Rivista Internazionale di Scienze Sociali, 1, 1997, pp. 3-18.

# A PLURALISTIC VIEW OF SCIENCE AND ITS UNCERTAINTIES*

To provide a framework for analyzing the state of science at the end of the 20th century and just before the beginning of the 21st century, I begin with an emphasis on the natural uncertainties that flow from the epistemology of experiments (Section I). I extend the analysis to errors of measurement, the great classical source of uncertainty in science, in Section II. Because pure mathematics is often taken as the ideal model of certainty, especially with reference to the clarity and definiteness of its results, in Section III, I contrast the intellectual climate of theoretical physics with that of pure mathematics.

In Section IV, I turn to some of the modern invariances that disturb our foundational certainties by disturbing the status of familiar scientific models of the world. In the final Section, Section V, I turn to the more general topic of freedom and uncertainty, but its bearing on the character of scientific inquiry is obvious.

## I - EPISTEMOLOGY OF EXPERIMENTS

What I want to stress is how far the whole activity of experimentation, including its reporting, is from standard theoretical model of science, full of mathematically definite and certain results.

The first point is that experimental reports, much more than theoretical papers, are not generally understandable. They are like detailed sports reporting, intelligible only to the initiated. There is, of course, the insistence of journals that reports be as brief as possible, and certainly there is no requirement they be generally intelligible. However, I think there is a deeper epistemological reason, which also applies to sports reporting. It is not possible to describe in ordinary language, even augmented by some technical terms, the backhand stroke of a tennis player with any accuracy. Similarly, it is not possible to describe in language the many activities of an experimentalist. This applies to his actions setting up and running experiments, but also to his perceptions in digesting the results in their first 'raw' form. The written account can only hint at the main features of what is done or what is observed.

The central epistemological point of these remarks is that this radical incompleteness of descriptions of experiments is not a mark of bad science, but is an essential,

* Paper presented at the Meeting of the Observatorio «Giordano Dell'Amore» on «Science Technology Society at the Threshold of the 21st Century» (Stresa, October 25th-26th, 1996).

unremovable feature of almost all science. The drastic descriptive limitation of what we have to say about experiments is, in my view, a fundamental limitation of our scientific knowledge, possible or actual.

Moreover, this radical incompleteness of the experimental reporting leads to more appeals to authority in experimental work than in theoretical work. It is common to hear, in every part of experimental science: «Well, we know those results are right and can be trusted because we know X and his lab». Of course, there is the answer that experiments can be repeated by others, and this is the great empirical check against being overwhelmed by authority. But it is still an important point, which can be easily amplified by various historical examples, that in many respects theoretical results can be evaluated for error much easier and more directly than experimental work. To bring this up to current science, computational experiments and simulations need to be included as well.

Part of the epistemology of experiments recognized by everyone is the presence of experimental error, but the theory of error has not crept into the philosophical foundations of science, but remains on the ground floor of actual experimental work, with only an occasional philosophical nod to its importance. Yet error is a central concept of a proper general epistemology, and, on the other hand, has a long technical history of theoretical development, to which I turn in the next section.

Finally I emphasize the hierarchy of models used in the analysis of experimental data. Modern statistics has developed within the set-theoretical view of mathematics, as can easily be seen by perusing the pages of «The Annals of Statistics» and other leading journals. The many levels of data reduction usually needed to get to detailed statistical analysis is a epistemological problem, not a statistical problem as such, and needs more philosophical analysis, with close attention to the varying practices in different parts of science. Skepticism about explicit objectivity holding «all the way down» is certainly one of the reasons for the recent increase in acceptance of a Bayesian viewpoint in statistics, although it was already very explicit at many points in Laplace's writing. The arguments about Bayesian ideas, both by philosophers and statisticians, are highly relevant to the epistemology of experiments.

## II - ERRORS OF MEASUREMENT

As scientific methodology and probability theory developed toghether in the burst of intellectual activity that occurred in the eighteenth century, it was recognized not only that errors in measurement occur, but also that a systematic theory of these errors can be given. Fundamental memoirs on the analysis of errors in observation were written by Simpson, Lagrange, Laplace, Gauss, and others. What is important about these memoirs, however, is that there was no examination of the question of the existence or nonexistence of an exact value for the quantity being measured. It was implicit in these eighteenth-century developments, just as it was implicit in Laplace's entire philosophy, that probabilistic considerations, including errors, arise from ignorance of true causes, and that the physical universe is so constituted that in principle we should be able to achieve the exact true values of physical quantities we desire to measure.

