FEW 2007, CMU, Pittsburgh

A confirmation-theoretic guide to explanation

Jan Sprenger, University of Bonn

I. Explanation and Confirmation

Motivation: The history of confirmation theory

- Modern confirmation theory started with **syntactic and qualitative accounts**, e.g. Hempel´s satisfaction criterion, hypothetico-deductive confirmation and bootstrap confirmation (Glymour)
- However, all these accounts fail to give a **comparative** account of confirmation
- None of these accounts is **complete** in the sense that all aspects of theory confirmation are captured

The solution: A Bayesian approach

- Epistemic interpretation of probabilities (as rational degrees of belief)
- **Evidential relevance**, defined by

  $P(T|E.K) > P(T|{\sim}E.K)$

  is picked out as the subject of investigation!

  [T = theory, E = evidence, K = background knowledge]
- Decouples a precise and formally tractable concept (evidential relevance) from a general and vague concept (confirmation)

The benefits of Bayesian confirmation theory

- More powerful modelling tools, better chances to tackle paradoxes of confirmation
- Measures of evidential relevance enable us to quantify confirmatory power
- Connection to statistical measures of evidence

The introduction of Bayesian probabilities in confirmation theory has been a success story!

Classical accounts of explanation

- Most classical accounts of explanation are purely qualitative: DN-model (Hempel), argument patterns (Kitcher), etc.
- They are not able to quantify explanatory power or to compare competing explanations

Structural similarity in the debates about explanation and confirmation!

Causation accounts of statistical explanation

- Redundancy of probabilistic analyses of causation => no clear insight into explanation either?
- Lack of operational verification of an causal/explanatory relationship
- And some more problems (to be resumed later...)

Confirmation and explanation

- Explanation: both causal and non-causal accounts are not completely satisfactory
- Confirmation: The loss of specific features of qualitative accounts is more than compensated by the newly gained power of Bayesian accounts
- Suggestion: replace existing accounts of explanation with a Bayesian framework?

Quantifying explanatory power

- Natural framework for quantification: probabilities as rational degrees of belief (Bayesianism)
- A probabilistic approach to the degree of explanatory power is particularly helpful for **statistical explanation**

Statistical explanation naturally possesses a probabilistic dimension!

Explanatory relevance

- In the same way that confirmation makes a theory more assertable, explanation involves **rationalization** of the explanandum
- I call this (rationalizing) aspect of explanation **explanatory relevance** – the positive effect of C on E
- Rationalization can be expressed in terms of positive probabilistic relevance (= increase in subjective probability): **P(E|C.K) > P(E|~C.K)**!
  - E = explanandum, C = candidate explanation,
  - K = background knowledge

Explanation = explanatory relevance?

- I do not claim that explanatory relevance captures the essence of explanation
- Rather, I believe that a **particular aspect of explanation can be captured in Bayesian terms**
- This offers a benchmark to compare competing explanations and leads to greater precision in the evaluation of explanatory relations

Knowledge relativity

- Try to make closer links between explanation and understanding
- Thesis: **Explanation is knowledge-relative**
- E.g. for a savage other things count as an explanation than for a highly educated western physicist
- Accepting knowledge relativity of explanation would explain that consider **understanding** to be crucial for explanation.

II. Measures of explanatory relevance

The Bayesian framework

- **Knowledge relativity** implies conditionalization on the background knowledge *K*
- **Explanatory relevance** (rationalization) implies that conditional on K, the explanans is positively (probabilistically) relevant to the explanandum:

  $P(E|C.K) > P(E|{\sim}C.K)$

  - E = explanandum, C = candidate explanation

Explanatory relevance

- What is the best **measure of explanatory relevance** r?
- r(C, E, K) > 0 if and only if C is positively relevant to E relative to K
- This still admits a lot of measures!

We have to develop a set of **adequacy conditions** that helps us to make a choice

Adequacy conditions

- **Maximality I:** C is maximally explanatory relevant to E only if (not: if and only if) $P(E|C.K) = 1$
- **Maximality II:** It is not the case that: if $P(E|C.K) = 1$ then C is maximally explanatory relevant to E
- **Non-Subjectivity:** r does not depend on the probability of the candidate explanation C, $P(C|K)$.

The maximality conditions

- $P(E|C.K) = 1$ should be necessary for maximal explanatory relevance because only then the explanandum maximally rationalized
- $P(E|C.K) = 1$ should not be sufficient because the explanandum could have been very probable anyway. In such cases we would not say that the candidate explanation C is particularly **relevant** to the explandum E.
- Compare this to cases where $P(E|{\sim}C.K)$ is low.

The non-subjectivity condition

- The „catchall problem" is looming – the calculation of $P(E|\sim C.K)$ requires prior probability assignments!
- Given non-subjectivity, does explanatory relevance merely depend on the likelihoods of the explanandum under the various potential explanations?
- No – the calculation of $P(E|\sim C.K)$ requires only the relative (prior probability) weights of the explanations that compete with C. We do not need an assessment of $P(C|K)$ itself!

A catalogue of measures (I)

The following measures are intuitively plausible relevance measures:

- $r_a = P(E|C.K)$
  - (absolute explanatory relevance)
- $r_l = c_l = \log [P(E|C.K)/P(E|\sim C.K)]$
  - (log-likelihood measure)
- $r_s = P(E|C.K) - P(E|\sim C.K)$
  - (comparative difference measure)

A catalogue of measures (II)

- $r_o = \log [P(E|C.K)/(1- P(E|C.K))]$
  $- \log [P(E|K)/(1- P(E|K))]$
  - (betting odds measure)
- $r_r = c_r = \log [P(E|C.K) \,/\, P(E|K)]$
  - (log-ratio measure)
- $r_d = P(E|C.K) - P(E|K)$
  - (difference measure)

Review of the measures

- r_a admits *negative* probabilistic relevance!
- r_l violates Maximality I
- r_d, r_o and r_r are **monotonously decreasing functions** of the prior probability of the candidate explanation *ceteris paribus* (given that C is positively relevant to E at all)

$$P(E|K) = P(C|K)\,[P(E|C.K) - P(E|{\sim}C.K)] + P(E|{\sim}C.K)$$

The measure r_s – our best game in town?

- r_s satisfies both maximality constraints.
- Does not violate non-subjectivity – for the computation of $P(E|{\sim}C.K)$, we require only the **relative weights** of the candidate explanations that compete with C.
- Case studies are required to support that evaulation!

A classical example

- A person contracts paresis
- To our best knowledge, the person must have had latent untreated syphilis
- However, the chance to develop paresis with latent untreated syphilis is quite small ($P(E|C.K) = 0.25$)
- r_s recognizes the explanatory relevance of the alleged syphilis infection (but admits that it is not a perfect explanation).

The intermediate results

- Analyzing explanatory relevance (instead of ``explanation´´) paves the way for Bayesian models
- We can compare competing explanations and quantify the degree of explanatory relevance
- The shift to epistemic probabilities opens a new perspective on statistical explanation
- r_s currently the most adequate measure of explanatory relevance

III. Objections and Applications

Objections of causal theorists

- Explanatory asymmetry
- Probability-lowering explanations
- Distinguished role of causation in scientific inference
- No strucural similarity between confirmation and explanation
- Forfeit of objectivity of scientific explanations

Explanatory asymmetry

- Positive probabilistic relevance is a symmetrical relation
- „Causes explain their effects, but not vice versa“
- Flagpole example: we can infer but not explain

Explanatory asymmetry

The quantification of explanatory relevance resolves the symmetry problem in many cases of statistical explanation: r(C, E, K) is not equal to r(E, C, K).

„Coherentist“ accounts of explanation – the explanandum is a bundle of sentences whose conjunction is rendered more likely.

Problem relies on misidentification of explanandum and background knowledge. Furthermore: Failure to make type/token distinction explicit.

Probability-lowering explanations

- There may be relevant explanations that actually lower the expectation of the explanandum
- Salmon‘s classical example: radioactive decay (genuine indeterministic laws) – to be investigated later

Probability-lowering explanations

Birth control pills / thrombosis example

- We make an implicit conditionalization on the background knowledge (pregnant or not pregnant)
- Conflating conditional and unconditional explanation?
- In the unconditional case (negative net effect), the birth control pills do not enhance our understanding (and not count as an explanation either)
- Assume we cannot find out whether X is pregnant...

Scientific explanation and causation

Mere correlation (positive relevance) is not enough for scientific explanation – scientists strive for the discovery of causal relations

- Scientists start with a huge bulk of data (large number of variables)
- Systematical isolation in the search for cause and effect

Causation and correlation

- Search for cause/effect relationship between (cause) X and (effect) Y
- Elimination of confounding factors
- Usually implies a gradual increase in explanatory relevance, too!

