THE FOUNDATIONS OF PROBABILITY

Arnold M. Faden
Iowa State University, Ames, U.S.A.

1. From Economics of Space and Time to the Foundations of Probability.

My book (Faden, 1977, abbreviated EST) scarcely mentions probability, and yet in retrospect there is a clear path from EST to the present essay. Let me begin, then, by tracing this path.

What is statistical data? This is the question that set me off. It turns out that most data--population, income, production, trade, etc.--may be represented as measures, in the technical sense of measure theory: as distributions of mass over physical space, time, resources and activities (where "mass" is used in a suitably generalized sense; EST, Chapter 2). The point is that problems may now be formulated in terms of measures, and the resources of measure theory brought to bear to solve them. This program was carried out with great success in classical location theory, so that one can say that measure theory is the natural language for this subject, and, in principle, for any subject based on statistical data. (I was inspired by several papers of Martin Beckmann, 1952, 1953, and also by Kantorovitch, 1942).

EST dealt almost exclusively with deterministic situations, and yet the link to probability was there, though unobtrusive. One of the great foundational schools is the frequentist, which identifies probabilities with relative frequencies, perhaps in a limited or hypothetical form. (In fact all schools attach great importance to frequency data). Now relative frequencies are special cases of the physical measures that are the stock-in-trade of EST. These physical measures participate in a network of relations which is ignored in frequency

Operations Research and Economic Theory
Edited by H. Hauptmann, W. Krelle and K. C. Mosler

theories. The insight into the architectonic of measures that I gained from writing EST gave me an unparalleled vantage point into the merits of these approaches. At the same time, insofar as I had thought about foundational problems, I was of Bayesian persuasion. This provided a certain creative tension in my thought, which has ripened with increasing rapidity of late and of which this essay is a progress report.

2. The State of the Foundations. The aim of all science, all inquiry, is knowledge. Yet criticism reveals that little is known with certainty. But uncertainty itself comes in different degrees, from generally accepted facts at one end to wild speculations at the other. Probability theory is the formal apparatus that quantifies degrees of uncertainty. Probability is the central concept of epistomology, the basis of induction and statistical inference, and "the guide of life" (Butler). Or so it should be.

Yet the foundations of probability are in disarray. The various schools--classical, Bayesian, frequentist, subjective, logical--contend with each other about the meaning of probability statements, to what sorts of things probability statements apply, whether there is one or several distinct probability concepts.

The situation is peculiar. Examination of individual writers--say Laplace, Von Mises, Carnap, Reichenbach, de Finetti--reveals glaring weaknesses. At the same time, the major schools of thought have echoes going back three centuries to the very founding of the subject. (See Fine, 1973, for a survey of schools; Hacking, 1975 for a critical history).

What is one to make of this? First, if a point of view keeps being resurrected despite repeated "refutations" it is likely to have grasped a part of the correct interpretation, though perhaps in a one-sided way misleading to those attending to a different side. Second, the persistence of major disagreement indicates that no single interpretation has arisen and been sufficiently developed to incorporate the valid insights of all schools.

3. Conditions for an Adequate Theory. Thus a study of the history of doctrines provides necessary conditions that an adequate theory must meet. I will single out four.

(i) Symmetry must play a basic role, as enshrined in the vague classical "principle of insufficient reason".

(ii) Frequencies must be intimately connected with (if not identical to) probabilities.

(iii) Probabilities must reflect personal systems of beliefs.

(iv) The ideal form of inference is Bayesian (i.e., conditioning on observations).

In addition, there are other conditions of adequacy.

(v) The intuitive notion of causation must be dealt with. There is a basic distinction between one state-of-affairs, A, merely providing information about another, B, and causally influencing B. A comprehensive probability approach might represent these both by a conditional probability statement Pr(B|A) = x. There must, then, be something in the structure of the theory that captures this distinction.

(vi) An adequate theory should be nondogmatic, in the sense that anything possible, or at least anything observable, should be given positive prior weight.

(vii) It should be sufficiently rich to find a place for the great diversity of stochastic models that have found application in one area or another.

Also, an adequate theory must take account of certain general features of "human epistemology:"

(viii) The fact that few of our probability judgments are numerically quantified, but instead involve terms like "likely", "plausible", "doubtful", etc. (At the same time they are not totally vague, and a cottage industry of assessing probability judgment has arisen).

(ix) The fact that much of our thinking seems to work by conjectures and refutations (Popper,1982) rather than by conditioning on evidence.