This same attitude persisted in the nineteenth century, and these was an atmosphere of confidence in being able to measure true values exactly, so that in important treatises (for example, Laplaces's *Celestial Mechanics* or Maxwell's *Electricity and Magnetism*) little attention was given to the analysis of errors. In particular, there occurred no detailed quantitative or mathematical analyses of errors. It was implicitly assumed that it was simply a matter of tedious and time-consuming effort to refine the values reported to one more significant digit. The assumption that the fixed true value of a physical measurement can be found is conceptually linked to the attitude of strict determinism enunciated by Laplace and also to the view that every event has a strict determinant cause.

But there is no hope of actually carrying out a program of strict certainty in the sense of obtaining the exact and true value of a continuous physical quantity because of the subtlety and complexity of the interacting causes.

From a working standpoint, there is little doubt that this was Laplace's attitude. Yet it would be a mistake to emphasize the primacy of this working attitude of Laplace and other astronomers and physicists, for the theoretical attitude that with sufficient effort a next significant digit can always be obtained is constantly implicit in nineteenth-century science. A curious and conceptually interesting fact is that hardly anyone enunciated the sensible thesis that this was all a mistake — that there were continual random fluctuations, and that the concept of an exact value had no clear meaning.

A notable and philosophically important exception was C.S. Peirce (1891 and 1892)[1]. The almost overwhelming adoption of the possibility of achieving certainty in science seems from the vantage point of the latter part of the twentieth century naive and unsupported by the evidence. The development of statistical mechanics and the kinetic theory of heat provided the right setting for sound views like those of Peirce to become dominant, but nothing of the sort happened. Certainty about the true value of a continuous quantity requires infinite precision of measurement, and the issues about certainty in classical physics all involve such quantities. The hegemony

[1] After citing Epicurus approvingly, Peirce says «For we now see clearly that the peculiar function of the molecular hypothesis in physics is to open an entry for the calculus of probabilities» (1892, p. 322). A few pages later Peirce has these wise words to say about errors of observation or measurement: «Try to verify any law of nature, and you will find that the more precise your observations, the more certain they will be to show irregular departures from the law. We are accustomed to ascribe these, and I do not say wrongly, to errors of observation; yet we cannot usually account for such errors in any antecedently probable way. Trace their causes back far enough, and you will be forced to admit they are always due to arbitrary determination, or chance» (1892, p. 329).

of theory over practice is reflected in the faith that there were no barriers to approaching asymptotically such precision.

The development of quantum mechanics in the first three decades of the twentieth century finally mustered strong enought evidence against this fantasy of certainty to dislodge it. It was reluctantly, but conclusively, recognized that it did not make sense to claim that any continuous physical quantity could be measured with arbitrary precision in conjunction with the simultaneous measurement of other related physical quantities. Thus, in the strict determinism of Laplace, to predict the deterministic course of events, we require the simultaneous exact measurement of the position and momentum of each particle. It is precisely the point of the fundamental fact of quantum mechanics, however, that such a simultaneously precise measurement of position and momentum is not possible, and, most important, it is not possible in principle. The inability to make exact measurement is not due to technological inadequacies of measuring equipment; rather it arises from the fundamental principle of uncertainty first enunciated by Heisenberg about half a century ago.

There is a remark that needs to be made about the relation between inexact measurement and apparently certain qualitative statements. I have no quarrel with the ordinary use of the word 'certain' and its other forms in ordinary talk, as when, for example, I say 'Of course, it is certain that Abraham Lincoln was once President of the United States'. It is just that when it comes to detailed and quantitative knowledge the search for certainty is bound to be frustrated in principle. It is worth emphasizing that I do not intend that there be a direct application of Heisenberg's uncertainty principle to the qualitative statements of certainty so often made in ordinary discourse. Quantum-mechanical principle cannot be used to prove as much as some people would like about ordinary experience and about the incorrect character of ordinary talk. On the other hand, the philosophical search for certainty does necessarily run up against the central features of one of the most important scientific theories of the twentieth century. If, in the grand tradition of Aristotle, Descartes, and Kant, we want to ground our knowledge of the physical world in a detailed theory, then the kind of certainty asked for above all by Descartes is mistaken. On the other hand, if we look at Descartes and others from a psycological standpoint, not from the standpoint of our quantitative knowledge of the physical world, we run up against a similar set of difficulties. The more detailed we get in our phenomenological effort at certainty, the more that certainty recedes in front of us.

## III - THEORETICAL PHYSICS IS NOT PRIMITIVE MATHEMATICS

I continue to endorse the set-theoretical view of theories as an important foundational viewpoint, but I think there is much about theoretical science that is left out of this view. I especially have in mind the detailed comparison of the methods of theoretical physics and the methods of pure mathematics. The latter come close to satisfying the set-theoretical view of theories. The former do not. It is possible to take the view that it is the job of philosophers of physics to fill in the set-theoretical foundations of physics (certainly it is the kind of enterprise I have engaged in myself in the past, and it is not my point here to denigrate it). What I want to emphasize here is that it is also philosophically important to understand the difference in the way mathematical methods are used in theoretical physics and in pure mathematics, or for that matter, standard applied mathematics — from the standpoint of the discussion here I make no real distinction between the practices of pure and applied mathematics. Even the most superficial empirical examination of «Physics Review Letters», the journal that perhaps more than any other publishes important new physics results, both experimental and theoretical, will confirm what I have to say, or at least so I claim. The articles, strictly limited as to lenght, do not read at all like articles in mathematics journals. Mathematics is used in an informal way, but that phrase is too vague. The real characterization, I think, is that theoretical physics is concerned with problem solving, in strong contrast to theorem proving. The essential skill is knowing just what assumptions or empirical facts to draw upon. Formal axioms seldom if ever play a role.

I should say at once mathematical physics, as opposed to theoretical physics, does use axioms. Look at the heroic efforts to provide an axiomatic foundation for quantum field theory. Essentially all the sustained efforts to provide such an analysis were made by mathematicians, not theoretical physicists.

Perhaps my favorite exhibits for my thesis about the difference are the papers of the two great magicians of 20th century physics, Dirac and Feynman. Any fundamental paper by either one of them being mathematically complete, either in terms of assumptions or argument. But these are only wonderful examples. I think the difference holds for almost any of the first theoretical papers in any part of modern physics. (I say «almost any» although I believe «all», because I have not looked at a wide enough range of candidates).

The saliency of the methodological differences between theoretical physics and mathematics is even more evident in physical papers and treatises further removed from what we all tend to regard as fundamental theory. For example, papers and books on quantum optics or solid state physics are full of mathematical equations and mathematical calculations. I say «calculations» deliberately, for it roughly characterizes the restricted kind of mathematical arguments used. One can look from one physical paper or book to the next and no trace of even an elementary *reductio ad absurdum* argument or of a pure existence proof will be found. (Undoubtedly isolated counter examples can be found to what I am saying, but they will be very isolated, *i.e.*, of very low frequency).

So the three main charateristics of the uses of mathematics in theoretical physics are these. First, they are highly constructive, mainly calculational in character — the calculus is not called the calculus by some inadvertent mistake. Second, they are fragmentary, a calculation, then a physical principle or argument, then another calculation, all weaved together into something physically plausible but far from being an exemplar of purely mathematical argument. Third, the broad organization of the argument fits a problem-solving not theorem-proving style. It is not that theoretical physicists forget to prove the theorems, or don't know how to, it is that they are doing something different, something that is in its own way just as wonderful and impressive as sustained proofs in mathematics.

It is a problem for the philosophy of science to understand the detailed nature of the methods of theoretical physics better than we now do. What I have sketched is only a superficial beginning.

## IV - ANCIENT ANTINOMIES AND MODERN INVARIANCES

Because of his view of the nature of the heavens, Aristotle famously held that the world is eternal. Almost a thousand years later the Christian philosopher Philoponus launched a major attack against Aristotle on the eternity of the world. Although Christians needed to believe in the creation, not the eternity, of the world, Aquinas wisely concluded in Question 46, Part I of the *Summa Theologica* that it must be an article of faith, not demonstration, that the world had a beginning. This long and passionate debate, which I am only sketching, formed the subject of Kant's First Antinomy in the *Critique of Pure Reason*. A modern view, so I would argue, is that a decisive demonstration of the finiteness or infinitude of space and time is out of reach. It is unlikely we can ever prove the Big Bang was the absolute beginning; we certainly seem to have no clue as to how such a proof would be constructed. Note also that we even have a revivial by Hoyle and others of the view that the universe may have been through endless cycles of creation, and the Big Bang was only the beginning of the current cycle.

Ancient arguments between Aristotelians and atomists about the continuity or discreteness of matter are well-known, and were the focus of Kant's Second Antinomy. From a mathematical standpoint continuity and discreteness are radically different, and we ordinarily think of a particular feature of a mathematical model of a theory as exemplifying one or the other. On the other hand, it is difficult to conceive of an experiment that would distinguish between a physical quantity being continuous or discrete if a sufficiently small grid of discreteness is permitted. Experimental proof could only go one way: in some cases there is discreteness of a given, established coarseness. So we expect indistinguishability at a very fine level of discreteness between it and continuity. An appropriate invariance should hold, in spite of the central role of continuity in classical mathematical analysis and many other parts of mathematics. The role of continuity, so I would argue, is not as an inaccessible ontological principle, but as a powerful assumption to support the calculations, symbolic and numerical, of mathematical physics that require continuity. They work, because they are such fine approximations. If there is discreteness, it is not too coarse. Here we can take the view that a discrete and a continuous model, one each of a given set of phenomena, must satisfy theories that are mutually contradictory when restricted to the same domains of interpretation.

