give the computer guidance as to which assumptions to choose and which rules of inference to apply. This guidance constitutes the Control component. Thus the Control component might specify at each stage of the derivation, which of the assumptions should be employed, and which of the rules of inference should be applied to these assumptions or to previously obtained results. More generally, the Control component would be designed to help in the construction of a derivation or proof of a conclusion.

I further suggested (in [Gillies, 1996, Ch. 5, pp. 98–112]) that this formula enables one to defend the possibility of an inductive logic. The development of machine learning has led to the formulation of inductive rules of inference, while confirmation theory constitutes the control component. In the case of automated theorem proving, the heuristics used could be formulated as part of the control component, and could then, using the formula above, be considered part of a logic of mathematical discovery. This criterion suggests therefore that a heuristic can be considered a logical principle if it can be formulated with precision and incorporated into a successful computer system for automated theorem proving. This criterion implies the Kvasz criterion since, at least as things stand at present, any automated theorem proving system has to operate within a fixed formal language specified at the beginning.

The kind of heuristics which I have considered in this paper (use of philosophical ideas, consideration of new practical problems, and domain interaction) are too vague in character to be suitable for precise formulation and implementation in programs for automated theorem proving. I would therefore argue that they are not logical in character.

A critic might say at this point that heuristics which are not precise enough to become part of logic are unlikely to provide much useful guidance. However such a comment would be unfair. The somewhat vague non-logical heuristics considered in this paper are certainly not precise enough to guide a computer in the execution of a program. However they are precise enough to suggest strategies for a human mathematician carrying out mathematical research. Moreover the kind of strategies suggested by the three heuristics given are rather different from those commonly adopted by human mathematical researchers. It is all too common for research mathematicians to become exclusively absorbed in their own small field and to devote themselves to reading only the literature of that specialty. The analysis given of the discovery of Bayesian networks suggests a quite different sort of research strategy, one which would involve a broader more interdisciplinary approach, with the study of some philosophy, an interest in areas which might require new techniques for successful practical applications, and a knowledge of several branches of mathematics which could be brought together for ‘domain interaction’. The example even suggests some more specific recommendations. Mathematicians preparing for research in the area of probability and statistics would normally take a master’s degree

in this speciality. The discovery of Bayesian networks suggests that it might be worth including a course on the philosophical and foundational aspects of probability and statistics as part of the preparation of the future researcher. Yet this is rarely if ever done. In effect the heuristics considered in this paper, though not precise enough to guide a computer, do definitely suggest strategies for humans who want to carry out research in mathematics.

BIBLIOGRAPHY

- [Adams, 1976] J.B. Adams. A probability model of medical reasoning and the MYCIN model. *Mathematical Biosciences*, **32**, pp. 177–86, 1976.
- [Barker, 1957] S.F. Barker. *Induction and Hypothesis: A Study in the Logic of Confirmation*. Ithaca, New York, Cornell University Press, 1957.
- [Buchanan:Shortliffe, 1984] B.G. Buchanan and E.H. Shortliffe (eds). *Rule-based expert systems: the MYCIN experiments of the Stanford heuristic programming project*. Reading, Mass, Addison-Wesley, 1984.
- [Carnap, 1950] R. Carnap. The two concepts of probability. In *Logical Foundations of Probability*, Chicago, University of Chicago Press, pp. 19–51, 1950.
- [Cellucci, 2002] C. Cellucci. *Filosofia e matematica*. Roma-Bari, Laterza, 2002.
- [Corfield, 2003] D. Corfield. *Towards a Philosophy of Real Mathematics*. Cambridge, Cambridge University Press, 2003.
- [Dauben, 1979] J.W. Dauben. *Georg Cantor. His Mathematics and Philosophy of the Infinite*. Cambridge, Massachusetts and London, England, Harvard University Press, 1979.
- [Descartes, 1641] R. Descartes. *Meditations on First Philosophy*, 1641. English translation in *The Philosophical Works of Descartes*, Volume 1, translated by Elizabeth S. Haldane and G.R.T. Ross, Cambridge, Cambridge University Press, pp. 131–99, 1970.
- [De Finetti, 1972] B. De Finetti. *Probability, Induction, and Statistics — the Art of Guessing*, New York, Wiley, 1972.
- [Duda, Hart and Nilsson, 1976] R.O. Duda, P.E. Hart and N.J. Nilsson. Subjective Bayesian methods of rule-based inference systems. In *Proceedings of the National Computer Conference (AFIPS)*, **45**, pp. 1075–82, 1976.
- [Gaschnig, 1982] J. Gaschnig. Prospector: an expert system for mineral exploration. In Donald Michie (ed.) *Introductory readings in expert systems*, New York, Gordon and Breach, pp. 47–64, 1982.
- [Gillies, 1987] D.A. Gillies. Was Bayes a Bayesian? *Historia Mathematica*, **14**, pp. 325–46, 1987.
- [Gillies, 1992] D.A. Gillies. The Fregean Revolution in Logic. In Donald Gillies (ed.) *Revolutions in Mathematics*, Oxford, Oxford University Press, pp. 265–305, 1992.
- [Gillies, 1993] D.A. Gillies. *Philosophy of Science in the Twentieth Century*. Oxford UK & Cambridge USA, Blackwell, 1993.