(x) The existence of apparent systematic illusions in probability judgments (Tversky and Kahneman,1974).

(xi) The fact that we believe a lot not justified by critical judgment (in effect, assign probability one)--cf religion, politics or just ordinary perception.

(xii) The limited information-processing capacity of the human mind (Simon, 1982).

(xiii) The limited intake capacity and sensitivity of our senses.

4. My Approach. What follows is the sketch of a program to found a theory meeting the above criteria, and the carrying out of a portion of it. I stress the word "program" because I think many even of the basic principles remain to be discovered. It is important not to commit oneself prematurely to a formalism. I shall outline the main ideas and indicate what the theory will look like if succesfully completed.

5. Random Variables. The concept of random variable is the key to the whole theory, and the greatest single source of confusion in foundational writing. In probability theory one writes, "Let (x_t), t = 1, 2. ..., be a family of random variables," so that each moment of time indexes a different random variable. In applied statistics, on the other hand, the idea of repetition is part of the concept, so that the entire sequence x_t is thought of as one random variable, taking a series of values in a common "sample space". In the first case, the variation is over the possible uncertain values at one time; in the second case the variation is over time itself. Call the first kind "either-or" variation, since exactly one value will be realized, we don't know which. Call the second kind "both-and" variation since a realization occurs for each time point.

We adopt the first approach exclusively, for two reasons. First, it yields a more flexible language. Anything expressible in "both-and" terms can also be expressed in "either-or" terms, but not conversely. Second, and more fundamental. although repetition is indispensible for building a foundational theory, exactly what is being repeated, and how, is itself not to be decided a priori but is a part of concept formation. A theory should not be built into a definition.

More concretely, think of a random variable as representing a question about the world, and its range of values as the possible answers to that question. Thus "what is the population of the world at time t?". corresponds to a random variable for each specific t and its range of values is the set of all propositions of the form, "the population of the world at time t is n", n = 0, 1, 2, ...

Note several points. First, the range of a random variable is a set of propositions (or, better, the states-of-affairs that they respectively affirm) that are exhaustive and exclusive. Often this set codes naturally into abstract mathematical objects--into the natural numbers

in the case above. Second, random variables almost always fall into natural families. In the case above there is the obvious indexing by time t, but also individual countries could be substituted for the world, and other species for homo sapiens. Third, random variables are often defined from others, or have logical relations to them: y is an abstraction of x if the value of x determines the value of y; e.g., inserting the phrase "population to the nearest million" gives an abstraction of the random variable above. Given a collection of random variables (x_t) we can think of them collectively as a single combined random variable x. (A special case is thinking of a stochastic process as a single random function. Note that x is still an "either-or" variable). Fourth, when formulated in ordinary language, most random variables are not sharply defined: a question does not always fix the set of exclusive exhaustive alternatives that answer it. This vagueness is an aspect of "human epistemology" that must be lived with. Any analogy, any metaphor, establishes a random variable. The human mind is stocked with a wealth of variables, from the soft, dealing with subtle shades of feeling and perception, to the hard, dealing with the adamantine concepts of pure science.

Think of the concept of "parameter" as used in statistics. This has an unknown value in a range of specified possibilities, and is therefore a random variable by our definition. But it seems to stand alone. not being one of a natural family of similar variables. (It is precisely this isolation that persuades frequentists that parameters do not have probability distributions). However, a parameter is rarely if ever completely isolated: there are similar models applying to different times, places or situations. These again have corresponding parameters, and the parameters again form a family, more sparsely indexed than the variables per se.

Consider, for example, models of economic change at different time scales. For short-run inventory models, we may treat capital stock, population, tastes, and political institutions as fixed: they are parameters. For growth models capital and population may vary in a fixed institutional framework, while in the very long run economic systems themselves vary. A similar phenomenon arises for different spatial scale levels. In summary, what looks like an isolated parameter at one level becomes one of a family of variables at a higher level. Variables fall into a hierarchy, a nested system of classes in

which the higher-order ones are constant over a broader realm of space, time or objects than the lower-order ones.

Now consider a person at a given time. He has a limited stock of concepts, of questions that he can understand--and thus, in our terms, a limited stock of random variables. Furthermore, this stock changes over time, becoming enriched as one grows up. An expert in a field acquires a richer stock of relevant concepts than a layman. An inhabitant of a region has a richer stock of local concepts than a stranger. Animals, too, have in effect a stock of concepts, certainly poorer than that of man overall, but richer in certain species-specific ways, like von Uexkull's tick that responded only to the presence of butyric acid.