This same kind of contradiction arises in connection with Kant's Third Antinomy, that between deterministic and indeterministic theories, the latter having a component of absolute spontaneity, in Kant's insightful terminology. Here we have modern support of the same kind mentioned in connection with the Second Antinomy: two kinds of theories, but indistinguishable from an observational standpoint, no matter how many observations are taken, if the accuracy of the quantitative observations is bounded, *i.e.*, not infinitely precise. Perhaps the simplest example is the ergodic one of Sinai billiards. We have a single billiard ball moving without loss of energy and satisfying the law of reflection, but with a convex reflecting obstacle in the middle of the table. The motion of the ball is ergodic. If we finitely partition the space of the table on which the ball can move and correspondingly make time discrete, starting with the simple Newtonian model, then this discrete model cannot be distinguished by observation at the discreteness level of the elements of the partition from a finite-state discrete Markov process. (The result can also be formulated in terms of flows for continuous time).

So what I am arguing for are new invariances in old bottles, to show how enduring these philosophical and conceptual antinomies are, ready for new ideas and new interpretation on issues central to the natural sciences.

With these ideas will come new uncertainties. The great tradition of classical analysis in the mathematical sense has dominated the applications of mathematics to the empirical sciences for more than 200 years. But the tradition that stretches from Euler to Schroedinger and beyond is finally being displaced by new ideas and new methods that are strongly reinforced by the ever increasing use of computers in every aspect of science.

The old verities of continuity and determinism in nature are being replaced not by new models that show they are wrong, but by new models that show they do not exhibit the proper invariance. To put it bluntly, continuity is a fact of computational convenience, and as such, often of great importance; it is not a fact about nature as such.

## V - FREEDOM AND UNCERTAINTY[2]

The close connection between freedom and uncertainty is the main focus of this section. Entropy as the measurement of freedom is also a focus of this section. The deeper reasons, derived from ergodic theory, for using this particular measure of uncertainty are developed later. The central idea is that two elections or markets as

[2] Much of the content of this section is taken from my recent article, Suppes (1996).

processes have the same freedom if their uncertainty structures are isomorphic. The technical details are given in what follows, but what is to be emphasized to begin with is that even the suggestion that uncertainty is central to the fact of freedom is missing in the classical philophical analyses mentioned above, and in the main philosophical successors to Hume and Kant, such as John Stuart Mill in his famous essay *On Liberty*. This omission continues in the standard literature of this century. Throughout the rest of this paper I try to show that this omission is mistaken, and that intuitive features of freedom in many economic, political and social settings implicitly take some form of uncertainty for granted. The social organization of science takes such freedom as given, and its sense is caught by the familiar call in science for freedom of inquiry.

To put the focus on uncertainty, I propose entropy as *the* measurement of freedom. Entropy is already used as a measure of uncertainty in mathematical statistics and statistical mechanics. Other features of freedom may also be subject to measurement, by my claim is that uncertainty, which is particularly susceptible of measurement, is, as a measure of freedom, *primus inter pares*.

Entropy as a proposed measurement of freedom is phenomenological and result, rather than procedurally, oriented. Consider two elections. The first, $E_1$, has three candidates and each receives about 1/3 of the votes. The second, $E_2$, has two candidates and the winner of the two receives about 3/4 of the votes. Almost all of us would agree, I think, that the results as such are evidence of $E_1$ being more free than $E_2$. In saying this we are assuming the usual *ceteris paribus* conditions. Moreover, in matters political, economic, or scientific, there is a strong skeptical tradition that looks to results rather than intentions in judging the character of an institution or procedure.

I propose that we measure the freedom of a set $A$ of alternatives by the entropy $H$ of the actual chosen proportions, or relative frequencies, of the various alternatives, that is,

$$H(A) = -\sum_{i \in A} p_i \log p_i$$

where log is to the base 2, $p_i \geq 0$ and if $p_i = 0$ then $0 \log 0 = 0$.

To give a feeling for the numbers, so to speak, let us consider the entropy of some American presidential elections. According to the measure proposed the elections since 1850 with the maximum freedom, *i.e.*, maximum entropy, of popular vote were those of 1860 and 1912. The tallies were as follows[3].