- [Gillies, 1996] D.A. Gillies. *Artificial Intelligence and Scientific Method*. Oxford, Oxford University Press, 1996.
- [Gillies, 1999] D.A. Gillies. German Philosophy of Mathematics from Gauss to Hilbert. In Anthony O'Hear (ed.), *German Philosophy since Kant*, Cambridge, Cambridge University Press, pp. 167–92, 1999.
- [Gillies, 2000] D.A. Gillies. *Philosophical Theories of Probability*, London and New York, Routledge.
- [Gillies, 2002] D.A. Gillies. Causality, Propensity, and Bayesian Networks. *Synthese*, **132**, pp. 63–88, 2002.
- [Grosholz, 1981] E.R. Grosholz. Wittgenstein and the Correlation of Logic and Arithmetic. *Ratio*, **23**, pp. 31–42, 1981.
- [Grosholz, 1985] E.R. Grosholz. Two Episodes in the Unification of Logic and Topology. *British Journal for the Philosophy of Science*, **36**, pp. 147–57, 1985.
- [Grosholz, 1991] E.R. Grosholz. *Cartesian Method and the Problem of Reduction*. Oxford, Oxford University Press, 1991.
- [Grosholz, 1992] E.R. Grosholz. Was Leibniz a Mathematical Revolutionary? In Donald Gillies (ed.) *Revolutions in Mathematics*, Oxford, Oxford University Press, pp. 117–133, 1992.
- [Grosholz, 2000] E.R. Grosholz. The Partial Unification of Domains, Hybrids, and the Growth of Mathematical Knowledge. In Emily Grosholz and Herbert Breger (eds.) *The Growth of Mathematical Knowledge*, Dordrecht, Boston, London, Kluwer, pp. 81–91, 2000.
- [Harré, 1970] R. Harré. Probability and confirmation. In *The Principles of Scientific Thinking*, Chicago, University of Chicago Press, 1970.
- [Heckerman, 1986] D. Heckerman. Probabilistic interpretations for MYCIN's certainty factors. In L.N. Kanal & J.F. Lemmer (eds.) *Uncertainty in Artificial Intelligence*, Amsterdam, North-Holland, pp. 167–96, 1986.
- [Helmer and Rescher, 1960] O. Helmer and N. Rescher. On the epistemology of the inexact sciences. *Project Rand R-353*, February 1960.
- [Hempel, 1965] C.G. Hempel. Studies in the logic of confirmation. In *Aspects of Scientific Explanation and other Essays in the Philosophy of Science*, New York, The Free Press, pp. 3–51, 1965.
- [Howard and Matheson, 1984] R.A. Howard and J.E. Matheson. Influence diagrams. In R.A. Howard and J.E. Matheson (eds.) *The principles and applications of decision analysis*, Vol. 2, Menlo Park, CA, Strategic Decisions Group, pp. 721–62, 1984.
- [Jackson, 1986] P. Jackson. *Introduction to expert systems*. Wokingham, England, Addison-Wesley, 1986.
- [Keynes, 1921] J.M. Keynes. *A Treatise on Probability* 1921. New York, Harper and Row, 1962.
- [Kim and Pearl, 1983] J.H. Kim and J. Pearl. A computational model for combined causal and diagnostic reasoning in inference systems. *Proceedings of the 8th International Joint Conference on AI (IJCAI-85)*, pp. 190–3, 1983.

