Reasoning and Argumentation in Science

PI: Prof. Stephan Hartmann

Host institution: Center for Advanced Studies, LMU Munich

Reasoning and argumentation are essential for the progress of science. Scientists draw inferences from hypothesis, they extrapolate data, and they aim at convincing others with their ideas and insights. There are different ways to study scientific reasoning. One way to proceed is to study scientific reasoning and argumentation descriptively. One can conduct case studies, and one can try to extrapolate from case studies and identify patterns of reasoning and argumentation. Many such patterns are captured by the various types of deductive and inductive reasoning. We infer, for example, from the law that copper conducts electricity that this piece of copper in front of me conducts electricity. This is a deductively valid argument. We may also infer from the observation that various tested pieces of copper conduct electricity that copper *always* conducts electricity. This is an inductive inference, for which David Hume challenged us to find a rational justification. While deductive and inductive inferences are important in science (see Adler and Rips 2008), they do not exhaust the inferences used in science. Other types of reasoning and argumentation are at least equally important. One very popular type of reasoning is the inference to the best explanation (Lipton 2004). Consider the following example from ordinary reasoning. You leave a piece of cheese on the kitchen table in the evening. On the next morning, you observe a few crumbles of cheese and a little hole on the bottom of the kitchen wall. The best explanation for these observations is that a mouse visited the kitchen last night and you infer, by an inference to the best explanation (IBE), to the truth of this hypothesis. There seems to be nothing wrong with this. Note, however, that this inference is neither a deductive inference nor an inductive inference. It is a different type of reasoning. But is it good reasoning? Does it always work? These questions are all the more pressing as IBE is a very popular type of reasoning in the sciences. For example, no one has ever seen an electron. And yet, electrons figure in our best explanations for certain phenomena. The assumption of electrons arguably yields the best explanation, and so one infers, via IBE, that electrons exist. Such an inference seems to be much more problematic than the inference in the cheese example (van Fraassen 1990, Douven 2011) which suggests that certain types of reasoning work only well under certain conditions or in certain contexts. What is therefore needed is a normative study of reasoning and argumentation. While theories of deductive and inductive reasoning and argumentation are well developed, reasoning and argumentation patterns that are not of this kind raise problems and challenges. This project therefore has the following aims, combining descriptive and normative approaches in a hopefully fruitful way: (1) identify certain reasoning and argument patterns beyond inductive and deductive reasoning. (2) Develop a normative theory that goes beyond inductive and deductive reasoning. We will see that the Bayesian framework can be adopted to these cases so that no new framework is needed. (3) Relate the descriptive and normative parts with the aim of reaching a reflective equilibrium.

The project is divided into four parts. I will now describe each part, its motivation and goals as well as the expected outcomes in detail.

Project A: Reasoning and Argumentation with "Non-Empirical" Evidence

The traditional model of science works roughly like this. Scientists come up with a theory or hypothesis H. They then derive empirically testable consequences E from H and then evaluate H in light of E. If E is observed, H is confirmed (Carnap) or corroborated (Popper). If E is not observed, then H might have to be given up (and there is a long discussion about when this should be done). But what if there are no empirically testable consequences? This is the situation in fundamental physics, and it is also the situation in some parts of the social sciences, which raises the question how, if at all, these theories can be assessed. Are they just metaphysical speculations, as some authors conjecture? Or is there a way to nevertheless attribute a scientific status to them? Does science enter a new phase in the light of the absence of empirical data?

Recently, some prominent physicists including George Ellis and Joe Silk (2014) stated that physics needs a new methodology in order to deal with this situation. Indeed, standard deductivist accounts such as Popper's Critical Rationalism do not have much to say here; but this project will argue that we do not need a new methodology beyond Bayesian Confirmation Theory, which can be adapted to the new situation.

To substantiate this claim, we consider String Theory. This theory is extremely ambitious: It promises to unify all fundamental forces of Nature and to be the theory of everything. However, it is difficult, if not impossible (at least at the moment), to derive empirically testable consequences from it. And yet, scientists would like to assess String Theory in a scientific way. What can be done? Is there a way to assess String Theory (or any other such theory) without reference to empirical data? Is there something like "non-empirical" evidence for a theory? One popular way to proceed is the following: Try to find alternatives to H which also satisfies the various conditions H satisfies. In the case of String Theory, it turned out that scientists could not find such an alternative, despite a lot of energy and effort. This observation (which we call F), i.e. the observation that the respective scientific community has not (yet) found an alternative despite a lot of effort, has then been used as one reason in favor of String Theory. We ask: Is this proper reasoning? Are physicists justified in taking the meta-observation F as evidence for

String Theory? In a previous publication, Dawid, Hartmann and Sprenger (2015) provided an assessment of this so-called No Alternatives Argument (NAA) in the framework of Bayesian Confirmation Theory and studied under which conditions the NAA is a good argument.

Project A builds on this work and extends it in two directions. First, we will examine two detailed case studies and analyze them from the point of view of the above-mentioned formal analysis (A1 and A2). Second, it turns out that the formal machinery developed in Dawid, Hartmann and Sprenger (2015) can be used to analyze other argument patterns (such as IBE), which have a very similar formal structure. Our plan is to do just this.

A1. The No Alternatives Argument: Case Studies from Physics

The formal analysis of Dawid, Hartmann and Sprenger (2015) makes a number of presuppositions, which need to be checked for concrete cases. For example, it makes the empirical assumption that scientists have beliefs about the number of alternative theories. This may be true or false, and it therefore needs to be checked. The success of the NAA depends on it. It also depends on a number of other details, and the plan of this subproject is to look at one or two case studies in more detail. Our first case study will be String Theory (see, e.g., Cappelli et al. 2012), which is the most important application of the NAA. We will also look at the Standard Model and the Higgs boson. The Higgs and its corresponding mechanism were conjectured in 1964 and scientists accepted it for several decades, supported by the NAA, until the Higgs was finally discovered in 2012. We will examine this case in detail and confront it with our formal analysis.

A2. The No Alternatives Argument: Case Studies from the Social Sciences

We suspect that NAA reasoning also plays a role in the social sciences and we would like to focus on a case study from this field. We have a number of ideas, but we have not yet decided in which direction to go.

A3. The No Alternatives Argument and Related Argument Patterns

It turns out that IBE and the NAA have a very similar formal structure and we would like to use our results about the NAA to learn something about the validity of IBE. Hence, we will provide a formal analysis of IBE which will illuminate why IBE works fairly well in ordinary reasoning contexts (such as the above-mentioned cheese example) and why it is much more controversial in scientific context where the notorious underdetermination thesis is a real threat for IBE. Again, we will have to take beliefs about the number of alternative theories seriously. A further application of our formal model concerns the reasoning pattern “no good reason for X is a good reason against X." This pattern is used, for example, in the philosophy of religion: Hanson (1971), for example, examines all proposed proofs of the existence of God. It turns out that he finds non of them convincing and he then concludes that this cumulative failure is a reason against the existence of God. We ask: Is this a good argument? Is it good reasoning? Our analysis will shed new light on this.

Project B: Model-Based Reasoning

Theoretical models play an enormous role in science, and they are used for many different purposes (Frigg and Hartmann 2012). Here we focus on models as a reasoning tool. More specifically, we investigate (i) *analog models*, which are currently enormously popular in fundamental physics, and (ii) *toy models*.

Analog models work as follows: We consider, as in Project A, a theory H that cannot be tested empirically. It now turns out that the mathematical structure of H is analogous (or similar) to the mathematical structure of a theory H' that can be tested empirically and ask: Is it possible to infer from the confirmation of H' that H is also confirmed? This move is made, for example in the physics of black holes (see below). In previous work, Dardashti et al. (2015) gave a first account of this case and the present project continues this research in several directions.

Toy models (such as Schelling's model of segregation) are highly idealized models, which typically do not fare well when it comes to empirical testing. This makes them something like a mystery as it is not obvious why they are used at all. And yet, scientists use toy models quite intensively. They claim that toy models provide them with understanding, and they use toy models as a reasoning tool. We will examine these issues (and their justification) in detail and on the basis of case studies.

B1. Reasoning with Analogic Models: Case Studies

We consider analogue experiments and black holes. In 1975 Stephen Hawking proposed that black holes are actually not black at all but radiate. Since black holes are empirically inaccessible and the reliability of the theory on which Hawking's prediction relies is hard to assess, the important question whether black holes do radiate remained unanswered. Based on a proposal by Unruh (1981) scientists have recently built table-top experiments which are to some extent analogous to black holes and have shown, or so they argue, that black holes do radiate thermally. Dardashti et al. (2015) showed that the observation of "Hawking radiation" in these analogue models may provide evidence for the radiation in the black hole case and we will reconsider this case and, time permitting, related cases in detail and, as in Project A1, in the context of our formal analysis provided in B2.

B2. Reasoning with Analogic Models: Confirmation

On the basis of the case studies examined in B1, we will examine the argumentative structure of analog reasoning in fundamental physics within Bayesian Confirmation Theory. Our methodology is inspired by the methodology developed in Dizadji-Bahmani, Frigg and Hartmann (2011), which focuses on a Bayesian analysis of intertheoretic reduction.

B3. Reasoning with Toy Models

This project will address the following questions: (i) How, if at all, do toy models provide understanding? We will do this by confronting toy models with theories of understanding such as the ones developed by Dieks and de Regt and by Strevens (see Frigg and Hartmann (2012)). A critique of these accounts will lead us to our own theory, which will be a modification of the theory of Strevens. (ii) Can toy models be confirmed? If so, how? Can one develop a formal theory of confirmation of highly idealized models? Note that this is a challenge for Bayesian Confirmation Theory as toy models involve false assumptions and false assumptions have a prior probability of zero, which cannot be updated to a non-vanishing value. (iii) How are toy models used as a reasoning tool (Magnani 2014), and how is this justified?

