Occam's Razor in the Theory of Theory Assessment

August Fenk
Alpen-Adria Universität Klagenfurt

31st International Wittgenstein Symposium, Kirchberg am Wechsel, August 10-16, 2008

Overview

- From the point of view of theories as hypothetical representations, with predictive success as their real touchstone, the paper argues in favour of a three-dimensional model of theory assessment, including the dimensions generality, precision, and parsimony.
- The question: Are such virtues, and especially the principle of parsimony (Occam's razor), also applicable to those meta-theories that have invented such criteria?
- The focus of the respective analysis will be on lawlikeness which is most commonly viewed as a precondition of both, prediction and anticipation as well as explanation and reconstruction. Laws turn out to be mere projections of the relative frequencies observed so far.
- Such projections can be justified - if at all, and irrespective of the weakness of the "regularity" and the number of observations - by applying some sort of Occam's razor: Do without the assumption of a change as long as you can't make out any indication that a system's output might change!

Occam's razor
lex parsimoniae or "law of parsimony"

- The most-cited version attributed to Occam (c 1285-1347/49):
  Entia non sunt multiplicanda praeter necessitatem
  "Entities are not to be multiplied beyond necessity"
- Actually to be found in Occam's writings:
  Numquam ponenda est pluralitas sine necessitate
  "Plurality ought never be posited without necessity"
- "It is vain to do with more what can be done with fewer"

Applications in Model Selection

Forster (2000): Model selection "is a relatively new branch of mathematical statistics", and all its standard methods "provide an implementation of Occam's razor, in which parsimony or simplicity is balanced against goodness of fit."

Applications in Model Selection

- These standard methods include (Forster 2000) or are based on the principle of Minimum Description Length (Grünwald 2000).
- Grünwald on MDL's philosophy:
  - "the goal of model selection is not the hunt for any true model, but rather for a simple model that gives a reasonable fit to the data;
  - both goodness-of-fit and model complexity contribute to the number of bits needed to describe the data;
  - and finally, a model that is much too complex is typically worthless, while a model that is much too simple can still be useful."

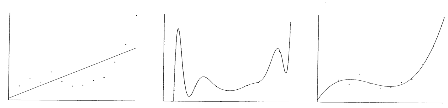


FIG. 1. A simple (a), complex (b) and a trade-off (3rd degree) (c) polynomial.

The Theory of Theory Assessment

is the philosophical counterpart of Model Selection. The differences:

- Rather abstract heuristic maximes instead of tools for coping with the stochastic complexity in given data-sets.
- Both disciplines show a trend to three-dimensional models. But while Model Selection recently tries to integrate, besides simplicity and fit, “generality” as the third dimension (Forster 2000), the role of “simplicity” is perpetually under debate in the philosophy of science.

Two-dimensional models of theory assessment

- In Laszlo’s (1972) model “empirical adequacy” is balanced against “integrative generality” which is explained as a measure of *elegance* and *neatness* (→ Occam‘s razor)
  The problem: “Empirical adequacy“ seems to be a confusion of different things.
- Coombs’ (1984) model balances “power” against “generality”.
  This model might be viewed as an appropriate decomposition of Laszlo’s “empirical adequacy”.

Coombs’ model

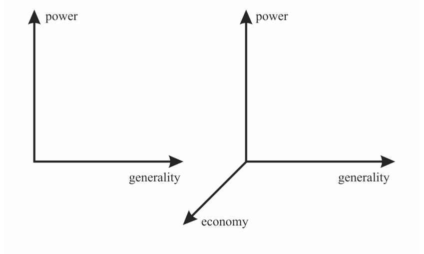
includes the idea that progress can **either** be achieved along the dimension of power **or** along the dimension of generality. For example (Fenk & Vanoucek 1992):

- Iron dilates when its temperature rises.
- Metals dilate when their temperature rises.
- Iron dilates by 10 to 15 millionth per cent by volume when its temperature rises from 18 to 19 degrees Celsius.
- Metals dilate by 10 to 15 millionth per cent by volume when their temperature rises from 18 to 19 degrees Celsius.

