Studies in History and Philosophy of Science 56 (2016) 168–174

Contents lists available at ScienceDirect

Studies in History and Philosophy of Science

journal homepage: www.elsevier.com/locate/shpsa

Empirical agreement in model validation

Julie Jebeile*, Anouk Barberousse

Université Paris-Sorbonne, SND, Paris, France

ARTICLE INFO

Article history:
Available online 26 October 2015

Keywords:
Empirical agreement;
Scientific modeling;
Model validation;
Confirmation holism;
Calibration

ABSTRACT

Empirical agreement is often used as an important criterion when assessing the validity of scientific models. However, it is by no means a sufficient criterion as a model can be so adjusted as to fit available data even though it is based on hypotheses whose plausibility is known to be questionable. Our aim in this paper is to investigate into the uses of empirical agreement within the process of model validation.

Empirical agreement is often used as an important criterion in model validation (Oberkampf & Trucano, 2002; Oberkampf, Trucano, & Hirsch, 2002). However, it is by no means a sufficient criterion as a model can be so adjusted as to fit available data even though it is based on hypotheses whose plausibility is known to be questionable. Our aim in this paper is to investigate into the uses of empirical agreement within the process of model validation as it is performed in scientific practice.

In order to do so, we first present the main reason why empirical agreement is not a sufficient criterion for model validation, namely, Duhem problem of refutation and confirmation holism. What we here call "Duhem problem" is the model-oriented version of the Duhem–Quine thesis (Lenhard & Winsberg, 2010, Winsberg, 2010): When a model's outputs are not the expected ones, the modeler has usually no way to cut the model into pieces that could be confirmed or refuted isolatedly. As a result, she cannot identify which part is responsible for the failure. Conversely, when a model's outputs do agree with available empirical data, it is not easy to tell whether it is only due to adjustments or to the model's core hypotheses. The model faces the court of experimental data as a whole in such a way that it is not easy to determine the precise role of each of its components.

However, models do not all suffer in the same way from Duhem problem. According to their goal and component hypotheses, it is more or less easy to overcome Duhem problem. Accordingly, empirical agreement is endowed with different meanings in different modeling situations. In order to account for these differences, we put forward a typology of models in the following.

At last, we put forward a special type of models that illustrates another difficulty in interpreting empirical agreement. This new difficulty is perhaps even more troublesome for the use of models than is Duhem problem.

1. Duhem problem

Even though empirical agreement does play an important role in the activity of model validation, it cannot be considered a straightforward criterion for model validation. Here, we understand validity as a purpose-relative notion. A valid model is one that performs the task for which it has been designed, whether predictions, experiment planning, prototype construction, etc. Even though validity is so construed as to be purpose-relative, empirical agreement seems to play an important role it assessing it. Why isn't empirical agreement a simple criterion for model validity, though? Because when a model's outputs are consistent with data acquired by observation or measurement, it is usually not possible to assess to which element within the model this match is due. More precisely, it is not possible to tell whether it is due to adjustments in the model or to the fact that the model's hypotheses accurately

---

* Corresponding author.
E-mail addresses: julie.jebeile@gmail.com (J. Jebeile), anouk.barberousse@paris-sorbonne.fr (A. Barberousse).

http://dx.doi.org/10.1016/j.shpsa.2015.09.006
0039-3681/

represent the underlying processes accounting for the investigated phenomena. Adjustments are mainly of two sorts: a model is “adapted” to the phenomenon at hand either by calibrating it or by introducing *ad hoc* terms into it. Calibration is a (usually long) process consisting in tuning some parameters—i.e., numerical constants in the model—in order to progressively guarantee that the model outputs better fit the database. Generally, the outputs to fit are associated with observable variables while the parameters to tune are neither known from observation and measurement nor from theories. This process has been described and analyzed by a number of people, including Winsberg (1999), Hitchcock and Sober (2004), and Epstein and Forber (2013). The second sort of adjustment consists in introducing *ad hoc* terms into the model's equations in order to compensate initially omitted target features in the model. These terms are either correlations between variables of the model, which are not derived from theories, or measured values. (For instance, air friction is often neglected in order to describe a free falling body, but an *ad hoc* term to account for air friction can be added in order to study light objects). This process has been extensively described and analyzed by Cartwright (1983, Essay 6 “For phenomenological laws”).

In order to better characterize Duhem problem, we first give a short analysis of these complex objects called models by describing their usual components. This will allow us to present a more precise diagnosis of Duhem problem.

