Modeling in Philosophy of Science*

Stephan Hartmann†

October 30, 2007

Abstract

Models are a principle instrument of modern science. They are built, applied, tested, compared, revised and interpreted in an expansive scientific literature. Throughout this paper, I will argue that models are also a valuable tool for the philosopher of science. In particular, I will discuss how the methodology of Bayesian Networks can elucidate two central problems in the philosophy of science. The first thesis I will explore is the variety-of-evidence thesis, which argues that the more varied the supporting evidence, the greater the degree of confirmation for a given hypothesis. However, when investigated using Bayesian methodology, this thesis turns out not to be sacrosanct. In fact, under certain conditions, a hypothesis receives more confirmation from evidence that is obtained from one rather than more instruments, and from evidence that confirms one rather than more testable consequences of the hypothesis. The second challenge that I will investigate is scientific theory change. This application highlights a different virtue of modeling methodology. In particular, I will argue that Bayesian modeling illustrates how two seemingly unrelated aspects of theory change, namely the (Kuhnian) stability of (normal) science and the ability of anomalies to over turn that stability and lead to theory change, are in fact united by a single underlying principle, in this case, coherence. In the end, I will argue that these two examples bring out some metatheoretical reflections regarding the following questions: What are the differences between modeling in science and modeling in philosophy? What is the scope of the modeling method in philosophy? And what does this imply for our understanding of Bayesianism?

*A shortened version of this paper will appear in: M. Frauchiger and W.K. Essler (eds.), *Representation, Evidence, and Justification: Themes from Suppes (Lauener Library of Analytical Philosophy; vol.* 1*)*. Frankfurt: ontos Verlag 2008.

†Contact information: Stephan Hartmann, Tilburg Center for Logic and Philosophy of Science, Tilburg University, 5000 LE Tilburg; email: S.Hartmann@uvt.nl

1 Introduction

Philosophers of science want to understand how science works. We want to know, for example, how scientific theories and models are confirmed, how evidence for or against a theory is evaluated, how different theories hang together, and how changes in scientific theory can be understood philosophically. Questions like these have been discussed extensively over the years, and a look at the literature shows that two main approaches can be identified. On the one hand, we find scholars who formulate and defend grand normative accounts that are then used to reconstruct and, sometimes, criticize what scientists have been doing. Popper's falsificationism and its influence especially on the development of the social sciences is a case in point. Bayesianism is another. On the other hand, we find so-called naturalized philosophers of science who examine, often in great detail and inspired by the work of Thomas S. Kuhn and others, specific episodes from the history of science and from contemporary science from a philosophical perspective and give us a much more realistic picture of science and its workings than, for example, the one that Popper and Bayes gave us (see Giere (1988)).

Grand approaches often run into problems when confronted with episodes from real science. Popper's falisificationism, for example, is arguably challenged by the existence of Kuhnian normal science, and Bayesianism has not gone much beyond explaining a few simple features of scientific theorizing. But naturalized accounts are not without their problems either. Most importantly, they fail to provide generalizable insights as well as normative standards that help us to separate good science from bad science and, perhaps, science from non-science. In short, while grand approaches tend to be "too far away" from real science, naturalized accounts are "too close" to it. An acceptable account has to be located somewhere in the middle and satisfy the following two conditions: (i) It should be normative and provide a defensible general account of scientific rationality. (ii) It should be "empirically adequate" and provide a framework to illuminate "intuitively correct judgments in the history of science and explain the incorrectness of those judgments that seem clearly intuitively incorrect (and shed light on 'grey cases')", as John Worrall (2000: 32) once put it. The goal of this paper is to formulate, illustrate and defend such an account, which will include normative as well as descriptive elements.

To achieve this goal I propose to mimic what I take to be successful scientific methodology: the construction of models within the framework of a theory. The theory, in my case Bayesianism, will provide the normative framework, while the models include additional assumptions about the specific situation at hand. To motivate my approach, I'll start off in Section 2 by pointing out some of the merits of modeling in science. Section 3 outlines and criticizes *Textbook Bayesianism*, a version of Bayesianism that I will later, in Section 7, replace with *Naturalized Bayesianism*. Before, however, I will introduce in Section 4 my main modeling tool, the machinery of Bayesian networks, and present two examples of how the modeling methodology can be applied in philosophy of science in Sections 5 and 6. I'll conclude with some general reflections regarding the scope and value of philosophical modeling (Section 8).

2 Modeling in Science

Models are widely used in the sciences as well as in the teaching of science. Already in high school, students are acquainted with several models such as the model of a pendulum and Bohr's model of the atom. Later, at university, one learns about various so-called standard models such as the Standard Big Bang Model and the Standard Model of Particle Physics. And if we look at the frontiers of science, we find researchers such as Harvard's acclaimed physicist Lisa Randall studying remarkably idealized models in order to answer the most fundamental questions about the nature of space and time (regarding, for example, the dimensionality of spacetime). One of Randall's goals is to get hints as to what a *Theory of Everything* might look like (Randall 2005). Similarly, researchers in other disciplines such as biology, psychology and economics almost exclusively use models, along with experimentation, as a tool to learn something about the objects or systems under investigation. It is this ubiquity of models – as opposed to theories – in science that led philosophers of science to shift the focus of their attention from theories, which dominated the discussion from Logical Empiricism onwards, to models over the last twenty five years (see Frigg and Hartmann 2006, Morgan and Morrison 1999).[1] But why are models so popular in science?
Before addressing this question, a word is in order about what a model actually is and how theories and models differ from each other. A look into science texts is not of much help here as scientists often use the terms "theory" and "model" interchangeably. The Standard Model of Particle Physics, which many consider to be our most fundamental theory, is a case in point. So how can one distinguish between theories and models? While there is no clear-cut difference in the use of the words "theory" and "model", scientific theories, such as Newtonian Mechanics or Quantum Mechanics, are associated with predicates like 'general', 'abstract', and 'universal in scope'.

[1]Patrick Suppes was one of the pioneers of this development. See Suppes (1960, 1962) and Suppes (2002: ch. 2) for a more recent discussion.

Moreover, theories are meant to not involve idealizations (although many, if not all, of them do). On a more practical level, theories are often hard to solve. Take quantum chromodynamics (QCD), which is a part of the Standard Model of Particle Physics and considered to be *the* fundamental theory of strong interactions. QCD cannot be solved analytically, not even approximately, in the regime of low energies where many nuclear phenomena occur. This, of course, does not stop physicists from exploring the physics of atomic nuclei and their constituents. Instead of solving QCD directly, they construct easier-to-solve alternatives such as effective field theories or models such as the MIT Bag Model (Hartmann 1999). I have argued in (Hartmann 2001) that each of these accounts provides us with an understanding of some aspect of a complex phenomenon, though none of them tells us the whole story.

But models do not only have a practical advantage over theories. They are also more intuitive, visualizable, and ideally capture the essence of a phenomenon in a few assumptions. This is why models are sometimes compared with caricatures. Unlike theories, models, such as the model of the pendulum or the MIT Bag Model, are specific and limited in scope. Moreover, they clearly involve idealizations, and it does not do any harm, at least for the purpose for which the model was constructed, that some of the model assumptions are strictly speaking false. Planets, for example, are of course not point masses, and yet, our models of the planetary system work perfectly fine with this apparently crude idealization.

The models of the planetary system and the models of a pendulum are examples of what is called a *model of a theory*. They are embedded into a scientific theory, here Newtonian Mechanics, which functions as a *modeling framework*. Note that Newtonian Mechanics can hardly be tested without the specification of a model (i.e., in this case, the specification of a force function plus assumptions about the geometry of the system etc.). However, not all scientific theories are modeling frameworks, and not all models are models of a theory. Sometimes the word "theory" is used with the meaning "fundamental model" (as in the Standard Model of Particle Physics), and many scientific models are not embedded into a theory. These models are called *phenomenological models*, and Bohr's Model of the atom and the MIT Bag Model are good illustrations of this type. Some of the assumptions made in a phenomenological model may be *inspired* by a scientific theory (such as Newtonian Mechanics or QCD), while others may even contradict an accepted theory.

