Physics Today

A mathematical framework for falsifiability

Ilya Nemenman

Citation: Physics Today 68(10), 11 (2015); doi: 10.1063/PT.3.2929
View online: http://dx.doi.org/10.1063/PT.3.2929
View Table of Contents: http://scitation.aip.org/content/aip/magazine/physicstoday/68/10?ver=pdfcov
Published by the AIP Publishing

gestion of fission makes all the more remarkable the assurance she gave to Otto Hahn in late 1938, as Sime mentions. That assurance is discussed in more detail in Sime's biography of Meitner (*Lise Meitner: A Life in Physics*, page 235), but one wonders whether Meitner was recalling Noddack's proposal from long before, probably without even identifying the source of her memory.

As for Sime's last point, I did not intend to suggest that Hahn had explicitly promised to include Meitner's name on the paper with Strassmann, had she come up with a physical explanation. I certainly believed, though, that such a promise was implicit in his request to the exiled Meitner seeking her advice on his puzzling results. However, a closer reading of Hahn's letter of 19 December 1938 to Meitner (see, for example, pages 233–34 of Sime's biography) shows that I was wrong: Hahn expresses the hope that Meitner will have something to publish on her own, so that "it would still in a way be work by the three of us!" Presumably Hahn wrote that for precisely the reasons Sime states in her letter; he acted in the only way that was open to him at the time. Any suggestion of deceit on his part at this stage would be inappropriate.

Reference

1. E. Fermi, *Nature* **133**, 898 (1934).

J. Michael Pearson
(pearson@lps.umontreal.ca)
University of Montreal
Montreal, Quebec, Canada

A mathematical framework for falsifiability

Paul Steinhardt's criticism[1] that multiverse inflationary cosmology, which was flexible enough to explain both negative and positive results of the BICEP2 experiment (see the Commentary by Mario Livio and Marc Kamionkowski, PHYSICS TODAY, December 2014, page 8), is unfalsifiable has resulted this past year in a renewed interest in the old debate: What defines the scientific method?[2] What makes a good physical theory? While the underlying inflationary theory is mathematically sophisticated and modern, the current debate itself has been surprisingly qualitative, similar to what it could have been five decades ago, when Karl Popper brought falsifiability into the spotlight. Such data-less arguments that are often binary to the extreme—for example, whether falsifiability should be retired altogether[3]—seem out of place in the data-driven, nuanced scientific world.

In fact, we scientists already have a mathematical framework to deal with falsifiability quantitatively. It is based on statistical principles that have long been a part of science. In particular, falsifiability is not an independent concept: Its graded, real-valued generalization emerges automatically from the empirical nature of science, much like the way Occam's razor transformed itself from a qualitative philosophical principle into a statistical result.[4,5]

The emergence of falsifiability from statistical inference is easiest seen in the language of Bayesian statistics. Suppose we want to decide which of two theories, T_1 and T_2, explains the world better. Our *a priori* knowledge of that is summarized in Bayesian priors, P_1 and P_2. After experimental data $\mathbf{x}$ are collected, the ratio of posterior probabilities of the theories is given by Bayes's theorem, $P(T_1 | \mathbf{x})/P(T_2 | \mathbf{x}) = P(\mathbf{x} | T_1)P_1/P(\mathbf{x} | T_2)P_2$, where $P(\mathbf{x} | T_1)$ and $P(\mathbf{x} | T_2)$ are the likelihood terms—the probabilities to get the observed data within the theory. The likelihood increases when the theory "fits" the data. However, because probabilities must be normalized, the likelihood scales inversely with the total number of typical data sets that could have been generated within the theory. The tradeoff between the quality of fit and the statistical complexity is known as Bayesian model selection, and it is used routinely in modern statistics. Against statistically complex theories it provides an *automatic* Occam's razor that depends only weakly on specifics of the priors.

At an extreme, any data set is equally compatible with an unfalsifiable theory and hence can come from it with the same probability. Thus the likelihood is the inverse of the total possible number of experimentally distinct data sets. In contrast, a falsifiable theory is incompatible with some data and hence has a higher probability of generating other, compatible data. The difference between the theories grows with the number of conducted experiments. Thus within Bayesian model selection, any falsifiable theory that fits data well wins eventually, unless the unfalsifiable theory had astronomically higher *a priori* odds. For example, as pointed out by biologist J. B. S. Haldane, evolution cannot generate "fossil rabbits in the Precambrian." Thus Bayesian model selection leads to an immediate empirical, quantitative choice of evolutionary theory over creationism as the best explanation of the fossil record, without the need to reject creationism *a priori* as unscientific.