- [Kvasz, 1998] L. Kvasz. History of Geometry and the Development of the Form of its Language. *Synthese*, **116**, pp. 141–86, 1998.
- [Kvasz, 2000] L. Kvasz. Changes of Language in the Development of Mathematics. *Philosophia Mathematica*, **8**, pp. 47–83, 2000.
- [Kvasz, 2002] L. Kvasz. Lakatos' Methodology Between Logic and Dialectic. In George Kamps, Ladislav Kvasz and Michael Stöltzner (eds.) *Appraising Lakatos. Mathematics, Methodology and the Man*, Vienna Circle Library, Dordrecht, Boston, London, Kluwer, pp. 211–41, 2002.
- [Lauritzen and Spiegelhalter, 1988] S.L. Lauritzen and D.J. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems (with discussion). *Journal of the Royal Statistical Society B*, **50**, pp. 157–224, 1988.
- [Neapolitan, 1990] R.E. Neapolitan. *Probabilistic reasoning in expert systems. Theory and algorithms*. New York, John Wiley, 1990.
- [Ng and Abramson, 1990] K. Ng and B. Abramson. Uncertainty management in expert systems. *IEEE Expert*, **5**, pp. 29–48, 1990.
- [Pearl, 1982] J. Pearl. Reverend Bayes on inference engines: a distributed hierarchical approach. *Proceedings of the National conference on AI (ASSI-82)*, pp. 133–6, 1982.
- [Pearl, 1985a] J. Pearl. How to do with probabilities what people say you can't. *Proceedings of the Second IEEE Conference on AI Applications*, Miami, Fl., pp. 6–12, 1985.
- [Pearl, 1985b] J. Pearl. Bayesian networks: a model of self-activated memory for evidential reasoning. *Proceedings of the Cognitive Science Society*, Ablex, pp. 329–34, 1985.
- [Pearl, 1986] J. Pearl. Fusion, propagation and structuring in belief networks. *Artificial Intelligence*, **29**, pp. 241–88, 1986.
- [Pearl, 1988] J. Pearl. *Probabilistic reasoning in intelligent systems. Networks of plausible inference*. San Mateo, California, Morgan Kaufmann, 1988.
- [Pearl, 1993] J. Pearl. Belief networks revisited. *Artificial Intelligence*, **59**, pp. 49–56, 1993.
- [Pearl, 2001] J. Pearl. Bayesianism and causality, or, why I am only a half-Bayesian. In David Corfield and Jon Williamson (eds.) *Foundations of Bayesianism*, Dordrecht, Boston, London, Kluwer, pp. 19–36, 2001.
- [Popper, 1959] K.R. Popper. Corroboration, the weight of evidence, and statistical tests. In *The Logic of Scientific Discovery*, New York, Scientific Editions, pp. 387–419, 1959.
- [Ramsey, 1931] F.P. Ramsey. *The Foundations of Mathematics and other Logical Essays*. London, Kegan Paul, 1931.
- [Reichenbach, 1949] H. Reichenbach. *Theory of Probability*. Berkeley, University of California Press, 1949.
- [Reichenbach, 1956] H. Reichenbach. *The Direction of Time*. Berkeley, University of California Press, 1956.

- [Riemann, 1854] B. Riemann. On the Hypotheses which lie at the Foundations of Geometry, 1854. English translation in David E. Smith (ed.), *A Source Book of Mathematics*, Volume Two, New York, Dover Publications, pp. 411–25, 1959.
- [Salmon, 1966] W.C. Salmon. *The Foundations of Scientific Inference*. Pittsburgh, Pennsylvania, University of Pittsburgh Press, 1966.
- [Salmon, 1973] W.C. Salmon. Confirmation. *Scientific American*, May 1973, pp. 75–83, 1973.
- [Savage, 1954] L.J. Savage. *The Foundations of Statistics*. New York, Wiley, 1954.
- [Savage, 1962] L.J. Savage (ed). *The Foundations of Statistical Inference*. London, Methuen, 1962.
- [Shortliffe and Buchanan, 1975] E.H. Shortliffe and B.G. Buchanan. A model of inexact reasoning in medicine. *Mathematical Biosciences*, **33**, pp. 351–79, 1975.
- [Suppes, 1970] P. Suppes. *A probabilistic theory of causation*. Amsterdam, North Holland, 1970.
- [Swinburne, 1970] R.G. Swinburne. Choosing between confirmation theories, *Philosophy of Science*, **37**, pp. 602–13, 1970.
- [Swinburne, 1973] R.G. Swinburne. *An Introduction to Confirmation Theory*. London, Methuen, 1973.