To sum up, there are two types of theories – theories as modeling frameworks and fundamental models – and two types of models – models of a theory and phenomenological models – in science. In the remainder, I'll focus on theories as modeling frameworks and models of a theory (and will present my own Bayesian account of what a scientific theory is in Section 6.1).

Unlike their colleagues in the sciences, many philosophers are exclusively concerned with the construction of what one could call theories – universal claims that are typically criticized or refuted by other philosophers who will then, in turn, put forward another universal claim, and so on. While such a sequence of seemingly non-converging universal claims may lead to some insight (we learn, at least, what does not work!), I submit that we can sometimes do better. More specifically, I suggest to follow Lisa Randall's lead and construct models also in philosophy. Similarly to many philosophers, Randall addresses some of the most fundamental questions one can ask, yet she does so by constructing and analyzing models, which may help her to explain phenomena and systematize existing regularities. It may also point the way to a more fundamental theory.

Before applying the modeling methodology to questions in the philosophy of science, however, a disclaimer or two is in order. Models are not always the preferred tool of the philosopher – there is a lot of space for the traditional analytic method – but there is no reason to exclude the modeling methodology from the toolbox of philosophy. Just as good scientists use many tools and continually learn about new ones, philosophers can profit from expanding their toolbox as well. This is what philosophers such as Brian Skyrms and Clark Glymour have been doing. While Skyrms applies evolutionary game theory to learn about the social contract and the evolution of cooperation (Skyrms 1996), Glymour and his colleagues use Bayesian network models to "discover" causes from statistical data (Spirtes et al. 2000). Following their lead, I will show that modeling can also be applied in the methodology of science. To do so, one can choose from a variety of modeling frameworks ranging from various applied logics, game theory, and Bayesianism to alternatives to probability theory such as Dempster-Shafer theory. All of these frameworks have their value and can be used to address many interesting problems. In this paper, though, I will focus exclusively on the construction and analysis of Bayesian network models.

3 Textbook Bayesianism

Bayesianism is a quantitative confirmation theory that was developed in the middle of the last century in light of the problems (such as the Raven Paradox) that qualitative theories of confirmation (such as hypothetico-deductivism) face. While qualitative theories formulate criteria that inform us whether or not a piece of evidence E confirms a hypothesis H, quantitative theories of confirmation also tell us how much E confirms H. Clearly, learning that the butler owns a Smith and Wesson, that he had an affair with the victim's wife and that he was in the castle when the count was murdered confirms the hypothesis that he was the murderer more than learning only that he was in the castle when the murder happened. A quantitative theory of confirmation accounts for this difference and assigns degrees of confirmation. Bayesianism is the most developed and most popular quantitative theory of confirmation. According to Bayesianism, scientists have subjective degrees of belief about certain hypotheses and, as rational agents, change their degrees of belief in the light of new evidence. These degrees of belief have to obey the axioms of probability theory, as several arguments aim at showing. The most popular of these arguments are the so-called Dutch book arguments, which demonstrate that any violation of the axioms of probability theory will necessarily lead to a loss of money of an agent in a betting situation (see, e.g., Skyrms 1999). So we have to make sure that our degrees of belief obey the axioms of probability theory (or are coherent) if we do not want to loose money. Coherence is a rather weak requirement, but it is a requirement all the same, and it makes Bayesianism a normative philosophical theory. Unlike naturalized philosophies of science, which identify scientific rationality with the practice of good scientists (assuming that they never err), Bayesianism formulates constraints on the beliefs of a scientist.

Confirmation, then, amounts to this for the Bayesian: A scientist starts with a subjective degree of belief that a certain hypothesis H is true. This degree of belief is called the *prior probability* of the hypothesis: It is denoted by $P(\mathrm{H})$. The prior probability is informed by the scientist's background knowledge and the assignment may differ from scientist to scientist. If H is the proposition "The patient has tuberculosis", then doctors will typically fix $P(\mathrm{H})$ by drawing on statistical (i.e. frequentist) data, provided that such data is available. Differing assignments may be made as long as they are coherent. In the next step, a piece of evidence, E, say a positive result of an X-ray scan, comes in. E prompts an update of the probability of H and the scientist assigns a *posterior probability* $P_{new}(\mathrm{H})$. E confirms H if the posterior probability $P_{new}(\mathrm{H})$ is greater than the prior probability $P(\mathrm{H})$, and E disconfirms (or falsifies) H if the posterior is smaller than the prior.

But how are probabilities updated? How do we get from a prior probability of the hypothesis $P(\mathrm{H})$ to the posterior probability $P_{new}(\mathrm{H})$? There is a lot of debate about this question amongst Bayesians and there is clearly no consensus. Many, however, hold that the new probability measure should be the conditionalized old probability measure:

$$P_{new}(\mathrm{H}) := P(\mathrm{H}|\mathrm{E}) \tag{1}$$

I'll not defend this choice, but will use it in the remainder. Note, however, that this choice is not dictated by mathematics. What is dictated by mathematics is how one evaluates the expression on the right hand side of eq. (1). Applying Bayes' Theorem to the right hand side of eq. (1), one obtains

$$P_{new}(\mathrm{H}) = \frac{P(\mathrm{E}|\mathrm{H})P(\mathrm{H})}{P(\mathrm{E})}, \tag{2}$$

according to which the posterior probability of H is proportional to the prior probability $P(\mathrm{H})$ and the *likelihood* P(E|H), i.e. the probability that the evidence obtains given that the hypothesis is true, and inversely proportional to the *expectancy* $P(\mathrm{E})$ of the evidence.

Textbook Bayesianism, as presented in Howson and Urbach (2006) and Earman (1992), uses eq. (2) to account for a number of methodological rules. To make things easy, let's assume that the evidence is deductive, i.e. that E can be deduced from H. This assumption often holds in science.[2] For example, if we want to confirm the hypothesis that all metals conduct electricity, we deduce from this hypothesis that a certain piece of copper conducts electricity. Whether or not this obtains in an experiment confirms or disconfirms H. If E is deductive evidence for H, then the likelihood is $P(\mathrm{E}|\mathrm{H}) = 1$ and our updating rule reduces to

$$P_{new}(\mathrm{H}) = P(\mathrm{H})/P(\mathrm{E}). \tag{3}$$

Eq. (3) shows that surprising (i.e. unexpected) evidence confirms a hypothesis better than expected evidence for, if E is surprising, then $P(\mathrm{E})$ is small and hence $1/P(\mathrm{E})$ is large. And so $P_{new}(\mathrm{H})$ is large as well.

The updating rule can also account for the claim that more varied evidence confirms better than less varied evidence. This is the variety-of-evidence thesis, which is, or so it seems, a truism of scientific theorizing. Let us, for example, drop a rock drop 25 times in a row and measure the time it takes it to hit the ground. This procedure is expected to confirm Newtonian Mechanics less than a more diverse collection of evidence, such as letting the rock drop only once, doing a pendulum experiment in London, another one on the moon, observing the orbits of various planets, etc. A way to proof the variety-of-evidence thesis in the Bayesian framework starts with the following explication: more varied evidence is less correlated evidence. Correlation can then, in turn, be explicated probabilistically. Let $\mathrm{E} = \mathrm{E}_1 \wedge \mathrm{E}_2 \wedge \cdots \wedge \mathrm{E}_n$ and expand $P(\mathrm{E}) = P(\mathrm{E}_1)P(\mathrm{E}_2|\mathrm{E}_1)P(\mathrm{E}_3|\mathrm{E}_1,\mathrm{E}_2)\ldots P(\mathrm{E}_n|\mathrm{E}_1,\ldots,\mathrm{E}_{n-1})$, which

[2] Paul Teller reminds me on Duhem's insight of the importance of background assumptions B. He is right. To account for this, I suggest to use a probability measure that takes B implicitly into account, i.e. to use $P(\cdot) := P^*(\cdot|\mathrm{B})$ instead of $P^*(\cdot|\mathrm{B})$.

follows from the axioms of probability theory and the definition of conditional probability. Obviously, $P(\mathrm{E})$ is small if the conditional probabilities in this product are small, and this is exactly the case if the different pieces of evidence are not much correlated (or even anti-correlated) probabilistically. And again, if $P(\mathrm{E})$ is small, then $P_{new}(\mathrm{H})$ is large, which proofs that more varied evidence confirms a hypothesis better than less varied evidence, i.e. the variety-of-evidence thesis.

