

## Introduction text

When told of the plan to devote an issue of *Synthese* to the principle of maximum entropy (PME), my first reaction was puzzlement as to why philosophers should be so interested in it. Then a plausible reason appeared to be that PME has been perceived (correctly, in my view) as a small but nonnegligible part of a fundamental revolution in thought now taking place.

We have the analysis of Kuhn describing how new foundation concepts make their way into science; but today in order to study this one does not need to go back to the history books to read of epicycles and ellipses. We can observe in our midst the phenomenon of one paradigm in the process of being replaced by another.

In one respect, our present revolution is not like those of cosmology, evolution, or relativity — which, however grand their concepts, were specialized to one particular area of science. What we lack in grandness of concept we make up for in generality; the new revolution concerns the principles of all human inference, and it applies with equal force to all areas of science. The conceptual disorientation and resulting controversy are being exhibited now in the scientific journals of half a dozen different fields.

But we have to admit that the spectacle unfolding today teaches us very little that could not have been learned from those history books; it is astonishingly like what happened in the time of Galileo, down to such small details as to give one a spooky feeling. However, this is not the place to develop that theme.

### How the revolution started

In particular, Cox (1946) proved by theorem what Jeffreys (1939) had demonstrated so abundantly by example; the equations of probability theory are not merely rules for calculating frequencies. They are also rules for conducting inference, uniquely determined by some elementary requirements of consistency. de Finetti, Wald, and Savage were led to the same conclusion from entirely different viewpoints.

As a result, probability theory, released from the frequentist mold in which Venn, von Mises, and Fisher had sought to confine it, was suddenly applicable to a vast collection of new problems of inference, far beyond those allowed by "orthodox" teaching. One might think that an explosive growth in new applications would result. But it did not, until quite recently, because half of probability theory was missing.

Orthodox statistics had developed only means for dealing with sampling distributions and did not even acknowledge the existence of prior probabilities. So it had left half of probability theory how to convert prior information into prior probability assignments — undeveloped. Jeffreys (1948) recognized that this half was a necessary part of any full theory of inference, and made a start on developing these techniques. Indeed, his invariance principle was closely related to PME, starting from nearly the same mathematical expression.

Much earlier, both Boltzmann (1877) and Gibbs had invoked the mathematical equivalent of PME as the criterion determining Boltzmann's "most probable distribution" and Gibbs' "grand canonical ensemble". But this fact had been completely obscured by generations of textbook writers who sought to put frequentist interpretations on those results, not realizing that in so doing they were cutting Statistical Mechanics off from generalizations to nonequilibrium problems.

It was only at this point that I entered the field, with the suggestion that Shannon's Information Theory provided the missing rationale for the variational principle by which Gibbs had constructed his ensembles. In my view, Gibbs was not assuming dynamical properties of an "ergodic" or "stochastic" nature, but only trying to construct the "most honest" representation of our state of information. One can find much support for this view in Gibbs' own words.

Pragmatically, in equilibrium problems this could not lead to any new results because P ME yielded the same algorithm that Gibbs had given and which was already in use as the basis of our calculations. The value of the principle lay rather in the realization of its generality; when seen in this light it was clear that the Gibbs "canonical ensemble" formalism was not dependent for its validity on the laws of mechanics; it would apply as well to any problems of inference, in or out of physics, that fit into the same logical format.

The place that PME occupies in our present revolution is that it is one of the principles that has proved useful — and fairly general in the task of developing that missing half (logically, the first half) of probability theory. Other such principles are group invariance, marginalization, coding theory, and doubtless others not yet thought of; this appears to be a fertile area for research.

Turning to the present discussions of Bayesian methods and P ME, there is an appalling gulf between what scientists are doing with them and what philosophers are doing to them. What is needed most is not still more contention over the exact meaning of things that were written thirty years ago, but a kind of report from the outside world on what has happened since, how these methods have evolved, and what they are accomplishing today. However, since I. was asked to comment on other works published in this issue and elsewhere, only a little of this can be given here.

It is of course distressing that a few philosophers disapprove of the work of so many scientists, engineers, economists, and statisticians on Bayesian/ PME methods, out of misunderstanding. However, I can speak only for myself, and even then, if thirty commentators interpret one's work in thirty different ways it is not feasible to analyze them all and write or read — thirty separate replies. ¶I\*herefore, I can only restate my position and some technical points as clearly as possible, then reply to a few specific comments where it appears that failure to do so would encourage further confusion.

#### Terminology

One critic states that my terminology is nonstandard and can mislead. He fails to note that this applies only to the 1957 papers; and even there my terminology was standard when written. It is, for example, essentially the same as that used by Jimmie Savage (1954). It is not our fault that Latter-Day Commentators, in ill-advised attempts to "classify 1 18

Bayesians' have scrambled the meanings of our words, producing a language appropriately called NEWSPEAK. To translate, we note a few approximate equivalences between standard terminology of the 1950's and NEWSPEAK:

To this we may add the alarming spread in use of the terms "prior distribution" and "posterior distribution" (which had clearly established meanings as referring to the application of Bayes' theorem) to describe instead two different maximum entropy distributions. As a result, some writers are now unable to distinguish between PME and Bayes' theorem and are led thereby into nonsensical calculations and claims.

PME belongs to that neglected first half of probability theory; Bayes' theorem to the second half that starts only after a prior has been assigned by PME or one of the other principles (as we shall see, the second half, given one prior, can then determine other priors consistent with it).

Because of this utter confusion that has been visited upon us, it is today misleading to use the terms "subjective" and "objective" at all unless you supply the dictionary to go with them. My more recent works have used them only in such phrases as "subjective in the sense that

My papers of 1957 used the term "subjective" not only in the superficial sense of "not based on frequencies", but in a deeper sense as is illustrated by the following statement of position: