Modeling Scientific Theory Change

Stephan Hartmann
London School of Economics

Motivation

- There are many different attempts to understand theory change in science, ranging from formal reconstructions to analyses of case studies from the history of science and from current science.
- Main thesis: A philosophical understanding of theory change requires a combination of formal (normative) and historicist (descriptive) elements.
- I will show how this can be achieved by using *philosophical modeling*.

Overview

I. Understanding Theory Change

- In this section, I shall contrast (idealized and over-simplified) versions of the formalist and of the historicist traditions to account for theory change in science.

1. The Formalist Tradition

- What is it?
  - give a formal (logical, probabilistic, etc.) reconstruction of theory change
- What's good?
  - general
  - normative
- What's bad?
  - too far away from the practice of science

2. The Historicist Tradition

- What is it?
  - examine case studies from the history of science
- What's good?
  - good to falsify bad formal (normative) accounts (e.g. Popper on normal science)
- What's bad?
  - does not lead to a normative account

3. Desiderata

- What do we expect from a philosophical account of theory change?
  - normative: provide a defensible general account of scientific rationality.
  - illuminate "intuitively correct decisions and judgments made in the history of science and explain the incorrectness of those judgments (…) that seem clearly intuitively incorrect (and sheds light on 'grey cases')". (J. Worrall)
- How can this goal be approached? Modeling

II. Interlude: Modeling in Science and Philosophy

- Observation: Models are very popular in the sciences.
- Aims of this talk: Demonstrate that modeling techniques can also be used to tackle problems of philosophical interest.
- The method: Probabilistic modeling (esp. Bayesian Networks)
- The problems: Epistemology, political philosophy, and philosophy of science

Shift from Theories to Models

- Features of theories
  - Examples: Newtonian and Quantum Mechanics
  - General and unversal in scope
  - Abstract
  - No idealizations involved (ideally...)
  - Modeling framework
- Features of models
  - Examples: *models of a theory* (e.g. pendulum), *phenomenological models* (e.g. the Bohr model)
  - Specific and limited in scope
  - Concrete
  - Involve idealized assumptions

Why this Shift?

- Distrust general theories
- Model are often needed to apply a theory
- Models can be solved
- Models help us to solve problems (think where we were if people would have addressed foundational problems of Newtonian Mechanics before applying it?!)
- Models are more intuitive
- …

Maybe it is time to also construct models in philosophy…

III. Bayesian Network Models

- There is a plurality of modeling tools on the market (and this is good so).
- Focus here on probabilistic modeling as it seems to be especially suited for the problems I am interested in.
- To do so, I adopt Bayesianism as a framework and construct models within it ("models of a theory").

Textbook-Bayesianism

- Hypothesis *H*, evidence *E*
- Confirmation = positive relevance between *H* and *E*
- Probabilities reflect subjective degrees of belief
- Update-rule: $P_{new}(H) = P(H|E) = P(E|H)\,P(H)/P(E)$
- Direct consequences of the update-rule:
  - surprising evidence confirms better
  - variety-of-evidence is a good thing
- Compare the update-rule with Newton's 2nd Law (i. e. F = m · a): (*i*) Some results can be derived from it, but many interesting ones only obtain when one constructs models. (*ii*) Don't worry too much about foundations!

Textbook-Bayesianism (cont'd)

- Textbook Bayesianism may be able to illuminate some typical features of the scientific practice.
- It has also been used (e.g. by Jon Dorling) to reconstruct episodes from the history of science.
- However, textbook Bayesianism is often "too far away" from real science, and does not seem to provide an account for features such as theory change, or so I argue.
- By constructing models, on the other hand, the gap between a general theory (i.e. Bayesianism) and the scientific practice can be bridged.

How does one do this?

- I will use Bayesian Networks to construct philosophical models.
- Bayesian Networks were developed in the 1980ies in computer science (by Judea Pearl and others) to represent and manipulate probability distributions over a large number of variables (e.g. for expert systems)
- But what is a Bayesian Network?