To conclude, then, *Textbook Bayesianism* explains some apparent truisms about the methodology of science. I'd like to argue, however, that *Textbook Bayesianism* is too general as it does not take into account empirical constraints such as dependencies between partially reliable measurement instruments in the discussion of the variety-of-evidence thesis. Moreover, *Textbook Bayesianism* has an overly-simplistic account of a scientific theory. Theories are complex and highly interrelated objects, and scientists typically only test parts of a theory while the rest of the theory is only indirectly confirmed. This complexity can hardly be captured by a single proposition H that is correlated, as a whole, with a piece of evidence E.

While this version of *Textbook Bayesianism* is too general, there is another version that is too specific. The philosopher of science Jon Dorling gives Bayesian reconstructions of specific episodes in the history of science (Dorling 1979). To do so, he estimates probability values of various hypotheses, based on what he knows about the knowledge people had at a certain time and concludes, after having done the calculations, that the scientists were justified to consider a certain hypothesis (e.g. a gas law) as confirmed by the available evidence. Although I think that such results are interesting, Dorling's version of *Textbook Bayesianism* is, in the end, too specific. We do not learn much in general about the methodology of science and to engage in a debate about specific probability values seems, despite Dorling's efforts, somewhat arbitrary.

To sum up: We need a version of Bayesianism that is not too general (to connect to the practice of science) and not too specific (to gain philosophical insight). It would also be nice to have an account that has a somewhat wider scope, i.e. an account that reaches beyond confirmation theory. I'll show that the use of modeling techniques within the Bayesian framework will meet these desiderata. Models help bridge the gap between a general theory (here: Bayesianism) and the scientific practice. I'll take Bayesianism as a modeling framework just as Newtonian Mechanics is a modeling framework. It has to be supplemented by model assumptions if one wants to apply it to real cases. I will show that these models explain features such as the stability of normal science and help challenge widely held methodological views such as the variety-of-evidence thesis.

4 Bayesian Networks

The theory of Bayesian networks was developed by Judea Pearl and his colleagues at UCLA in the early 1980s, building on results by Philip Dawid and Wolfgang Spohn on conditional independence structures (Pearl 1988). The main goal of the theory of Bayesian networks is to provide an efficient graphical representation of a joint probability distribution over a large number of variables and to develop algorithms to compute functions of the joint probability distribution. To reduce the complexity of the problem, Bayesian networks encode knowledge about probabilistic independencies that hold among the variables. They play an important role in expert systems and are used in many parts of science, engineering and medicine where inferences have to be drawn on the basis of complex though uncertain information.

Medical inferences are a good case to illustrate Bayesian networks. Let's assume that we want to test whether a patient has a certain disease (say, tuberculosis) by making an X-ray scan. A representation of the testing situation has to take into account that the X-ray machine is only partially reliable. It will sometimes tell us that the patient has the disease when he in fact does not have it (a false positive), and it will sometimes not give us a positive report when the patient does have the disease (a false negative).

Let's represent the situation by two binary propositional variables T and X (in italics). The positive instantiation of T – T (in roman script) – represents the proposition "Patient has tuberculosis", and the negative instantiation ¬T stands for "Patient does not have tuberculosis". Similarly, X stands for "We obtain a positive X-ray report", and ¬X stands for "We obtain a negative X-ray report". We know from statistical data that $P(\mathrm{T}) = .01$ (i.e. that 1% of the patients who undertake an X-ray scan have tuberculosis) and that the likelihoods $P(\mathrm{X}|\mathrm{T}) = .95$ (corresponding to a false negative rate of 5%) and $P(\mathrm{X}|\neg\mathrm{T}) = .02$ (corresponding to a false positive rate of 2%). We can then calculate the probability that the patient has tuberculosis, given that we obtained a positive X-ray report. To determine the posterior, we apply Bayes' Theorem,

$$P(\mathrm{T}|\mathrm{X}) = \frac{P(\mathrm{X}|\mathrm{T})P(\mathrm{T})}{P(\mathrm{X})}, \tag{4}$$

and the law of total probability,

$$P(\mathrm{X}) = P(\mathrm{X}|\mathrm{T})\,P(\mathrm{T}) + P(\mathrm{X}|\neg\mathrm{T})\,P(\neg\mathrm{T}), \tag{5}$$

and obtain, after plugging in the numbers, that $P(\mathrm{T}|\mathrm{X})$ is .32, i.e. 32%.

This situation can be represented by the Bayesian network in figure 1. The two nodes represent the propositional variables X and T and the arrow indicates that there is a direct probabilistic dependence between them. This

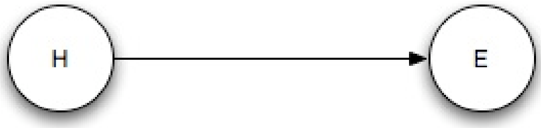

Figure 1: A Bayesian network with two nodes

dependence is sometimes read causally, but this is not necessary for our purposes. Nodes with only outgoing arrows are called *root nodes* or nodes without a *parent*, all other nodes are *child nodes*. A Bayesian network, then, is a *directed acyclic graph* (dag, for short) with a probability distribution defined over it. To fully specify the probability distribution of a Bayesian network, we have to assign prior probabilities of all root nodes (here only the propositional variable T), and conditional probabilities of all other (child) nodes given all instantiations of their parents. In our case, we have to assign $P(\mathrm{X}|\mathrm{T})$ and $P(\mathrm{X}|\neg\mathrm{T})$. With this information at hand, the Bayesian network is fully specified and we can use various algorithms to calculate, for example, $P(\mathrm{T}|\mathrm{X})$.

Clearly, the machinery of Bayesian networks is not required to deal with a case as simple as the present one. In real life, however, the situation is often much more complicated and more variables are involved. Lauritzen and Spiegelhalter (1988) give the following illustration that is known under the name "Asia example" in the literature. The network in figure 2 depicts a whole range of interrelated habits, diseases and symptoms. Given the probabilistic information encoded in the network, it is now not any longer possible to easily draw inferences like the one we drew for the two-node Bayesian network.

A Bayesian network encodes probabilistic information in an intuitive way; we see immediately which nodes are directly related, and we can read off conditional independencies by using the *Parental Markov Condition* that is built in the Bayesian network:

> (PMC) A variable represented by a node in a Bayesian network is independent of all variables represented by it non-descendent nodes in the Bayesian network, conditional on all variables represented by its parents.

Applying PMC, we can, for example, read off the Asia network that "lung cancer?" (L) is independent of "bronchitis?" (B) given their common cause "smoking?" (S), or, symbolically $L \perp\!\!\!\perp B|S$. Bayesian networks also have the

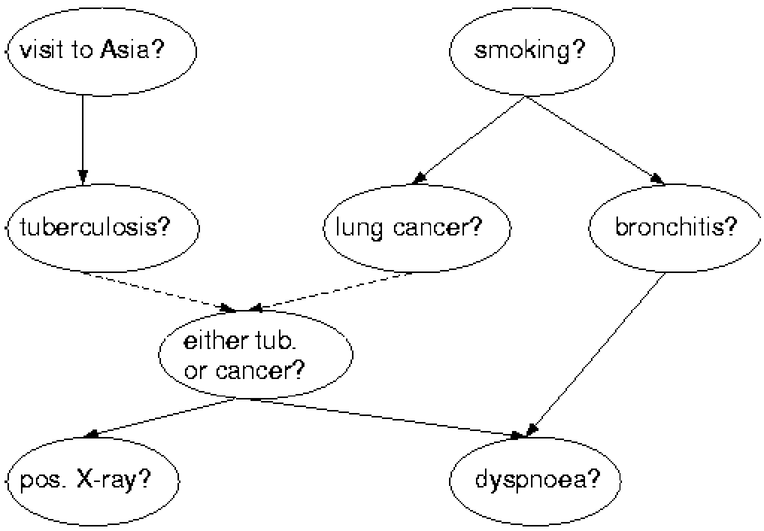

Figure 2: A more complicated Bayesian network

practical advantage, in that only a small number of probabilities have to be specified. Remember that $2^n - 1$ probabilities have to be fixed in order to fully specify the joint distribution over n binary variables. In a Bayesian network, we only have to specify the prior probabilities of all root nodes, and the conditional probabilities of all child nodes. This reduces the numbers considerably and makes much more complex situations manageable. For instance, in the Asia example, the knowledge about conditional independencies reduces the number of probabilities that have to be specified from 255 to 18 – an enormous reduction indeed.

The construction of a Bayesian network model proceeds in four steps: First, identify the set of all relevant variables. Second, specify all conditional independencies that hold between these variables. Third, construct a dag that incorporates these conditional independencies. Fourth, specify the prior probabilities of all root nodes and the conditional probabilities of all child nodes. Once all this is done, the joint probability distribution over the variables is specified and any probability of interest can be computed. Let's apply this methodology to problems from the philosophy of science.

5 Example 1: The Variety-of-Evidence Thesis

Let's return to the already mentioned variety-of-evidence thesis and ask whether the thesis holds in general or whether it has exceptions. Are there

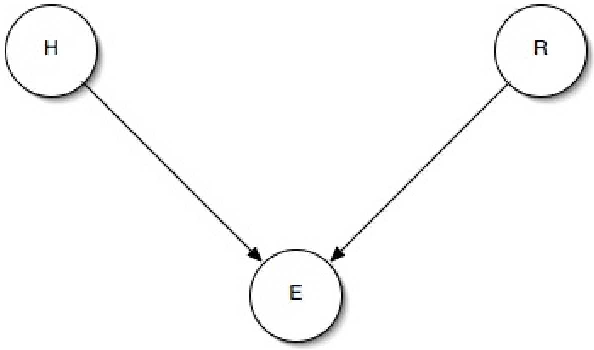

Figure 3: A Bayesian network representing the test of a hypothesis with a partially reliable instrument

testing scenarios in which less varied evidence confirms a hypothesis more than more varied evidence?[3]

To test the variety-of-evidence thesis, let's assume that hypothesis H was positively tested by an instrument I_1 and consider two options for a second test. In **option 1**, a different instrument I_2 is used to test the hypothesis. In this case, the second test is *independent* of the first test. In **option 2**, I_1 is used again, which renders the two tests dependent. Clearly, the variety-of-evidence thesis suggests that option 1 is to be preferred as the posterior probability of the hypothesis after two positive test reports will presumably be greater for independent tests than for dependent tests. Following our methodology, we construct a model to find out if this is the case.

To warm up, let's consider the test of a hypothesis H with one partially reliable instrument. Our model has three binary propositional variables: (i) H represents the hypothesis and has the instantiations H (hypothesis holds) and ¬H (hypothesis does not hold). (ii) E represents the evidence and has the instantiations E (evidence obtains) and ¬E (evidence does not obtain). (iii) R represents the reliability of the instrument and has the instantiations R (instrument is reliable) and ¬R (instrument is not reliable). E is dependant on H and R, but H and R are independent (in symbols: $H \perp\!\!\!\perp R$). This makes sense, for the truth value of the hypothesis does not depend on the reliability of the instrument. We can now construct a Bayesian network that incorporates these assumptions (see figure 3). Finally, we have to specify the prior probabilities of all root nodes (H and R) and the conditional probabilities of all child nodes (E). We set

$$P(\mathrm{H}) = h \quad , \quad P(\mathrm{R}) = \rho. \tag{6}$$

[3]This section follows Bovens and Hartmann (2004), ch. 4.

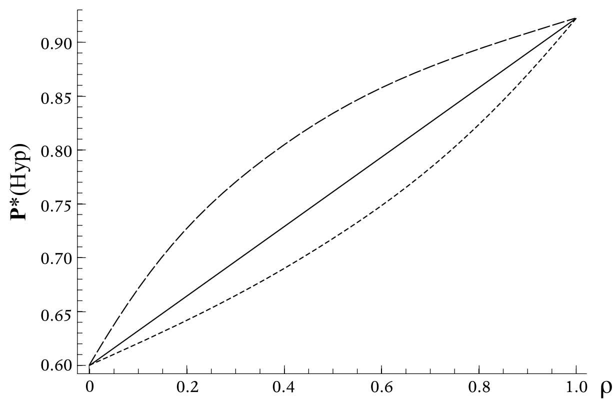

Figure 4: The posterior probability that the hypothesis (HYP) is true as a function of the reliability parameter ρ when the randomization parameter is set at $a = .5$ (full line), $a = .2$ (dotted line) and $a = .9$ (dashed line) and $P(\mathrm{H}) = .6$.

Specifying the conditional probabilities of E is a bit harder as we have to make assumptions about the reliability of the instrument. To keep things simple, we assume that the instrument is a truth-teller if it is in the reliable mode (i.e. if R is instantiated), and a randomizer, if it is in the unreliable mode (i.e. if $\neg$R is instantiated). So we set:

$$\begin{aligned} P(\mathrm{E}|\mathrm{H},\mathrm{R}) = 1 \quad &, \quad P(\mathrm{E}|\neg\mathrm{H},\mathrm{R}) = 0 \\ P(\mathrm{E}|\mathrm{H},\neg\mathrm{R}) = a \quad &, \quad P(\mathrm{E}|\neg\mathrm{H},\neg\mathrm{R}) = a \end{aligned} \tag{7}$$

a is called the *randomization parameter*. Note that the probability of E does not depend on whether H is true or not if the instrument is in the unreliable mode. This way of modeling a partially reliable instrument is clearly a strong idealization, which will not hold in many cases.

Applying the machinery of Bayesian networks, we can now calculate the posterior probability $P(\mathrm{H}|\mathrm{E})$ of the hypothesis after receiving a piece of evidence. The result is plotted in figure 4. We see that the posterior equals the prior if $\rho = 0$, which renders the test to be useless. Conversely, if $\rho = 1$, the posterior is 1, i.e., if God tells us that the evidence obtains, then the hypothesis must be true. For values of the reliability parameter in between 0 and 1, the posterior is an increasing function of ρ, with the slope depending on the randomization parameter a.

Let's now generalize this simple model to the more complicated case of two tests of a hypothesis. Our two testing options can be represented by the Bayesian networks depicted in figures 5 and 6. The network in figure 5 models testing option 1, i.e. repeated testing with two dependent instruments, as

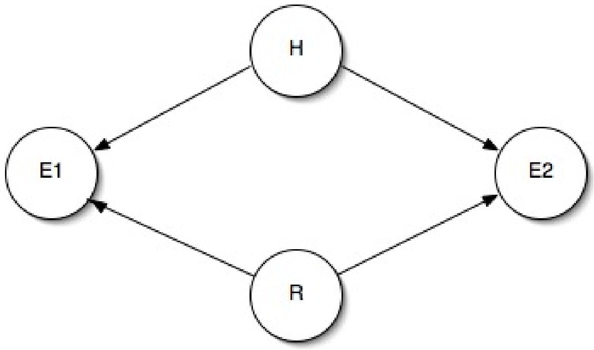

Figure 5: A Bayesian network representing the test of a hypothesis with two independent instruments

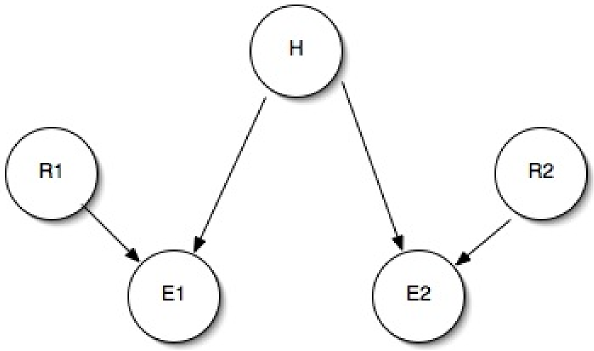

Figure 6: A Bayesian network representing the test of a a hypothesis with two dependent instruments

there is only one reliability node R that connects the two report nodes. E_1 and E_2 are conditionally dependent on each other: $P(\mathrm{E}_1, \mathrm{E}_2|\mathrm{H}, \mathrm{R}) \neq P(\mathrm{E}_1|\mathrm{H}, \mathrm{R})P(E_2|\mathrm{H}, \mathrm{R})$. The Bayesian network in figure 6 represents testing option 2. Now, each evidence node has its own reliability node (R_1 or R_2, respectively), and the two tests are conditionally independent of each other: $E_1 \perp\!\!\!\perp E_2|H$.[4]

We apply the methodology of Bayesian networks and calculate the posterior probability $P(\mathrm{H}|\mathrm{E}_1, \mathrm{E}_2)$ for the two cases. Let P_1 be the probability measure for the case of independent instruments, and P_2 be the probability measure for the case of dependent instruments. This probability will depend on the parameters a, h, and ρ which we assume to have the same values in both cases (without this ceteris paribus clause it would not make sense to compare the

[4]This independence follows from the $d-$separation criterion, which is explained, for instance, in Pearl (1988). Applying PMC, we only get $E_1 \perp\!\!\!\perp E_2|H, R_1$ and $E_2 \perp\!\!\!\perp E_1|H, R_2$.

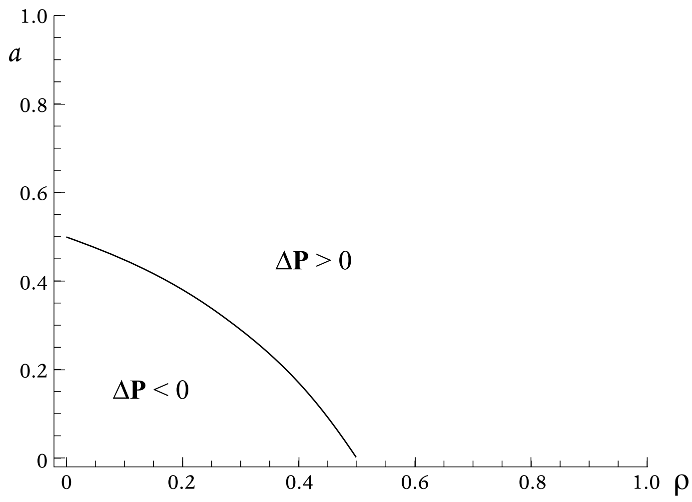

Figure 7: Phase plot illustrating Theorem 1.

two scenarios). To find out which testing option leads to more confirmation (i.e. to a higher posterior probability), we calculate the difference of the posteriors and ask when this difference is positive. The following theorem holds:

Theorem 1 *Let* $\Delta P = P_1(\mathrm{H}|\mathrm{E}_1, \mathrm{E}_2) - P_2(\mathrm{H}|\mathrm{E}_1, \mathrm{E}_2)$*, with* P_1 *and* P_2 *as defined in the text. Then* $\Delta P > 0$*, if and only if* $1 - 2(1-a)(1-\rho) > 0$.[5]

The phase plot in figure 7 illustrates this result. Below the phase curve in the $\rho{-}a$ diagram, the posterior probability for two tests with the same instrument is higher. For parameter pairs (a, ρ) above the phase curve, two independent tests result in a greater posterior probability, i.e. in more confirmation.
Do these findings make sense? Our model clearly makes a lot of idealizing assumptions and it is hence not clear at all whether we should accept what it tells us. Without a *story* that makes the consequences of the model plausible, the model clearly belongs in the waste bin. So let's try to understand our findings.
Obviously two conflicting intuitions are at work here. On the one hand, independence is a good thing. We are impressed by two positive reports from independent instruments. On the other hand, coherent test results from a single instrument increase our confidence in the reliability of the instrument. And this, in turn, increases our degree of confidence in the truth of the hypothesis.

[5]The proof of the theorem can be found in Bovens and Hartmann (2004: 144).

To sum up: Our model shows that the variety-of-evidence thesis is not sacrosanct. It can fail, and it does indeed fail under certain conditions. The model points to the existence of these cases, and it makes plausible why the thesis fails in these cases. It acts like a midwife as it helps us to discover a new feature and it leaves us with the job of making this new feature plausible. Ideally, we are able to tell a *model-independent story* that makes the results of the model plausible. While this works sometime, it does not fully work in this case as the details of the modeling of the partial reliability of the instruments matters for the story.

6 Example 2: Theories and Normal Science

Scientific theories change over time. Sometimes a theory is completely given up (as the phlogiston theory in chemistry), and sometimes (and presumably more often) elements of the old theory are carried over to the successor theory. This holds especially for science in the last one hundred years. But which elements are carried over depends on the specific case in question and I do not think that something general (such that only 'structures' survive theory change) can be said here (see Hartmann (2002)).

So the question arises whether a philosophical account of scientific theory change can be given. So far, the philosophical debate has largely been centered around the question as to whether a rational reconstruction of scientific theory change can be given or not. Challenged by Kuhn and Feyerabend, who stressed radical changes (so called scientific revolutions), quasi-religious conversion from one paradigm to another, and incommensurability, many philosophers have attempted to present a rational reconstruction of scientific theory change *tout cour*. They aimed at replacing Kuhn's grand approach by another grand approach and I think that this is too fast.

Instead I suggest to aim at explaining empirically established aspects of scientific theory change in a philosophical (here: Bayesian) framework. Which aspects become the subjects of explanation has to be determined by historical research. The challenge, then, is to see whether such an integration of modeling and empirical research can be facilitated. Note that the project I am suggesting is in the tradition of Wesley Salmon's programmatic article "Tom Kuhn Meets Tom Bayes" in which Salmon attempts to reconcile Bayesianism with the work of Thomas S. Kuhn (Salmon 1990). In this section, I want to take a modest step in this direction by explaining one aspect of the dynamics of science related to the work of Kuhn: the stability of normal science. Before, however, a word is in order about what a scientific theory is from a Bayesian point of view.

6.1 What is a scientific theory?

Scientific theories such as Newtonian Mechanics and Quantum Mechanics are highly complex and interrelated wholes. In them we find, for example, principles, empirical assumptions, laws, and models. How are they all related, and how are they integrated into a theory? A simple philosophical account such as *Textbook Bayesianism*, which only talks about a hypothesis H and a piece of evidence E, does not suffice to adequately address these questions and to capture the complexity of real scientific theories.

Unfortunately, the two major accounts discussed in the philosophical literature – the syntactic view and the semantic view – do not suffice either. According to the syntactic view, scientific theories are linguistic entities, sets of propositions formulated in, say, first-order logic. Clearly, the linguistic formulation of a theory matches well with the way scientists formulate a theory. However, it is important to note that inductive relations between elements of a theory cannot be captured adequately. This is a serious flaw, as inductive relations between the elements of a theory are crucial, for example, when it comes to assess a theory. According to the semantic view, championed by Suppes and others, scientific theories are non-linguistic entities, and so a theory is not equated with a specific linguistic representation. It is rather identified with the set of models, i.e. realizations of the abstract mathematical structure of the theory. Some adherents of the semantic view, such as Ronald Giere (1988) and the Munich structuralists, have stressed the interrelatedness of the models of a theory. But they lack a detailed account of the nature and evidential relevance of these relations (cf. Gähde 1996).