[3] Data taken from U.S. Bureau of the Census (1975).

| *1860* | | *1912* | |
|---|---|---|---|
| Abraham Lincoln | 1,865,593 | Woodrow Wilson | 6,296,547 |
| J.C. Breckinridge | 848,356 | Theodore Roosevelt | 4,118,571 |
| Stephen A. Douglas | 1,382,713 | William H. Taft | 3,486,720 |
| John Bell | 593,906 | Eugene V. Debs | 900,672 |
| | | Eugene W. Chafin | 206,275 |
| | | Arthur E. Reimer | 28,750 |
| | H = 1.87 | | H = 1.87 |

In contrast, the least free election as measured by popular vote was in 1964, with an entropy of 0,98.

| | *1964* | | |
|---|---|---|---|
| Lyndon B. Johnson | 43,129,566 | Clifton DeBerry | 32,720 |
| Barry M. Goldwater | 27,178,188 | E. Harold Muun | 23,267 |
| Eric Hass | 45,219 | | |

The measure of freedom I am proposing is, as I said at the beginning, mainly phenomenological. There is no suggestion that the measure itself says very much about the causal factors producing the measure at a given time, or a change in the measure from one period to another, whether in an election or in a market. There is, surely, an utter pluralism of causes of changes in entropy. Above all, increases in freedom occur not necessarily because of the intentional actions of individuals focusing on problems of freedom, but often because of what Aristotle termed incidental causation. This means that their intentions were focused on something else, but out of those intentions arose a mixture of results from the actions of many individuals that increased or decreased the freedom of a given institution, or political or social procedure.

It may well be said by some political philosophers, but not by politicians or voters, that we do not really care about the outcomes of elections. What we care about are the political conditions under which they take place. If there is good evidence prior to the elections that there was a serious campaign among alternative candidates and individuals could freely state their political opinions, then what we judge as important are these conditions and not the fact that there was a real landslide of 90% to 10% in the actual voting. In this sense, it would be argued, the entropy measure is inappropriate. There is something in this criticism. It means that the analysis of freedom should be displaced from the results of the election to the procedures or processes leading up to it. We should then attempt to measure the presence of genuine dissent in the political dialogue preceding the election, the opportunities for choosing in terms of external social and political pressures, the resources available to the various candidates, etc. In my own view the outcome of this investigation would be in most cases fairly consistent with the analysis of the election results. Moreover, it is difficult to get quantitative and objective data about much of the political process leading to elections, but assuming the elections are themselves not dishonestly run, excellent quantitative data can be found in the results alone.

When there is freedom in the sense of entropy as measured quantitatively and as proposed here, it would be surprising to have a high measure of freedom for the process and a low one for the result. Notice, of course, that it is a part of they rhetoric of politics that many people would say, even when very few resources were available, that is still the case that individuals under the law were free to speak their minds about the candidates and to campaign as they wished in favor of whomever they wished. This is an important aspect of freedom and one that may not be satisfactorily caught by the measure I am proposing, but it is also one that is a source of skepticism about a political process that permits the kind of freedom just described and yet produces almost no results to back it up.

These remarks apply to scientific inquiry in any areas of active investigation. For any open problems of scientific importance, we expect disagreement and diversity of opinion. As in the case of elections, without such diversity in the process of inquiry, we are skeptical of the scientific results.

## *Stochastic Freedom*

There is another sense of process that is central to the view of freedom being developed. We can observe successive elections and markets for a number of time periods. It is, above all, the entropy rate of these processes over time, rather than data for a single cross-section, that is central, for reasons I hope to make clear.

First, some technical details. A stochastic process $\chi$ is an indexed family $\{\mathbf{X}_n\}$ of random variables. The index, discrete or continuous, is usually interpreted as time, and so it will be here. For simplicity and without any real conceptual loss, I consider only the discrete case with $n = 1, 2, 3, \ldots$, although some remarks will concern the doubly infinite case, $n = \ldots -2, -1, 0, 1, 2, \ldots$ The usual assumption about the collection of joint probability distributions of any finite subsequence of the random variables being consistent is made.

The appropriate concept of entropy for a stochastic process $\chi$ is that of *entropy rate* $H(\chi)$ defined as follows

$$H(\chi) = \lim_{n \to \infty} \frac{1}{n} H(\mathbf{X}_1, \ldots, \mathbf{X}_n),$$

provided the limit exists. (Notice that $H(\mathbf{X}_1, ..., \mathbf{X}_n)$ is just the entropy of the first $n$ random variables. We convert to a rate by dividing by $n$).

A (discrete, finite) Bernoulli process is a stochastic process that is a sequence $\mathbf{X}_1, \mathbf{X}_2, ...$, or possibly a doubly infinite sequence, with the $\mathbf{X}_n$ 's independent and identically distributed random variables with a fixed finite range of values. It is easy to show that such a Bernoulli process $\chi$ has entropy rate

$$H(\chi) = \lim_{n \to \infty} \frac{H(\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n)}{n} = \frac{nH(\mathbf{X}_1)}{n} = H(\mathbf{X}_1) = -\sum p_i \log p_i.$$

We take the measure of freedom to be the entropy rate of the process.