Causation as a kind of
idealized statistical relevance!?

Objectivity of explanations

- Bayesian accounts of explanations sacrifice the objectivity of scientific explanations
- Ontic account of explanations (e.g. causal accounts): Explanations are completely subject-independent, they are elements of the natural world

This was heavily criticized by Bas van Fraassen!

Objectivity of explanations

- The meaning of an explanation-seeking why-question crucially depends on the *pragmatic interest of the subject* which determines the relevant *contrast class* for the explanandum
- The **strong objectivity thesis** (complete subject independence) is certainly too strong
- Van Fraassen: pragmatics of explanation

Objectivity in a Bayesian framework

- **Weak objectivity thesis**: conditional on the background knowledge, explanatory relevance is an objective relation
- Epistemic probability is not arbitrarily subjective probability – the weak objectivity of explanatory relevance can be rescued
- Bayesianism: „conditional objectivity“

Explanation and causation: potential problems

- Plurality of accounts of causation (probabilistic accounts, manipulative accounts, INUS accounts etc.) => plurality of accounts of explanations?
- Redundancy of analyses of causation => no clear insight into explanation either?
- Lack of operational verifiability of an explanatory relationship

Explanation and causation: further problems

- Almost homogenous populations (a single counterexample to the causal claim)
- Reference class dependency (fine-grainedness)
- Causal preemption and overdetermination (preemption of even stronger/more relevant causes)

Explanation and Prediction

- The Bayesian account draws strong parallels between explanation on the one side and prediction and confirmation on the other side
- In Bayesian confirmation theory, c_s (the confirmation-theoretic analogue to r_s) does not satisfy all reasonable adequacy constraints

Similarities present, but not to be overstressed

The virtue of unification

- Unification is often exhibited in **evidential diversity** for the unified theory

  => conditional probabilistic independence
- Bayesian confirmation theory: several independent pieces of evidence have a higher evidential relevance than any of the single pieces of evidence
- Does this lead to higher explanatory relevance, too?

Evidential diversity

- Assume that C is positively relevant to both E_1 and E_2 and that it screens off E_1 from E_2
- Then $r_s(C, E_1.E_2, K)$

  $= P(E_1 | C.K)\ P(E_2 | C.K) - P(E_1 | {\sim}C.K)\ P(E_2 | {\sim}C.K)$

  $> \max\ \{r_s(C, E_1, K), r_s(C, E_2, K)\}$

Certain aspects of unifying theories
boost the degree of explanatory relevance!

Thanks a lot for your attention!!!

IV. Does “explanatory relevance” fail to be explanatorily relevant?

A brief rejoinder to Justin Fisher‘s comment

Justin‘s concerns

- Non-subjectivity is no good basis for distinguishing between r_s and r_d (solved by modification of the non-subjectivity condition)

- The proposed explication of „explanatory relevance“ fails to be explanatorily relevant:

  (a) the explanandum may already be very likely

  (b) the correct explanations may lower the probability of the explanandum

  (c) irrelevant/inadequate explanations may also be explanatorily relevant

Explanatory relevance revisited

- First Caveat: many problems arise similarly in Bayesian confirmation theory (BCT)
- Second Caveat: type/token distinction (scientists interested in replicable explanation)
- Third Caveat: pragmatic factors affect the identification of background knowledge and explanandum

E is likely anyway

The blue sky example

- (a) Justin is interested in a type explanation, not in inductive inference of the sky being blue
- (b) Similar to BCT – take a highly expected piece of evidence
- (c) Knowledge about the sky being blue implies justification!

Unfortunate modelling!

Probability-lowering explanations

- Justin‘s example bears resemblance to Salmon‘s example
- Several kinds of explanations are conflated
- Token explanation: I bite the bullet. Above all, statistical explanation is relevant to large samples
- Type explanation: The „correct explanation“ fails if it continues to yield unexpected results

C is irrelevant

The case of the magician

- Three kinds of explananda: the actual sawing in a half, the general habit of sawing ladies in a hlaf and the mechanism behind the sawing trick
- Again, Justin‘s objection develops is force if the three explananda are not precisely distinguished

Conclusion

- It is possible to rebut Justin's objections by making explicit the distinction between the various types of explanation
- However, Justin has rightfully pointed to the underdetermination of K and E in superficial descriptions of explanation-seeking facts
- Identification of the relevant explananda and background assumptions requires the consideration of **pragmatic factors**
- Next research projects: systematization of these factors