As we will see, the Bayesian view combines elements of the syntactic view with elements of the semantic view while taking account of the deficiencies of both approaches. On the Bayesian view, a scientific theory has an empirical and a non-empirical (or heuristic) part. The *non-empirical part* of a scientific theory consists of the laws and principles of the theory, which are, perhaps, organized in an axiomatic structure. It also comprises certain tools and tricks that are used to construct models in the framework of the theory and assumptions about the domain of the theory. Note that the non-empirical part includes much that is captured in Kuhn's notoriously vague notion of a paradigm. The non-empirical part of a theory is not probabilified.

The *empirical part* of a scientific theory, however, is probabilified. It is explicated as a Bayesian network. The nodes of the network represent the models of the theory, which account for (and are confirmed by) empirical phenomena. The models themselves are conjunctions of propositions, some of which are instantiations of the laws included in the non-empirical part of the theory, others are additional assumptions that have to be made to

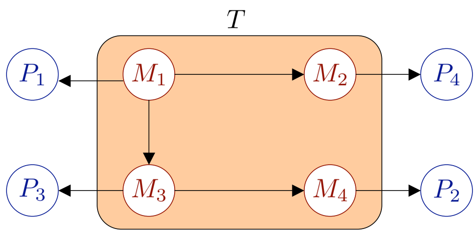

Figure 8: The empirical part of a scientific theory

account for a phenomenon. A probability distribution is defined over all model and phenomena variables, but the propositions within the model are not probabilified. An arrow connects two models in the Bayesian network if there is a direct probabilistic dependence between them. Such a dependence is generated, for example, by the instantiation of the same law in both models. Consider Newtonian Mechanics. In this theory, Newton's Second Law is instantiated in most, if not all, models. This makes it plausible that for any two models M_1 and M_2 of the theory, $P(\mathrm{M}_2|\mathrm{M}_1) \neq P(\mathrm{M}_2)$. But there might be other relations as well. Note that the empirical part of the theory is a dynamical entity. It grows in the course of time as more and more models are added to the empirical part of the theory to capture more and more phenomena.

Phenomena can relate to the models of a theory in different ways. One option is that each model M_i is accounted for by just one phenomenon E_i and that E_i is probabilistically independent of (and hence does not confirm) any other model in the theory (see figure 8).[6] This is a rather unrealistic assumption, which typically does not hold. For example, it does not allow for the indirect confirmation of a model M_2 by a phenomenon E_1 that is accounted for by a

[6] I represent phenomena by variables E_i instead of P_i to avoid confusion with the probability measure P and to indicate that the phenomenon E_i is *evidence* for model M_i.

model M_1 that is related to M_2 in the network representing the empirical part of the theory. For more on indirect confirmation, see Dietrich and Moretti (2005). I take it to be an empirical question how models and phenomena are related, and the corresponding Bayesian network has to be constructed accordingly.

6.2 The stability of normal science

Kuhn observed that science is characterized by long periods of what he calls normal science. In these periods, Kuhn provocatively wrote, it is not the scientific theory that is tested, but the scientist (Kuhn 1970: 5). If a scientist is able to solve a given problem (or a puzzle, as Kuhn put it), then she passed the test, and if she fails to solve it, then she does not pass the test. In this way, more and more models are added to the cluster of models that comprise, in our reconstruction, the empirical part of a scientific theory. Phenomena are fitted into the framework of the theory by constructing appropriate models, and typically no surprises arise. Understandably, Sir Karl Popper felt challenged and provoked by these claims. He could not accept the apparently uncritical attitude of scientists who do not seem to follow his conjectures-and-refutations methodology. However, Popper admitted that Kuhn's claims are descriptively correct and gave up his efforts to find a rationale for the stability of normal science (Popper 1970: 52). For Popper, scientists who act as Kuhn describes it are simply irrational, as no critical-rationalistic justification for the stability of normal science can be given. However, the failure to give a Popperian justification does not entail that no epistemic justification can be given at all. So let's explore the prospects for a Bayesian justification.
At first sight, a Bayesian justification of the stability of normal science seems to be hard to come by as normal science raises an immediate problem for a probabilist. According to probability theory the probability of a conjunction cannot be greater than the probability of one of the conjuncts. So by adding more and more models to the empirical part of a theory, the joint probability of all models will monotonously decrease. And this can't be right. Let's refer to this problem as the *conjunction problem* and ask how a Bayesian can deal with it.
To dissolve the conjunction problem, note that models are related to evidence, which has to be taken into account in an assessment of a theory. Each model M_i accounts for a phenomenon, which is, in turn, evidence E_i for the model. We therefore have to consider conditional probabilities, not unconditional ones. It is easy to see that the conjunction problem disappears if we conditionalize on the combined evidence, i.e. the conjunction of all relevant E_i's as it is a mathematical truism that $P(\mathrm{M}_1, \mathrm{M}_2, \ldots, \mathrm{M}_n | \mathrm{E}_1, \mathrm{E}_2, \ldots, \mathrm{E}_n)$

can be greater than $P(\mathrm{M}_1, \mathrm{M}_2, \ldots, \mathrm{M}_{n-1}|\mathrm{E}_1, \mathrm{E}_2, \ldots, \mathrm{E}_{n-1})$.
So let's conceptualize the empirical part of a theory as laid out in Section 6.1 and ask under which conditions $P(\mathrm{M}_1, \mathrm{M}_2, \ldots, \mathrm{M}_n|\mathrm{E}_1, \mathrm{E}_2, \ldots, \mathrm{E}_n)$ is greater than $P(\mathrm{M}_1, \mathrm{M}_2, \ldots, \mathrm{M}_{n-1}|\mathrm{E}_1, \mathrm{E}_2, \ldots, \mathrm{E}_{n-1})$. The answer to this question involves the epistemological notion of coherence, which should not be confused with the confirmation-theoretic notion of coherence mentioned in Section 3. The coherence of a set of propositions informs us how well the propositions in the set hang together. Coherence comes in degrees: some sets are more coherent, other are less coherent. Interestingly, we often make good judgments, which of two sets of propositions is more coherent, without examining the underlying probability distribution over the propositions. However, a probabilistic account of coherence can be given. I shortly lay out the account given by Bovens and Hartmann (2003, 2004). The reader may consult these publications for details and motivation.
Let's assume that we have a set of propositions $\mathcal{T} = \{\mathrm{M}_1, \ldots, \mathrm{M}_n\}$ and a set of pieces of evidence $\mathcal{E} = \{\mathrm{E}_1, \ldots, \mathrm{E}_n\}$ such that the following independence assumption holds:

$$E_i \perp\!\!\!\perp M_1, E_1, \ldots, M_{i-1}, E_{i-1}, M_{i+1}, E_{i+1} \ldots, M_n, E_n | M_i \text{ for } i = 1, \ldots, n \quad (8)$$

The Bayesian network in figure 8 incorporates this assumption. We also assume that each model M_i is supported with the same strength by its corresponding evidence E_i. This strength is measured by

$$r := 1 - P(\mathrm{E}_i|\neg\mathrm{M}_i)/P(\mathrm{E}_i|\mathrm{M}_i) \text{ for all } i = 1, \ldots, n. \quad (9)$$

Note that $r \in (0, 1)$. We define $a_i = P(n - i$ propositions of $\mathcal{T}$ are true) for $i = 0, \ldots, n$ and the *weight vector* of $\mathcal{T}$ by $\langle a_0, \ldots, a_{n-1}\rangle$. Then the posterior probability is given by

$$P(\mathrm{M}_1, \ldots, \mathrm{M}_n|\mathrm{E}_1, \ldots, \mathrm{E}_n) = \frac{a_0}{\sum_{i=0}^n a_i(1-r)^i} \ . \quad (10)$$

Moreover, Bovens and Hartmann argue that $\mathcal{T} = \{\mathrm{M}_1, \ldots, \mathrm{M}_n\}$ is more coherent than $\mathcal{T}' = \{\mathrm{M}'_1, \ldots, \mathrm{M}'_m\}$ if, for all values of r, $c_r(\mathcal{T}) > c_r(\mathcal{T}')$ with

$$c_r(\mathcal{T}) = \frac{a_0 + (1-a_0)(1-r)^n}{\sum_{i=0}^n a_i(1-r)^i} \ . \quad (11)$$

Note that $c_r(\mathcal{T})$ is not a coherence measure. It is rather a function that generates a coherence quasi-ordering. It does not generate a complete ordering as it may happen that $c_r(\mathcal{T}) > c_r(\mathcal{T}')$ for some values of r, and $c_r(\mathcal{T}) < c_r(\mathcal{T}')$ for other values of r. In this case it is indeterminate which of the two sets is more coherent.

