started this general shift in thinking. These works had laid the basis of the revolution by their demonstrations that probability theory can deal, consistently and usefully, with much more than "frequencies in a random experiment".

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 (1902) 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