1.1. Models and their components

A model is a representation of a class of phenomena that may have a variety of goals. According to the desired degree of generality of this representation and the modeler's goal, the importance of the various components may vary quite a lot. For instance, there are models whose inability to provide accurate predictions is not considered unacceptable, when their goal is mostly heuristic, whereas for others, precision is a major virtue. In order to shed some light on this diversity, we put forward a distinctive analysis of models based on the identification of the nature and function of their components.

In order to devise a fully general analysis, we both include models that are written and solved by hand, which we call “analytical models”, and computational models, which are written for, and solved by computers. This amounts to saying that we do not focus on the purely conceptual aspect of models, that is, their instantiating theoretical principles. On the contrary, we want to include a large variety of representations that are used for various tasks, from prediction to experimental design or artifact production.

The first distinction we introduce is between the conceptual components of models and the components that transform them into *usable tools* (of investigation, prediction, design, etc.). This distinction is meant to capture the intuition that contrary to theories, models encompass elements that do not possess strict justification but are required for these models to be *applied* to concrete situations.

The conceptual components of models are themselves of two sorts: first, the description of the target system's properties and second, a set of equations supposed to represent the behavior of the target phenomenon. The description of the system's properties consists in a selection of properties that are supposed to bear on the problem at hand; as models are representations whose scope and validity are determined by their purpose, this description is not supposed to hold absolutely, but only locally, that is, for the concrete situation at hand, and for a specific purpose. The same is true of the model's equations: they are not designed in the first place to hold for ever, but are only meant to help modelers solve the problem they face. Thus the validity of a model is assessed in terms of the accuracy of its predictions *or* the correctness of its core hypotheses *depending on* its specific purpose. For instance, some statistical models used in life insurances are only considered valid only because, in virtue of the adjustments they contain, they give accurate predictions; it is not expected from them to include any substantial hypotheses about subject matter.[1]

Both the description of the system's properties and the equations are written by relying on already established modeling practice in the relevant domain: they do not come out of nowhere. However, they need not be exclusively grounded on available theories. Sometimes, the equations are just meant to catch a basic type of behavior, like the linear dependence of one variable of interest against another. It is also important to emphasize that in many cases, the assessment of the model's quality is not based on the quality of the representation *per se*, but only relative to the variables that have been identified as interesting ones for the purpose at hand. Purpose-relativity is a major component of the way empirical agreement is taken into account when performing the validation task.

As emphasized above, if one is willing to account for models as *usable tools*, it is necessary to include other components than the conceptual ones and to mention simplifying assumptions, idealizations, approximations, to which we come back below, but also algorithms and computational schemes that are essential parts of computational models. Let us take the ballistic equation as an example to illustrate the difference between our two types of components. The ballistic equation is:

$$\mathrm{d}(\mathrm{v}\cos\tau)/\mathrm{d}\tau = \mathrm{c}/\mathrm{gvF}(\mathrm{v}) \tag{1}$$

where v is the projectile's velocity, τ is the angle between its direction and the horizontal, g is the gravitational constant, c and F express air resistance.

This equation can be derived from Newton's second law and is thus well justified. However, it only holds when the following idealizations are accepted: the projectile is a point mass, there is no wind, and the Earth is flat. For sure, in any concrete situation, at least the first two idealizations have to be dismissed and the equations transformed accordingly. But there is worse, as the ballistic equation is only integrable in very few cases. So there is still another reason why it has to be transformed in order to be applied to a concrete situation.

One may have the impression that the components allowing a model to be usable are inessential, because the genuine scientific content is carried by what we have called the conceptual components. However, this impression is erroneous. The characteristic feature of a model, as opposed to a theory, is precisely to include, as unremovable components, those elements that allow scientists to use it for whatever purpose they may have determined.

In order for models to be usable, model equations need to be (1) expressed in mathematical terms and (2) (analytically or numerically) tractable. Both requirements entail including simplifying assumptions. This both holds for models that are written and solved by hand and for computational models. Simplifying assumptions come in two sorts, approximations and idealizations. Approximations are modifications of the equations that are governed by tractability requirements: they are needed to find out the solutions to the equations (Laymon, 1989a, 1989b; Redhead, 1980; Ramsey, 1990, 1992). For instance, Equation (1) can only be integrated when F has special properties; otherwise it has to be replaced by e.g. polynomials. Why by

---

[1] We are grateful to our anonymous reviewer for mentioning this example to us.