

Confirmation and testing of scientific models, revisited

Wendy S. Parker¹

¹*Department of Philosophy, Ohio University, Athens, USA*

Keywords: *model evaluation; testing; confirmation*

In a well-known paper in *Science* magazine, Oreskes et al. (1994) counseled against the often-used terminology of model “verification” and “validation”. The concept of confirmation, however, remains on the table both in their discussion and in many subsequent treatments of model evaluation.

I argue that usually what we can sensibly aim to confirm are not scientific models themselves, but rather their adequacy for particular purposes. I illustrate how testing for adequacy differs from testing for the truth of modeling assumptions. Explicitly tying model evaluation to purposes in this way, I suggest, has the advantage of discouraging the mistaken assumption that successes had by a model in one context constitute good evidence of its general adequacy.

Nevertheless, a concept other than confirmation might be even more useful when it comes to the practice of model evaluation. On a confirmation-driven approach, attention is not necessarily focused on testing models in the most informative ways; we may favor tests of convenience, even if they have little chance of revealing that our model is inadequate for the purpose of interest (if in fact it is inadequate). An alternative is to approach model evaluation as an activity aimed at severely testing, rather than confirming, the hypothesis that the model is adequate for the purposes at hand. A “severe test” of some hypothesis H is a procedure that is likely to indicate that H is false, if and only if H is in fact false (Mayo 1996).

After some remarks on how severe testing for adequacy-for-purpose might proceed, I argue that even if severe testing is often not achievable, the concept of severe testing would be valuable as a regulative ideal. A focus on severe testing might encourage better use of the limited resources available for model evaluation and, in the face of predictive successes, would direct attention to important questions about the testing procedure that are often overlooked in model evaluation at present. I briefly illustrate how this might play out in the context of climate model evaluation.