Consider a market over time in which $m$ individuals are sellers and $n$ are buyers. At each period each buyer makes a purchase from exactly one seller. As before, the uniform probability distribution on the set of $m^n$ possible transactions would define a discrete (and finite-valued) Bernoulli process, which would be for $m^n$ possible transactions the stochastic process with maximum entropy rate and thus the one of this size with maximum freedom.

I simplify the analysis at this point by considering only the sellers as the states of the market process. The probabilities of each of the $m$ states, *i.e.*, sellers, represents the probability a random buyer will choose that seller at the given time. In application of these ideas to market data we would often need to estimate $p_{i,n}$ for seller $i$ at the end of time period $n$ by the relative proportion of the market seller $i$ had for that period and make no attempt to identify the behavior of individual buyers. This asymmetry in the treatment of buyers and sellers is common in the analysis of markets and correspondingly, in the case of elections for candidates and voters. However, it is to be emphasized that this limited kind of data analysis is not at all satisfactory for a study of market processes over time, when the entropy rate depends on the transition data for individual buyers, as will become clear in the sequel. I note here that a sample path for a buyer is the sequence of states occupied by the buyer from one time period to another, with the state representing the seller with whom the buyer has a transaction. Although I do not do it here, for actual data analysis it will be desirable to introduce a state corresponding to a buyer not making a transaction in a given time period.

The necessity of considering the time course of a market, and not just cross-section data, in measuring freedom can be well illustrated by a market with just three sellers. We can look at the three-state Markov market with the transition matrix

| | 1 | 2 | 3 |
|---|---|---|---|
| 1 | $1 - 2\varepsilon$ | $\varepsilon$ | $\varepsilon$ |
| 2 | $\varepsilon$ | $1 - 2\varepsilon$ | $\varepsilon$ |
| 3 | $\varepsilon$ | $\varepsilon$ | $1 - 2\varepsilon$ |

As $\varepsilon \to 0$, the entropy approaches zero, but the cross-sectional distribution remains $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$. I think it is intuitively obvious that a market or election with 100% loyalty, *i.e.*, with $\varepsilon = 0$ in the above analysis, is not free. Sellers or candidates need make no effort to compete. This is why merely cross-section data can be misleading.

More generally, for a stationary process the entropy rate as defined above, it can be shown, is equal to the conditional entropy rate, defined as

$$H'(\chi) = \lim_{n \to \infty} H(\mathbf{X}_n \mid \mathbf{X}_{n-1}, \ldots, \mathbf{X}_1),$$

provided the limit exists, which it does for stationary processes. For a (first-order) stationary Markov process, as in our example,

$$\begin{aligned} H'(\chi) &= \lim H(\mathbf{X}_n \mid \mathbf{X}_{n-1}, \ldots, \mathbf{X}_1) \\ &= H(\mathbf{X}_2 \mid \mathbf{X}_1) \\ &= -\sum_x p(x) \sum_y p(y \mid x) \log p(y \mid x) \end{aligned}$$

and so it is easy to show for our Markov market example as defined above that as $\varepsilon \to 0$, $H(\chi) \to 0$. (Hereafter, I drop the distinction between $H$ and $H'$ in view of their equality for stationary processes).

I now turn to the concept that is critical for making entropy rate the essential measure of the freedom of a market or election process — I add the word «process» to emphasize we are considering processes, not one-time cross-sections. The central question is this. How do two markets, or a market and an election, for that matter, compare in their intuitive sense of freedom if they have the same entropy, and contrariwise? As far as I know, this is not a question that has been previously addressed in economics or political science. There have been several prior uses of entropy to measure the one-time cross-section distribution of a market, as part of a more general consideration of indices of concentration (Encaoua - Jacquemin, 1980; Curry - George, 1983; Tirole, 1988, ch. 5; Foley, 1994), but not of a market as a stochastic process. More importantly, entropy, as an invariant feature of what structural properties of stationary stochastic markets, has not been examined. The answer lies ready at hand in the mathematical literature on ergodic theory. In many cases of conceptual interest two stationary stochastic markets or elections will have the same entropy rate if and only if they are isomorphic in the measure-theoretic sense. It is this latter concept that needs to be formally defined.