We have now the necessary prerequisites to formulate the following theorem (proven in the Appendix), which will be used to give an epistemic justification for the stability of normal science.

Theorem 2 *Let* $\mathcal{T} = \{\mathrm{M}_1, \ldots, \mathrm{M}_{n-1}\}$ *and* $\mathcal{T}' = \mathcal{T} \cup \{\mathrm{M}_n\}$ *be two theories and* $\mathcal{E} = \{E_1, \ldots, E_{n-1}\}$ *and* $\mathcal{E}' = \mathcal{E} \cup \{\mathrm{E}_n\}$ *the corresponding evidence sets. A probability measure* P *is defined over* $\{\mathrm{M}_1, \ldots, \mathrm{M}_n, \mathrm{E}_1, \ldots, \mathrm{E}_n\}$ *and the independence assumption (8) holds.* $\langle a_0, \ldots, a_{n-1}\rangle$ *and* $\langle b_0, \ldots, b_n\rangle$ *are the weight vectors of* $\mathcal{T}$ *and* $\mathcal{T}'$*, and* $r_0 := (a_0 - b_0)/[a_0(1-b_0)]$*. Moreover, all* M_i *are supported by* E_i *with the same strength* $r = 1 - P(\mathrm{E}_i|\neg\mathrm{M}_i)/P(\mathrm{E}_i|\mathrm{M}_i) > r_0$*, and* $\mathcal{T}'$ *is more coherent than* $\mathcal{T}$*. Then the posterior probability of* $\mathcal{T}'$ *is greater than the posterior probability of* $\mathcal{T}$*, i.e.* $P(\mathrm{M}_1, \ldots, \mathrm{M}_n|\mathrm{E}_1, \ldots, \mathrm{E}_n) > P(\mathrm{M}_1, \ldots, \mathrm{M}_{n-1}|\mathrm{E}_1, \ldots, \mathrm{E}_{n-1})$ *and* $\mathcal{T}'$ *is better confirmed by* $\mathcal{E}'$ *than* $\mathcal{T}$ *is by* $\mathcal{E}$*, if one uses the distance, ratio or likelihood measure of confirmation.*

How does this theorem help us explain and justify the stability of normal science? To address this question, let's see if the assumptions of the theorem are satisfied in real science. First, I take it to be empirically established that scientific theories become increasingly coherent in the course of scientific theorizing. The models in a theory are related to each other and support each other. And the more models are added to the theory, the better they will typically cohere.[7] Second, in a period of normal science, all models account for their corresponding phenomena, and it is plausible to assume that the strength of support a model gets from a phenomenon is more or less the same for all models. This is clearly an idealization, but I think that it is a fairly good one. Third, as scientific theories contain a large number of models, their prior probability does not decrease much when a new model is added. In this case, r_0 is small, and so the theorem holds for a large range of values of r. For example, if $a_0 = .7$ and $b_0 = .69$, then $r_0 = .046$. So, according to Theorem 2, the posterior probability of $\mathcal{T}'$ is greater than the posterior probability of $\mathcal{T}$, even if the models are only minimally confirmed by their corresponding evidence. At the same time, $\mathcal{T}'$ is better confirmed by $\mathcal{E}'$ than $\mathcal{T}$ is by $\mathcal{E}$. Hence, we have given an epistemic justification for adding more and more models to a scientific theory in the course of normal science.

[7]Note that this is an empirical claim that needs to be supported by case studies. The tricky part of the argument is, of course, to relate the intuitive notion of coherence with the form notion. Clearly, much more needs to be said about this.

7 Naturalized Bayesianism

After having presented two examples of the modeling methodology in philosophy of science, it is time to step back and lay out the general philosophical position behind this work. I shall call this position *Naturalized Bayesianism.* It combines elements of *Textbook Bayesianism* with findings from the naturalistic approach to the philosophy of science. From *Textbook Bayesianism* it takes the normative framework, which is, however, insufficient to get a full philosophical account of scientific theorizing. It is insufficient just as Newton's three laws are insufficient in mechanics. They have to be supplemented by additional elements, such as force functions, if we want to account for mechanical phenomena in Newtonian Mechanics.

Likewise, the normative framework of *Textbook Bayesianism* has to be supplemented by empirical elements. These additional (modeling) assumptions are typically inspired by generalizations from case studies, which makes *Naturalized Bayesianism* a philosophical project that requires the competence and the collaboration of philosophers and historians of science with a background in formal methods as well as knowledge in the history of science and of contemporary science. Building on this combined competence, naturalized Bayesians can attempt to explain the regularities that we find in the methodology of science.

8 Conclusions

I hope to have shown that the method of modeling has some value in philosophy. Two functions of modeling are especially worth mentioning: (i) Models are heuristically important. They suggest something that we might later be able to explain in a model-independent way (as the variety-of-evidence example shows); (ii) Models help us to explain features of science. My discussion of scientific theory change is a case in point. Models help us to deal with more realistic (and, as it happens, more complicated) situations. When different intuitions pull in different directions (as in the variety-of-evidence example), the philosophical model tells us which intuition "wins" in which part of the parameter space.

While there are many similarities between modeling in philosophy and modeling in science, there are also important differences. Most important is the issue of model assessment. Why should we believe in a philosophical model, accept its consequences and be content with an explanation that a philosophical model provides us with? After all, unlike in science, we cannot compare the model with "hard" data from experiments. All we have in philosophy is our intuitions, and all we can do is to aim at a *reflexive equilibrium* between our intuitions and the consequences of the model. That is, on the one hand, the model may correct our intuitions and make them more precise. On the other hand, our intuitions may suggest a correction of a philosophical model. Clearly, much more needs to be said about the assessment of philosophical models, but I have to leave this for another occasion.

I'd like to close with a final word about Bayesianism. *Naturalized Bayesianism*, or so I have argued, can explain some features of the methodology of science. But can it explain all of its features, i.e. is *Naturalized Bayesianism* a universal philosophy of science? Probably not. But it remains to be seen how far one gets with this extremely minimal and conceptually simple framework. And once we come across features of science that cannot be fit into the Bayesian framework, we have to find another account.[8] Just as in science.

Appendix

To proof theorem 2, we first define $a_0^* := P(\mathrm{M}_1, \dots, \mathrm{M}_{n-1}|\mathrm{E}_1, \dots, \mathrm{E}_{n-1})$ and $b_0^* := P(\mathrm{M}_1, \dots, \mathrm{M}_n|\mathrm{E}_1, \dots, \mathrm{E}_n)$. From eqs. (10) and (11) we obtain

$$c_r(\mathcal{T}) = \frac{a_0 + (1-a_0)(1-r)^{n-1}}{a_0} a_0^* \quad ; \quad c_r(\mathcal{T}') = \frac{b_0 + (1-b_0)(1-r)^n}{a_0} b_0^*$$

As $\mathcal{T}'$ is more coherent than $\mathcal{T}$, $c_r(\mathcal{T}') > c_r(\mathcal{T})$ for all $r \in (0,1)$. Hence

$$\frac{b_0^*}{a_0^*} > \frac{a_0 b_0 + (1-a_0) b_0 (1-r)^{n-1}}{a_0 b_0 + a_0(1-b_0)(1-r)^n}.$$

We conclude that $b_0^*/a_0^* > 1$ if $(1-a_0)b_0 > a_0(1-b_0)(1-r)$, i.e. if

$$r > r_0 := \frac{a_0 - b_0}{a_0(1-b_0)},$$

which proofs the first part of the theorem.