Project C: A Bayesian Theory of Reasoning and Argumentation

Projects A and B used already the Bayesian framework (for surveys, see Hartmann and Sprenger (2010) and Hajek and Hartmann (2010)) to address a number of methodological questions that show up in scientific reasoning and argumentation. This project provides a full-fledged Bayesian theory of argumentation (B2), which we will also test in psychological experiments (B3). Before, however, we have to better understand conditional information that plays a crucial role in argumentation. Every new piece of information can affect an agent's previous beliefs: some old beliefs may have to be withdrawn, some strengthen, and some new propositions may need to be accepted. Even though a lot of what we learn is conditional in form, it is not clear how exactly agents should respond to new conditional information. The main objective of Project C is to investigate what learning conditional information amounts to, and to verify various theoretical accounts against the empirical data.

C1. Learning Conditional Information

One of the problems any account of updating on conditionals faces is that our intuitions on how agents adapt their beliefs in response to conditional information seem to vary from case to case (cf. Douven 2012). We will investigate possible factors that can influence the way agents update their beliefs upon learning a conditional. Psychological research on Bayesian argumentation (e.g. Hahn and Oaksford 2007) suggests that the effect of a conditional on an agent's beliefs depends, among others, on their prior degrees of belief in the conditionals' consequent. Moreover, as shown in Krzyżanowska et al. (2013, 2014), conditionals do not make a homogeneous class, but they can be analyzed as corresponding to deductive, inductive, or abductive arguments. Agents may respond differently to conditionals of different inferential types. We will conduct an experimental study to investigate how prior beliefs and types of conditional information affect participants' posterior degrees of belief.

C2. A Bayesian Theory of Argumentation

Argumentation is an important interdisciplinary topic of study and much progress has been made over the last decades. For an extensive recent survey, see the monumental (van Eemeren et al. 2014). The central idea of our new theory is as follows: We consider an argument to be a set of assumptions (often, but not always, involving a conditional), which jointly make the conclusion of the argument more likely. The premises provide reasons for the conclusion, and the task, then, is to come up with a theory that accounts for this and makes it more precise.

In line with the Bayesian approach we are promoting in this project, we proceed as follows: Let us consider a single agent who has prior beliefs about a number of propositions. These beliefs are represented by a Bayesian network model, which is the basis for the agent's further reasoning. Reasoning, or so we argue, is always relative to a model. Next, the agent learns the premises of the argument from some (more or less reliable) information source. For example, the agent may learn that a certain proposition is true or that a certain conditional information obtains. To integrate this new information into her or his belief set, the agent updates her prior probability distribution to a posterior distribution leaving the causal structure of the Bayesian network unchanged. How the update works precisely is a matter of debate. We argue that it cannot be decided top-down, but requires empirical studies. One way to proceed is by minimizing a distance measure between the posterior and the prior probability distribution such as the Kullback-Leibler divergence. There are, however, alternative measures and we want to explore in empirical studies, which measure works best.

C3. Testing the Theory

This project tests the above outlined theory in various scientific and non-scientific scenarios.

Project D: Collective Reasoning and Argumentation

The projects A, B and C considered an individual agent who reasons and argues. Science, however, is a collective enterprise and the scientific community plays an important role in the complex and involved reasoning and decision-making processes that occur in real science. This has already been observed and stressed by Thomas Kuhn in *The Structure of Scientific Revolutions* and a lot of work on the social studies of science and on social epistemology has confirmed and elaborated this point. Hence, a theory of reasoning and argumentation in science would be incomplete without a consideration of at least some collective aspects. To start with, we propose to work on the following three subprojects, building on our earlier work.

D1. Deliberation in Science

In Hartmann and Rafiee Rad (under review, 1), we developed a model of deliberation of a group of epistemic agents, which has to make a yes-no decision. (Our example was a jury in court.) In this subproject, we want to extend this account in the following ways: (i) The group does not only have to make a yes-no verdict, but agree on a number, such as a probability. (ii) The group has to make a decision on logically interrelated propositions. That is, we want to extend our model to judgment aggregation (List and Polak 2010).

D2. Correcting Biases in Deliberation Processes

In Hartmann and Rafiee Rad (under review, 2), we constructed and analyzed a simple model of the anchoring effect in a group of (boundedly) rational agents. It turns out that the effect also occurs in such groups and that the agent who speaks first has the highest impact on the resulting group decisions. This is an unwanted effect and we would like to device simple procedures that can be easily implemented and get rid of the effect. We have already some ideas and would like to test them in computer simulations.

D3. Value Aggregation and Theory Choice

Choosing a theory typically involves different value judgment as Thomas Kuhn noted already long ago. It therefore does not come as a surprise that Arrow-style impossibility results can be derived (Okasha 2011). We ask: How can these results be avoided, and how can all this be captured and discussed in Bayesian Philosophy of Science?