Let us first begin with a standard probability space $(\Omega, \Im, P)$, where it is understood that $\Im$ is a $\sigma$-algebra of subsets of $\Omega$ and $P$ is a $\sigma$-additive probability measure on $\Im$. We now consider a mapping $T$ from $\Omega$ to $\Omega$. We say that $T$ is *measurable* if and only if whenever $A \in \Im$ then $T^{-1} A = \{ \omega : T\omega \in A \} \in \Im$, and even more important, $T$ is *measure preserving* if and only if $P(T^{-1} A) = P(A)$. $T$ is *invertible* if the following three conditions hold: (*i*) $T$ is $1 - 1$, (*ii*) $T\Omega = \Omega$, and (*iii*) if $A \in \Im$ then $TA = \{T\omega : \omega \in A\} \in \Im$. In the application we are interested in, each $\omega$ in $\Omega$ is a doubly infinite sequence and $T$ is the *right-shift* such that if for all $n$, $\omega_n = \omega'_{n+1}$ then $T(\omega) = \omega'$. Intuitively this property corresponds to stationarity of the process — a time shift does not effect the probability laws of the process, and we can then use $T$ to describe orbits or sample paths in $\Omega$.

We now characterize isomorphism of two probability spaces on each of which there is given a measure-preserving transformation, whose domain and range need only be subsets of measure one, to avoid uninteresting complications with sets of measure zero that are subsets of $\Omega$ or $\Omega'$. Thus we say $(\Omega, \Im, P, T)$ is *isomorphic in the measure-theoretic sense* to $(\Omega', \Im', P', T')$ if and only if there exists a function $\varphi: \Omega_0 \to \Omega'_0$ where $\Omega_0 \in \Im$, $\Omega'_0 \in \Im'$, $P(\Omega_0) = P(\Omega'_0) = 1$ and $\varphi$ satisfies the following conditions:

$$\varphi \text{ is } 1 - 1, \tag{i}$$

$$\text{If } A \subset \Omega_0 \text{ and } A' = \varphi A \text{ then } A \in \Im \text{ if } A' \in \Im', \tag{ii}$$

$$\text{and if } A \in \Im \quad P(A) = P'(A'),$$

$$T\,\Omega_0 \subseteq \Omega_0 \text{ and } T'\,\Omega'_0 \subseteq \Omega'_0, \tag{iii}$$

$$\text{For any } \omega \text{ in } \Omega_0 \quad \varphi(T\omega) = T'\varphi(\omega). \tag{iv}$$

I emphasize that the isomorphism in the measure-theoretic sense of two markets, two elections, or a market and an election seems at the right level of abstraction. The isomorphism expresses that the two structures have the same degree of uncertainty and thus the same structural freedom, even though they differ considerably in other characteristics. The fundamental point is that our conception of freedom needs to be at a rather high level of abstraction in order to be conceptually useful. It would be of little use if we ended up by making the freedom of each market of election *sui generis*, and thus not comparable to any other. What we need is a methodology for comparing degrees of freedom. The isomorphism is a measure-theoretic sense of two stationary stochastic processes provides the important step of giving us a meaningful basis in terms of uncertainty for judging equivalence in freedom. Note why this is so. The $\varphi$ function mapping one process into another is measure-preserving, so there is a structural isomorphism between corresponding events of the two processes such that they have the same probability. It is precisely the fact that the mapping carries events into events of the same probability that supports the claim that isomorphism in the measure-theoretic sense represents equivalence of uncertainty, and thus, of freedom of markets, elections, or areas of active scientific inquiry.

On the other hand, it is equally important to note that isomorphism in the measure-theoretic sense of two stochastic markets only means isomorphism in the structure of uncertainty, as I have called it. Such isomorphism does not imply observational equivalence, nor would we want it to. For example, a Bernoulli market and a Markov market with strong dependence from one period to the next can be isomorphic in the measure-theoretic sense but easily distinguishable by a chi-square test for dependence. What we want to be able to say about these two markets is that they are equivalent in terms of freedom, but clearly different in other respects.

To show how recent fundamental results are about the relation between entropy rate and measure-theoretic isomorphism, I note that it was an open question in the 1950s whether the two finite-state discrete Bernoulli processes $B\left(\frac{1}{2}, \frac{1}{2}\right)$ and $B\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$ are isomorphic. (The notation here should be clear, as explained earlier; $B\left(\frac{1}{2}, \frac{1}{2}\right)$ means that the probability for the Bernoulli process with two outcomes on each trial is that for each trial the probability of one alternative is $\frac{1}{2}$ and of the other $\frac{1}{2}$). The following theorem clarified the situation.

THEOREM 1 (Kolmogorov, 1958, 1959; Sinai, 1959): *If two finite-state, discrete Bernoulli or Markov processes have different entropies, then they are not isomorphic in the measure-theoretic sense.*

Then the question became whether or not entropy is a complete invariant for measure-theoretic isomorphism. The following theorem was proved a few years later by Ornstein.