An Example from Medicine

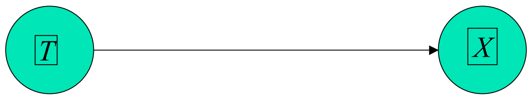
- T: Patient has tuberculosis
- X: Positive X-ray
- Given information:

  t := P(T) = .01
  p := P(X|T) = .95 = 1- P(¬X|T) = 1- rate of *false negatives*
  q := P(X|¬T) = .02 = rate of *false positives*

- What is P(T|X)? ⇒ Apply Bayes' Theorem

  P(T|X) = P(X|T) P(T) /[P(X|T) P(T) + P(X|¬T) P(¬T)] =
  = p t /[p t + q (1-t)] = t/[t + (1-t) x] with x := q/p
  = .32

A Bayesian Network Representation



P(T) = .1

P(X|T) = .95

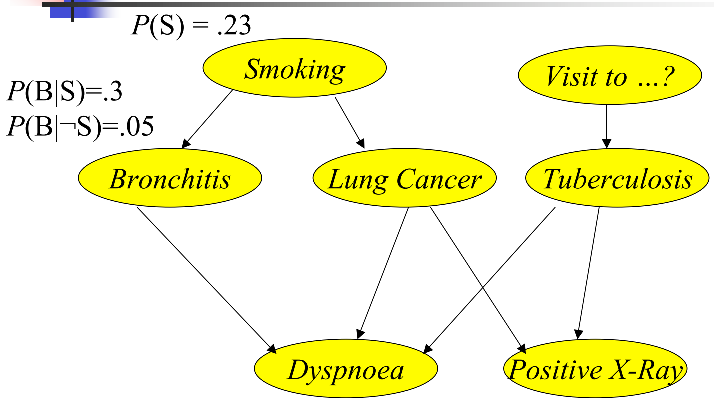
P(X|¬T) = .2

Parlance:

"*T* causes *X*"

"*T* directly influences *X*"

A More Complicated (=Realistic) Scenario



Bayesian Networks

- A BN is a *directed acyclic graph* (DAG) with a *probability distribution* defined on it.
- BNs represent conditional independencies that hold between (sets of) variables, which helps reducing the number of probabilities that have to be fixed to specify the joint probability distribution over all variables.
- Specifying the joint probability distribution over n variables requires the specification of 2^n-1 numbers. However, in a BN one only has to specify the probabilities of each node given the values of its parents: *P(A|par(A)).*
- There are efficient algorithms to compute whatever probability one is interested in.

How to Construct a BN Model

- Fix the set of (relevant) variables.
- Specify the (probabilistic) independencies that hold among them.
- Construct the corresponding Bayesian Network and fix a probability distribution.
- Calculate the wanted probabilities.

Role and Scope of BN Models

- Aim at a medium level of generality
- Explicate typical features of the scientific practice:
  - not too general (or far away from science), and
  - not too specific (à la Dorling)
- Models connect the general framework (viz. Bayesianism) with generalizations and taxonomies obtained from case studies.
- The historicist (or descriptivist) approach is needed to help us learn about what these features are.

What is a Scientific Theory?

- Textbook Bayesianism has no account of what a scientific theory is. This is a shortcoming.

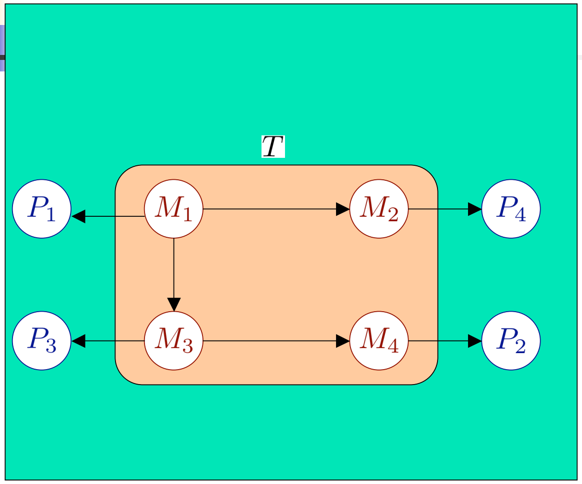
Formal Characterizations of Theories

- Syntactic view:
  - linguistic entities
  - sets of assumptions (and their consequences)
- Semantic view:
  - non-linguistic entities
  - realizations of an abstract formalism
- Probabilistic view:
  - theories are networks of interrelated models, and models are conjunctions of propositions

