Artificial Intelligence and Its Methodological Implications

Stephan Hartmann[1]

1. Introduction

Donald Gillies is one of the pioneers in the philosophical analysis of artificial intelligence (AI). In his recent book, Gillies (1996) not only makes a new and rapidly developing field of science accessible to philosophers; he also introduces philosophical topics relevant to researchers in AI and thereby helps establish a dialogue between the two disciplines. His book clearly and convincingly demonstrates the fruitful interplay between AI and philosophy of science.

The present paper continues the discussion in the book by focusing on methodological issues. I agree with a lot of what Gillies has to say, but commentaries tend to be more interesting if the commentator focuses on controversial issues. So I will discuss three topics on which our views differ. First I reconsider the issue of the relevance of scientific developments for a philosophical debate (Sec. 2). More specifically, I will address the following question: Do we have to change our philosophical views in the light of the development of new scientific *tools* (such as Bayesian Networks)? Contrary to Gillies, I will argue that the new developments in AI have no impact on what Gillies calls the Bayesian controversy. Second I discuss two of Gillies' concrete methodological recommendations. Gillies suggests that the use of objective probabilities and the construction of a Bayesian Network model should be supplemented by a Popperian testing methodology. I will argue that the process of finding the value of an objective probability is not an instance of a test (Sec. 3), and that there are alternative (and perhaps more efficient) strategies for the construction of a Bayesian Network model (Sec. 4). Third I am

concerned with the scope of Bayesianism. Gillies gives an argument to the effect that Bayesianism does not apply to general hypotheses. I will show that this argument does not go through (Sec. 5).

2. The Philosophical Implications of AI

There is no doubt that science has an effect on our philosophical views. Some scientific developments have inspired work in metaphysics (e.g. quantum statistics gave rise to a reconsideration of the concept of individuality), and other developments forced philosophers to rethink their positions (e.g. about absolute space). The situation in AI, however, is different. Here, a new scientific tool is developed (i.e. Bayesian Networks) and Gillies claims that this development has methodological implications for the Bayesian controversy. This controversy is concerned with two issues. The first is about the use of probabilities (as opposed to other ways of representing uncertainty in AI). The second is the interpretation of probability. Let me examine both issues in turn.

A survey of the literature reveals that the majority of approaches to AI use probabilities, while only a minority deals with alternatives such as the Dempster-Shafer theory, ranking functions, or Popper functions. This raises the question of whether the calculus of probabilities is appropriate to handling uncertainty. Probabilities are very popular in scientific practice. But can we conclude from this that probabilities are *better* than alternative options; i.e. can we conclude from this that the right way to represent uncertainty is by using probabilities? I do not think so. The popularity of Bayesian Networks in the AI community can be understood by noticing two points: First, there is an extremely fruitful research program centered around Bayesian Networks which gives researchers in the field a chance to make their own contributions to the development of the theory. Second, Bayesian Networks are very easy to apply. In fact, they are much easier to apply than any of the other available tools, which is partially rooted in the fact that the alternatives to Bayesian Networks are simply not developed well enough yet. These pragmatic factors play an important role when a decision has to be made as to which tool we

choose to work with. However, these pragmatic factors do not support any deeper philosophical claim.

When discussing non-Bayesian approaches of handling uncertainty, Gillies himself does not refer to alternatives such as the Dempster-Shafer theory, but to measures of confirmation $C(H, E)$ for a hypothesis H by evidence E that do not satisfy the standard axioms of probability. In symbols the claim is that $C(H, E) \neq P(H \mid E)$. This is confusing because different measures of confirmation (Fitelson 1999) have been discussed extensively within the Bayesian literature. These measures (such as the difference measure, the ratio measure, and the log-likelihood measure) do not satisfy the axioms of probability theory. Gillies also mentions the certainty factors used in the expert system MYCIN, which do not obey the axioms of probability theory. It should be noted, however, that Heckerman (1986) shows that a consistent probabilistic interpretation of certainty factors *can* be given. According to this analysis, certainty factors are monotonic transformations of the likelihood ratio $P(H \mid E)/P(H \mid \neg E)$. This demonstrates that there are many different ways of handling uncertainty also within the Bayesian paradigm.

The second aspect of the Bayesian controversy is the interpretation of probability. Gillies suggests that the development of Bayesian networks supports a subjective interpretation because they were developed within the tradition of subjective Bayesianism. I do not think that this argument holds. To see why, consider the following analogy. Quantum mechanics in its current form was developed within the tradition of the Copenhagen approach of quantum mechanics. Bohr, for example, entertained the ideas that later became part of the Copenhagen interpretation long before the development of the formal theory in the 1920s. And yet, no one would claim that the success of quantum mechanics provides an argument for this interpretation. Later, other interpretations (like the many-worlds or the many-minds interpretation) were developed which turned out to be completely compatible with quantum mechanics. And indeed, Gillies himself points out that Bayesian Networks can also be applied if the probabilities in question are objective. Hence, even if Bayesian Networks were created within the tradition of subjective Bayesianism, the theory of Bayesian Networks itself is *neutral* with respect to the interpretation of probability (Bovens and Hartmann 2002, Williamson 2001). In the first place,

Bayesian Networks are a highly efficient tool to represent and manipulate probabilistic structures, irrespective of how the probabilities are interpreted. One might reply that Bayesian Networks require the specification of some prior probabilities which is, after all, a typical Bayesian step. This, however, is not correct since Gillies' own examples from medicine show that prior probabilities can also be objective.

Let us now examine if the subjective interpretation of probability was heuristically important for the development of Bayesian Networks. In his personal account, Pearl (1993) mentions a lot of factors that played a role there, but he does not mention subjective probabilities. Apart from paying lip service to subjective Bayesianism, no deeper influence of this methodological account is acknowledged. Apparently the Bayesian framework is flexible enough to allow for an incorporation of the technical ideas and practical applications that guided Pearl's thinking. In any case, philosophical questions concerning the interpretation of the probability calculus do not seem to have been that important after all.

To sum up, I do not see that the development of Bayesian Networks has an effect on the Bayesian controversy in philosophy of science. The question whether Bayesianism is a good methodology or not is of course important. It is, however, not decided or even illuminated by pointing to a commonly used tool in AI.

3. Objective Probabilities

Gillies makes several methodological recommendations in his paper. He argues, for instance, that objective probabilities should be used whenever possible and that the use of objective probabilities requires a Popperian methodology. Let us examine these two claims. To support the first claim, Gillies discusses an example from medical diagnostics. It turns out that expert systems based on objective probabilities yield better results than systems that use averaged subjective probabilities provided by experts. I agree, but I do not find this particularly astonishing. After all, the subjective probabilities provided by the experts are nothing but guesses of the values of objective probabilities. These guesses are based on the experience of the expert,

on the experience of their colleagues, and, perhaps, on what the experts took from textbooks and journal articles. That is, the subjective probabilities discussed here are hypotheses about the values of objective probabilities. An expert estimates these probabilities and this, of course, may involve errors and inaccuracies. The error will be expected to be particularly high if the sample an expert uses for his judgements is small. And this is just what Gillies' examples show.

Let us now turn to Gillies' second claim. Do objective probabilities require a Popperian testing methodology? Gillies argues that the assignment of an objective probability is a conjecture, and conjectures have to be tested. If the hypothesis does not pass the test, it is falsified and a new conjecture has to be made. I doubt that this reconstruction of scientific practice is adequate. True, to claim that an objective probability has a certain value is a scientific hypothesis. This claim is based on the available data. If new data become available, the corresponding value is *updated* (and not revised), i.e. it is changed according to all available evidence. This is done on the basis of the new *and* old data jointly. Such an updating procedure does not correspond to a test of the original hypothesis, which was, perhaps, falsified. The original hypothesis was simply modified on the basis of more data.

It is worth pointing out that such a procedure is not a problem for a Bayesian. The modification of the value we claim an objective probability to have is a different procedure than the updating of the probability of a hypothesis in the light of new evidence. It can be done "by hand" without using Bayes' theorem (Hawthorne 1993, p. 134).

4. A Testing Methodology for Bayesian Networks?

According to Gillies, a testing methodology is not only needed for the specification of various objective probabilities, but also for the determination of the structure of a Bayesian Network. In his example from medicine, one starts with causal knowledge from experts. On the basis of this knowledge, a Bayesian Network model is constructed. Now, this network implies various probabilistic independencies which may turn out not to be reflected in the statistical data the network is supposed to

represent. In this case, the hypothesis that the network in question represents the statistical data is falsified and a new hypothesis has to be found. Hence, Gillies suggests the following methodology to construct a Bayesian Network model: (1) Start with the causal knowledge of experts and construct a Bayesian Network. (2) Use all available data to test the independence assumptions (and perhaps also other assumptions) made by the network model. These tests are attractive because they provide an answer to a yes/no question: an independence condition either holds or it does not. (It is assumed that the values of the objective probabilities were already tested.) (3) If the model does not pass a test, a modified network has to be suggested based on new knowledge about various independencies.

In which situations is this procedure a good methodology? Certainly if there is reliable causal knowledge and if the number of variables is small. Both conditions hold in Gillies' example from medicine. Moreover, they are often related for it is harder to formulate reliable causal claims if the number of variables is large. However, if the number of variables is large, there may not be enough computational power available to test all independencies represented by the network.

What alternatives are there to Gillies' Popperian methodology? I will discuss three. First, if there is no causal knowledge to start with, a methodology based on Bayesian Networks can be used to discover causal knowledge. Algorithms such as the ones developed by the CMU group aim at finding a Bayesian Network which fits a certain set of statistical data best (Spirtes et al. 2000). An additional Popperian testing methodology is not needed here. It should be mentioned however, that this methodology has a lot of problems of its own and so far not much causal knowledge has been gained.

A second alternative is to apply a Bayesian methodology. Here, tests are also possible. If the data is inconsistent with the independence assumptions represented by the network, the probability of the data given the network, P(data I graph), equals zero and hence, by Bayes' Theorem, we have P(graph I data) = 0. This in turn leads to a falsification of the proposed network model (Howson and Urbach 1996, p. 84f). A problem with this procedure is that it does not directly identify the faulty

independence assumption. However, given additional background knowledge, a negative test may suggest a better model.

A third strategy, which is also a Bayesian strategy, is to start with several possible candidate models and then eliminate them successively when new data become available. Hawthorne (1993) defended this methodology and proved a number of limiting theorems that aim at demonstrating the practicability of this procedure. This strategy does, however, not seem to be realistic given certain computational limitations.

To sum up, there seems to be a plurality of strategies which guide the construction of a Bayesian Network. Gillies' Popperian testing methodology is one of them. All of these strategies have problems. It will be interesting to learn more about the conditions under which the various methodological strategies can be successfully applied.

5. The Scope of Bayesianism

Gillies mentions in this paper, and argues in more detail elsewhere, that he accepts Bayesianism only if it is about specific hypotheses. He presents a new argument for the conclusion that general hypotheses have probability zero (Gillies 1998, p. 154f). This implies, by Bayes Theorem, that also the posterior probabilities of general hypotheses are zero which, in turn, renders the whole Bayesian program useless in these cases. I will argue that Gillies' argument is flawed.

The argument is based on the observation that one cannot win a bet on a general hypothesis. One can only lose it. Hence one should not bet on a general hypothesis, which implies that its probability is zero. Let us look at a specific case. Let H be the hypothesis "All ravens are black". Since H is a general hypothesis, P(H) = 0 and so P(not-H) = 1. Hence we should be prepared to bet *everything* on the hypothesis that there is a non-black raven. I am not prepared to do so, and I am not sure if Gillies is. I think that my argument shows that something is wrong with Gillies' de Finetti style argument. I take it to be much more plausible to follow the

programme outlined by Ramsey and to assign probabilities on the basis of utilities. This might indeed lead to a non-zero probability for a general hypothesis.

Having defended Bayesianism against the charges of Gillies, I should mention a problem which Bayesians and Popperians share. Both methodologies assume that a decision can always be made. For a Popperian, a hypothesis either passes a test or is falsified; for a Bayesian it can always be decided which hypothesis of a set of hypotheses should be accepted. This might not always be possible. One can imagine situations in which there is not enough evidence to make a decision because there is too much uncertainty. Instead of making a definite decision, scientists should withhold judgement and try to collect more relevant information. That is, I suggest to weaken the link between Bayesianism and decision theory.[2]

To sum up, I have argued in this paper that the remarkable developments in AI have no baring on the Bayesian controversy. New scientific tools such as Bayesian Networks do not provide arguments in favor of a specific interpretation of probability or a specific scientific methodology. It is up to philosophical analysis to provide arguments in favor or against one of the various options.

References

Luc Bovens and Stephan Hartmann, "Bayesian Networks and the Problem of Unreliable Instruments", in: *Philosophy of Science* 69, 2002, pp. 29-72.

Branden Fitelson, "The Plurality of Bayesian Measures of Confirmation and the Problem of Measure Sensitivity", in: *Philosophy of Science* 63, 1999, pp. 652-660.

Donald Gillies*, Artificial Intelligence and Scientific Method*. Oxford: Oxford University Press 1996.

Donald Gillies, "Confirmation Theory", in: Philippe Smets (ed.), *Handbook of Defeasible Reasoning and Uncertainty Management Systems. Volume 1, Quantified*

---

[2] I owe this point Jürg Kohlas.

Representation of Uncertainty and Imprecision. Dordrecht: Kluwer Academic Publishers 1998, pp. 135-167.

Colin Howson and Peter Urbach, *Scientific Reasoning: The Bayesian Approach*. Chicago: Open Court 1996.

James Hawthorne, "Bayesian Induction Is Eliminative Induction", in: *Philosophical Topics* 21, 1, 1993, pp. 99-138.
Howson Urbach

Judea Pearl (1993), "Belief Networks Revisited", in: *Artificial Intelligence* 29, 1986, pp. 241-288.

Peter Spirtes, Clark Glymour and Richard Scheines, *Causation, Prediction, and Search*. Cambridge, MA: MIT Press 2000.

Jon Williamson, "Foundations for Bayesian Networks", in: David Corfield and Jon Williamson (eds.), *Foundations of Bayesianism*. Dordrecht: Kluwer Academic Publishers 2001, pp. 75-115.