THEOREM 2 (Ornstein, 1970): *If two finite-state, discrete Bernoulli processes have the same entropy rate then they are isomorphic in the measure-theoretic sense.*

This result was then soon easily extended.

THEOREM 3 (Adler-Shields-Smorodinsky, 1972): *Any two irreducible, stationary, finite-state, discrete Markov processes are isomorphic in the measure-theoretic sense if and only if they have the same periodicity and the same entropy.*

We then obtain:

COROLLARY 1: *An irreducible, stationary, finite-state, discrete Markov process is isomorphic in the measure-theoretic sense to a finite-state, discrete Bernoulli process of the same entropy rate if and only the Markov process is aperiodic.*

Given a stationary stochastic market or election the case is a good one for accepting entropy rate as an appropriate measure of freedom. To take advantage of the intuitions and results of ergodic theory this rather drastic abstraction has been used, a practice not uncommon in economics, but not to be commended. What is critical is approximate stationarity, and fortunately this can be statistically evaluated for the finite sequence of time periods available.

### *Final Remarks*

The importance of competition in scientific inquiry can hardly be underestimated. Its role in inquiry is comparable to the one it plays in economic markets. It is unfortunate that as we look toward the character of science in the next century, conceptual models of scientific inquiry based on sociological ideas have assumed a greater importance that those based on economic ideas in recent theorizing about the nature of science.

I would claim that it is no accident that much of the best science, like much of the best engineering, gets done in direct response to intense competitive pressures that originate from many sources, economic, political, social and psychological. Uncertainty of outcome is a central characteristic of any serious scientific inquiry, but it is intense competition to get there first that is often the source of critical breakthroughs. Competition and uncertainty, not paternalism of ideas and funds, should be the ideal for scientific inquiry in the 21st century.

PATRICK SUPPES
*Department of Economics*
*Stanford University (USA)*

## REFERENCES

R.L. ADLER - P. SHIELDS - M. SMORODINSKY, *Irreducible Markov Shifts*, in «The Annals of Mathematical Statistics», XLIII, 3, 1972, pp. 1027-1029.

B. CURRY - K. GEORGE, *Industrial Concentration: a Survey*, in «Journal of Industrial Economics», XXX, 1983, pp. 203-255.

D ENCAOUA - A. JACQUEMIN, *Degree of Monopoly, Indices of Concentration and Threat of Entry*, in «International Economic Review», XXI, 1980, pp. 87-105

D.K. FOLEY, *A Statistical Equilibrium Theory of Markets*, in «Journal of Economic Theory», LXII, 1994, pp. 321-345.

D. HUME, *A Treatise of Human Nature*, Selby-Bigge Edition, Oxford 1888.

I. KANT, *Critique of Pure Reason* (1781-1787), N. Kemp Smith, Trans (1929), St Martin's Press, New York 1965.

A.N. KOLMOGOROV, *A New Metric Invariant of Transient Dynamical Systems and Automorphisms in Lebesgue Space*, in «Dokl. Akad. Nauk. SSSR», 119, 1958, pp. 861-864. (Russian) MR 21 # 2035a.

A.N. KOLMOGOROV, *Entropy per Unit Times as a Metric Invariant of Automorphism*, in «Dokl. Akad. Nauk. SSSR», 124, 1959, pp. 754-755. (Russian) MR 2 # 2035b.

J.S. MILL, *On Liberty*, Oxford University Press, London 1954.

D.S. ORNSTEIN, *Bernoulli Shifts with the Same Entropy are Isomorphic*, in «Advances in Mathematics», 4, 1970, pp. 337-352.

C.S. PEIRCE, *The Architecture of Theories*, in «Monist», 1, 1891, pp. 162-176.

C.S. PEIRCE, *The Doctrine of Necessity Examined*, in «Monist», 2, 1892, pp. 321-337.

YA.G. SINAI, *On the Notion of Entropy of a Dynamical System*, in «Dokl. Akad. Nauk SSSR», 124, 1959, pp. 768-771.

P. SUPPES, *The Transcendental Character of Determinism*, in «Midwest Studies in Philosophy», XVIII, 1993, pp. 242-257.

P. SUPPES, *Transzendentale Prinzipien: eine Neubetrachtung der Kantschen Antinomien*, in «Metaphysik», XI, 1, 1994, pp. 43-54.

P. SUPPES, *The Nature and Measurement of Freedom*, in «Social Choice and Welfare», XIII, 2, 1996, pp. 183-200.

J. TIROLE, *The Theory of Industrial Organization*, MIT Press, Cambridge (Mass.) 1988.

UNITED STATES BUREAU OF THE CENSUS, *Historical Statistics of the United States, Colonial Times to 1970*, Bicentennial Edition, Part II, Washington (D.C.) 1975.