The Probabilistic View

- Theories are networks of interrelated models.
- Models (M_i) are conjunctions of propositions that account for a specific phenomenon P_i. One model for each phenomenon.
- There is a joint probability distribution over all propositional variables M_i, P_i.
- From this, the posterior probability of the theory (given the phenomena) can be obtained.

Representing Theories by Bayesian Networks



IV. Modeling Theory Change

- I'll present *two examples* that (hopefully) show how modeling methods can be applied to illuminate typical features of "real" theory changes in science.
- There is already some work to Bayesianize Kuhn's criteria for theory choice. In *Bayesian Epistemology* (OUP 2003, with Luc Bovens), for example, we focus on Kuhn's consistency criterion which we explicate as the (internal) coherence of a theory.

1. The Stability of Normal Science

- In normal science, more and more applications of a theory are considered. The goal, however, is not to test the theory. It is taken for granted and applied.
- An immediate problem for the Bayesian:

  - $P(A, B, C) < P(A, B)$

  - so by adding more and more applications of a theory, the joint probability will sooner or later be below the threshold (if there is any). I refer to this as the *conjunction problem*.
- So can normal science be defended in a Bayesian framework?

Coherence

- The conjunction problem can be (dis-)solved by taking evidence into account. Then it is possible that

  $P(A, B, C|E_A, E_B, E_C) > P(A, B|E_A, E_B)$
- What are the determinants of the posterior probability?
- According to *Bayesian Coherentism*, the coherence of an information set (or theory) is such an indicator.
- Within normal science, the theories and models belonging to a certain paradigm become more and more coherent.
- It can be shown, that under certain assumptions coherence becomes the exclusive criterion of theory choice.

2. Why Anomalies Hurt so Much

- The theories and models in normal science are highly coherent.
- But what is coherence?
- In recent work, Dietrich & Moretti (forthcoming) showed that sufficiently coherent sets of propositions transmit confirmation, i.e. if *E* confirms one of the propositions of a theory, it also confirms any other proposition as well as the whole theory, i.e. the conjunction of all propositions.
- The same holds, of course, for disconfirmation.
- So an anomaly hurts the whole theory in normal science, and this can be fatal for it.

More Features to Explain?

- Here more case studies are needed…
- Current research project: Use the ideas presented in this talk to formulate an account of *intertheory relations* that avoids the extremes - reductionism à la Nagel and an untamed pluralism.
- Examine case studies and extract features that survive theory change, and those that do not, and try to explain this in a formal model.
- Hypothesis: Often more than structure survives…

V. Methodological Reflections

- Here are some reflections about the value of the methodology I suggested.
- Note: I will *not* address standard arguments against Bayesianism (see my remarks about Newtonian Mechanics before).

1. Philosophical Modeling

- Philosophical modeling works well for "on the one hand, on the other hand"-situations. Which "force" is stronger will then depend on the specific parameters.
- Real situations are typically like this!
- Mathematics helps us to draw conclusions from complex assumptions.
- The philosophical model acts as a midwife - it helps us find and establish qualitative claims that have to be understood independently of the model.
- Philosophical models are *toy models*.
- Acceptance criteria for a philosophical model?

2. The Methodology (in a Nutshell)

1. **Problem Specification**: Formulate a philosophical problem in ordinary language.
2. **Model Construction**: Choose a modeling framework and make modeling assumptions which suit the problem at hand.
3. **Translation**: Translate the problem into a question which can be posed within the mathematical model.
4. **Deduction**: Obtain an answer to this question by deduction within the model.
5. **Back-Translation**: Translate this answer back in ordinary language.
6. **Interpretation**: Give a *model-independent* explanation of the results of the model.