Problems with Coombs’ model

- One may doubt whether the either/or-principle is also applicable to more complex theories (→ periodic system)
- The model fails to account for Occam‘s razor

A completion of Coombs’ model



Economy

- It can be shown that such a model which integrates economy corresponds, after all, with three different facets of Popper’s concept of *testability*: generality (Allgemeinheit), precision (Bestimmtheit), and simplicity (Einfachheit).
- More generally, economy can be viewed as a dimension of, and thus also as a criterion for, performance and capacity – within and beyond cognitive efforts!

The Question

- Are virtues such as generality, precision, and in particular Occam's razor, also applicable to those meta-theories that have invented them?
- I now shall try doing so with the key-concepts of *LAW* and *LAWLIKENESS*.

The decomposition of *law*: the starting point

The Deductive-Nomological model by Hempel & Oppenheim (1948): The prediction or explanation is a logical consequence of the premises. The law(s) among the premises is (are) universal laws:

X is made of iron
X will be heated today at 8 pm
Iron dilates when its temperature rises

X will dilate today at 8 pm

The Problem

- The application of the D-N model is restricted to a restricted or even non-existing world of universal laws (Salmon 1971).
- In Goodman (1973) a hypothesis is lawlike only if it is projectible and projectible when and only when it is supported (some positive cases), unviolated (no negative cases), and unexhausted (some undetermined cases). But especially the criterion "unviolated" seems to be rather meant for universal laws. What should be considered the negative and the positive cases in view of a weak regularity such as a very severe side-effect of a new medicament showing in one of hundred patients in nine of ten studies?

The decomposition of *law*

- In face of a world of **statistical laws** we see a shift from the D-N model to Hempel's (1962) Inductive-Statistical model: Among the premises are statistical arguments rendering their extremely high probability ("close to 1") to the explanandum.
- The "plausability" (of predictions and explanations) is reduced to relative frequencies (Mises 1972).
- "Stable frequencies" are viewed as a sufficient basis for "objective chances" (Hoefer 2007).

Carried to the extremes

- 2, 3, 1, 4, 6, 3, 4, 5, 6, 5, 3, 1, 1, 1, 1. Next toss even or uneven?

Carried to the extremes

- 2, 3, 1, 4, 6, 3, 4, 5, 6, 5, 3, 1, 1, 1, 1. Next toss even or uneven? "Uneven!"
  If a dice had produced an uneven number in ten of fifteen cases I would, if I had to bet, bet on "uneven" for the 16th trial. For if there is a system it seems to prefer uneven numbers, and if there is none, I can't make a mistake anyway (Fenk 1992).
- 3. Next toss even or uneven? "Uneven!"
  If the "series" that had produced uneven has the minimal length of only one toss, I would again bet on "uneven".
- The only way I can see to justify such decisions is an application of Occam's razor, or a principle at least inspired by **Occam's razor**: Do without the assumption of a change as long as you can't make out any indication or reason for such an assumption!
- These arguments also apply to the relative frequency of an association (co-occurrence, succession) of the events X and Y in series of any length (→ conditional probabilities; → „transitional probabilities" in information theory).

Conclusions

- Hardly anybody would talk about laws or about relative frequencies in the example with the fifteen dices or in the 1-toss-"series" - despite an ideal "relative frequency" of 1 in this last example!
- But the examples reflect a principle as ***simple*** as ***general***: Use the slightest indication and all your contextual knowledge to optimize your decision but bet on continuity as long as you see no reason to assume that a system might change its output-pattern; generalize the data available to unknown instances!
- "Laws", "probabilities", and "objective chances" are – beyond a purely mathematical world – nice names for such generalizations and projections, usually based on large numbers of observations. But there is no lower limit regarding the strength of a regularity or the number of data available that ceases the admissibility of this way of reasoning!
- I can't resist quoting Hempel (1968:117) when he admits that "no specific common lower bound" for the probability of an association between X and Y "can reasonably be imposed on all probabilistic explanation."

Points of Discussion
(in a future and extended version of the paper)

- Parallels in cognitive psychology (e.g. Brunswik 1957)
- Evolutionary perspectives and implications for an evolutionary epistemology
- Relations to Hume's (1740) problem of induction and to Whewell's (c 1840) principle of the "consilience of inductions"