In our notation, the distance measure d and ratio measure r of confirmation are defined as follows:

$$\begin{aligned} d(\mathcal{T},\mathcal{E}) &:= a_0^* - a_0 \quad , \quad d(\mathcal{T}',\mathcal{E}') := b_0^* - b_0 \\ r(\mathcal{T},\mathcal{E}) &:= a_0^*/a_0 \quad , \quad r(\mathcal{T}',\mathcal{E}') := b_0^*/b_0 \end{aligned}$$

[8] Some think that the Bayesian framework is too flexible and that everything can be fit into it. I doubt that this is true. The philosophical community will not accept every model. There are standards of model acceptance – just as in science – and it is our task to make these standards explicit.

Hence,

$$\begin{aligned} d(\mathcal{T}', \mathcal{E}') - d(\mathcal{T}, \mathcal{E}) &= (b_0^* - b_0) - (a_0^* - a_0) \\ &= (b_0^* - a_0^*) + (a_0 - b_0) \end{aligned}$$

and

$$\begin{aligned} r(\mathcal{T}', \mathcal{E}')/r(\mathcal{T}, \mathcal{E}) &= \frac{b_0^*/b_0}{a_0^*/a_0} \\ &= \frac{b_0^*}{a_0^*} \cdot \frac{a_0}{b_0} . \end{aligned}$$

From $b_0^* > a_0^*$ (for $r > r_0$) and $a_0 > b_0$, we obtain that $d(\mathcal{T}', \mathcal{E}') > d(\mathcal{T}, \mathcal{E})$ and $r(\mathcal{T}', \mathcal{E}') > r(\mathcal{T}, \mathcal{E})$.
The likelihood measure l is defined as follows:

$$l(\mathcal{T}, \mathcal{E}) := \log \left(\frac{P(\mathrm{E}_1, \ldots, \mathrm{E}_{n-1} | \mathrm{M}_1, \ldots, \mathrm{M}_{n-1})}{P(\mathrm{E}_1, \ldots, \mathrm{E}_{n-1} | \neg\mathrm{M}_1, \ldots, \neg\mathrm{M}_{n-1})} \right)$$

After some algebra, $l(\mathcal{T}, \mathcal{E})$ can be written in the following form:

$$l(\mathcal{T}, \mathcal{E}) = \log\left(a_0^*/(1 - a_0^*)\right) - \log\left(a_0/(1 - a_0)\right)$$

A similar expressions obtains for $l(\mathcal{T}', \mathcal{E}')$. We now note that $f(x) := \log[x/(1-x)]$ is an increasing function of x on the interval $(0, 1)$ and obtain that $l(\mathcal{T}', \mathcal{E}') > l(\mathcal{T}, \mathcal{E})$, if $b_0^* > a_0^*$ and $a_0 > b_0$. This completes the proof of theorem 2.

Acknowledgement

It is a pleasure to dedicate this paper to Patrick Suppes with whom I had many inspiring discussions over the last ten years. Earlier versions of this paper were given at various universities in Australia, Europe, the UK, and the US, and I'd like to thank the audiences for their valuable feedback and criticism. Special thanks go to Max Albert, Horacio Arlò-Costa, Luc Bovens, Philip Dawid, John Earman, Pascal Engel, Ludwig Fahrbach, Roman Frigg, Alan Hájek, Colin Howson, Jenann Ismael, Sebastian Lutz, Kevin Mulligan, Fiona McPherson, Margaret Morrison, John D. Norton, Huw Price, Paul Teller and Neil Tennant.

References

Bovens, L. and S. Hartmann (2003). Solving the Riddle of Coherence, *Mind* 112, 601–634.

Bovens, L. and S. Hartmann (2004). *Bayesian Epistemology.* Oxford: Oxford University Press.

Dietrich, F. and L. Moretti (2005). On Coherent Sets and the Transmission of Confirmation. *Philosophy of Science* 72(3): 403-424.

Dorling, J (1979). Bayesian Personalism, the Methodology of Research Programmes, and Duhem's Problem, *Studies in History and Philosophy of Science* 10, 177-87.

Earman, J. (1992). *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory.* Cambridge, MA: MIT Press.

Frigg, R. and S. Hartmann (2005). Models in Science. In: *The Stanford Encyclopedia of Philosophy* (Spring 2006 Edition).

Gähde, U. (1996). Anomalies and the Revision of Theory-Nets. Notes on the Advance of Mercury's Perihelion. In: M.L. Dalla Chiara et. al. (eds.), *Structures and Norms in Science.* Dordrecht: Kluwer.

Giere, R. (1988). *Explaining Science: A Cognitive Approach.* Chicago: University of Chicago Press.

Hartmann, S. (1999). Models and Stories in Hadron Physics, in: Morgan and Morrison (1999), 326–346.

Hartmann, S. (2001). Effective Field Theories, Reduction and Scientific Explanation, *Studies in History and Philosophy of Modern Physics* 32B, 267-304.

Hartmann, S. (2002). On Correspondence, *Studies in History and Philosophy of Modern Physics* **33B**, 79–94.

Howson, C. and P. Urbach (2006). *Scientific Reasoning: The Bayesian Approach.* Open Court Publishing 2006.

Kuhn, T.S. (1970). Logic of Discovery or Psychology of Research? In: Lakatos and Musgrave (1970), pp. 5-23.

Lakatos, I. and A. Musgrave, *Criticism and the Growths of Knowledge.* Cambridge: Cambridge University Press.

Lauritzen, S.L. and D.J. Spiegelhalter (1988). *Local computations with probabilties on graphical structures and their application to expert systems.* J. Royal Statistical society B, 50:154-227.

Morgan, M. and M. Morrison (1999). *Models as Mediators.* Cambridge: Cambridge University Press.

Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* San Mateo, CA: Morgan Kaufman Publishers.

Popper, K. (1970). Normal Science and its Dangers. In: Lakatos and Musgrave (1970), pp. 51-58.

Randall, L. (2005). *Warped Passages: Unraveling the Universe's Hidden Dimensions.* Ecco.

Salmon, W.C. (1990). Rationality and Objectivity in Science, or Tom Kuhn Meets Tom Bayes. In: C. W. Savage (ed.), *Scientific Theories.* Minneapolis: University of Minnesota Press, 175–204.

Skyrms, B. (1996). *Evolution and the Social Contract.* Cambridge: Cambridge University Press.

Skyrms, B. (1999). *Choice and Chance: An Introduction to Inductive Logic.* Wadsworth Publishing.

Spirtes, P, C. Glymour, and R. Scheines (2000). *Causation, Prediction, and Search.* Cambridge, MA: MIT Press.

Suppes, P. (1960). A Comparison of the Meaning and Uses of Models in Mathematics and the Empirical Sciences, *Synthèse* 12: 287-301.

Suppes, P. (1962). Models of Data, in E. Nagel, P. Suppes and A. Tarski (eds.), *Logic, Methodology and Philosophy of Science: Proceedings of the 1960 International Congress.* Stanford: Stanford University Press, 252-261.

Suppes, P. (2002). *Representation and Invariance of Scientific Structures.* Stanford: CSLI Publications.

Worrall, J. (2000). Kuhn, Bayes and "Theory-Choice": How Revolutionary is Kuhn's Account of Theoretical Change? In: R. Nola and H. Sankey (eds): *After Popper, Kuhn and Feyerabend: Recent Issues in Theories of Scientific Method*, 25-51.