In other words, there is no need to require falsifiability of scientific theories: The requirement emerges automatically from statistical principles, on which empirical science is built. Its statistical version is more nuanced, as has been recognized by philosophers.[5] The practical applications are hard and require computing probabilities of arbitrary experimental outcomes. In fact, it was an error in such a computation that rekindled the current debate. In addition, there is an uncomfortable possibility that statistics can reject a true theory that just happens to be unfalsifiable. Yet, crucially, statistical model selection is quantitative and evidence driven; it potentially moves the inflationary multiverse debate and similar discussions from the realm of philosophy to that of empirical, physical science. Whereas inflation predicts many different worlds, it is incompatible with others—the theory is not completely unfalsifiable. One can hope to end the long-running arguments about its scientific merits by calculating the relevant likelihood terms.

References

1. P. Steinhardt, *Nature* **510**, 9 (2014).
2. G. Ellis, J. Silk, *Nature* **516**, 321 (2014).
3. S. Carroll, "2014: What scientific idea is ready for retirement?" https://edge.org/response-detail/25322 (2014).
4. D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*, Cambridge U. Press (2003).
5. E. Sober, *Ockham's Razors: A User's Manual*, Cambridge U. Press (2015).

Ilya Nemenman
(ilya.nemenman@emory.edu)
Emory University
Atlanta, Georgia

In praise of CETUP*

Recently I thumbed through the April 2015 issue of PHYSICS TODAY and came across the story (page 22) about the South Dakota underground laboratory. I had a chance to stay there for a few weeks last summer and came away with an excellent impression of the lab's potential. That is something that was described quite well in the story.

What I did not see is any mention of the highly successful operation of the Center for Theoretical Underground Physics and Related Areas (CETUP*), which was established only a few years ago and has attracted excellent groups of scientists for summer programs. Especially worthy of mention are the two organizers, Barbara Szczerbinska at Dakota State University and Baha Balantekin with the University of Wisconsin–Madison. Szczerbinska in particular has taken a lot of initiative and done a great deal of work to get CETUP* off the ground. She deserves to be mentioned in an article about the underground lab and its impact on the state as a whole.

Ulrich Mosel
(ulrich.mosel@theo.physik.uni-giessen.de)
University of Giessen
Giessen, Germany

A note on the neutron–proton mass difference

The Search and Discovery story "The neutron and proton weigh in, theoretically," by Sung Chang (PHYSICS TODAY, June 2015, page 17), reports on very important research determining the neutron and proton masses and mass difference. However, the interpretation in the penultimate paragraph, based on hypothetically varying the neutron–proton mass difference—or the electromagnetic coupling strength or other fundamental parameters—is too narrow.

There is strong phenomenological and theoretical motivation for an underlying theory in which the couplings are unified at short distances—a grand unified theory. If they are, then, for example, after Big Bang nucleosynthesis the number of neutrons, and most other relevant quantities, are affected by all the couplings and would change too.[1] Without a calculation of all the combined effects, one cannot draw any reliable conclusions.

Reference

1. G. L. Kane, M. J. Perry, A. N. Zytkow, *New Astron.* **7**, 45 (2002).

Gordon Kane
(gkane@umich.edu)
University of Michigan
Ann Arbor

Material to capture stardust

The cover of the October 2014 issue of PHYSICS TODAY recently caught my eye. I fondly remember participating in the early discussions and conferences to find a physical medium to capture the high-velocity particles that would be encountered during NASA's *Stardust* mission. It was clear from the beginning that a low-density material, some type of foam, was necessary. In 1987 Peter Tsou of NASA's Jet Propulsion Laboratory visited me at Los Alamos National Laboratory to see some of the foams that we were producing for our physics experiments. Most were opaque and polymeric. Included, however, were some silica-based aerogel foams. It was readily apparent that although the aerogel foams did not have the mechanical tenacity and capture capability of the polymeric foams, they had two unmatched properties: The first was very low carbon and hydrogen content as a result of the preparative process. The second was transparency, the property that would lead to aerogel's ultimate selection. The trajectory of the captured particle could easily be determined and the particle could be found at the end of the visible capture track.

It was gratifying to be recognized for my role in the development of the stardust capture media when Tsou wrote about the history of the search and testing of various foam media and the ultimate selection of aerogel to perform the task.[1]

Reference

1. P. Tsou, *J. Non-Cryst. Solids* **186**, 415 (1995).

Joel M. Williams
(science@swcp.com)
Los Alamos, New Mexico