Let (x_i), $i \in I$, be the stock of random variables pertaining to a person at a given time. In probability theory one defines random variables as (measurable) functions in an underlying space Ω. We have not yet mentioned the latter, and indeed for foundations the variables come first. Formally x_i is identified by the set of possible answers to its corresponding question. Ω may be defined as the cartesian product of these sets, and x_i may be redefined as the function that assigns to $\omega \in \Omega$ its i-th component. We now have the familiar probability set-up. Each $\omega \in \Omega$ may be thought of as a "possible world", giving a complete set of answers to all the questions that one understands.

Now suppose one's stock of concepts gets enriched. The prescription just outlined requires that we now change Ω to Ω': Each point of Ω has been split into a multiplicity corresponding to the possible answers to the new random variables. Splitting of this sort is inevitable in any process of concept formation. Probabilists usually postulate an underlying fixed Ω which can accommodate as many random variables as needed.

(Since there are logical relations among random variables, some "possible worlds" involve incompatible answers, and therefore commit us to Meinongian impossible objects. We have no room to explore the deep implications of this fact, but merely mention that it opens the way to introduce probability into mathematical reasoning itself. Polya's "plausible reasoning" (Polya, 1954) is actually informal Bayesian inference, which means that in the process one assigns positive probabil-

ity to statements that turn out false--i.e., are inconsistent. They will, of course, end up with probability zero).

6. Probability. Let random variable x be in a person's conceptual system. It corresponds to a question with its set of possible answers. His state of belief concerning the true answer is represented (ideally) by a probability distribution over the set. $Pr(x \in A) = c$ states that the "degree of belief" that the true value of x lies in A (a measurable subset of the set of possible answers) is c ($0 < c < 1$). Controversy flares up at once concerning the meaning and justification of such statements.

The view advocated here will be called the perspectival interpretation of probability. Probability is thought of as a relation between a person in a certain cognitive situation, with a certain pattern of life experiences, and that person's stock of random variables. It is time to transcend the sterile clash between "subjectivists" and "objectivists". The elliptical perspective of a penny from a certain point of view is a fact of nature, but one depending on the relative orientation of the eye and the penny.

The speech and writing of others becomes part of one's own life experiences and in this way others' life experience, filtered and distorted to be sure, becomes part of one's own. Thus a certain convergence of world views arises within groups that communicate intensively among themselves (schools of thought).

The hard questions remain: What is the quantitative meaning of probability? Why should it be countably additive, or even additive? Why should it be revised by conditioning on observations? And finally, why should it be used as a guide to action? What follows is a synthesis of Bayesian and frequentist views, with the concept of ergodicity playing a key role.

7. Ergodic Processes. Consider an experiment with two possible outcomes, 0 and 1, repeated indefinitely at times $t = 1, 2, \ldots$. Given the outcomes up to time t, one is to give the distribution of the remaining outcomes. We have a family of random variables indexed by t. each two-valued. The prior probability assignment P is over the space Ω of all 0-1 sequences. What restrictions should be placed on P? As a

start, the principle of nondogmatism requires that $P > 0$ for every realization of $x_1, \ldots, x_t$: what can be observed should be given some weight. Thus we may condition on observations without meeting 0/0 (the so-called Kolmogorov paradox will be discussed later). Let P' be the conditional probability on $x_{t+1}, x_{t+2}, \ldots$. With no further restrictions on P, it is easily seen that <u>any</u> (nondogmatic) P' is compatible with <u>any</u> realization of $x_1, \ldots, x_t$, even with the realization of all 0's or all 1's.

Since this freedom contravenes all experience of how people make inferences, it constitutes a crisis for the extreme subjectivists (above all de Finetti) for whom indeed any distribution P is as good as any other. De Finetti's response, 1937, is well-known: P should be <u>exchangeable</u> (i.e., invariant under finite permutations of random variables) and then the posterior P's are not only sensible, but converge to an iid distribution determined by the limiting relative frequency. This result follows from de Finetti's theorem which states that there exists a distribution Q on [0, 1] such that

$$P(x_1, \ldots, x_t) = \int_0^1 y^s (1-y)^{t-s} Q(dy)$$

for all t, where $s = x_1 + \ldots + x_t$. That is, the exchangeable distributions are precisely the mixtures of iid distributions. With experience, posterior Q sharpens toward $\delta(r)$, r the relative frquency of 1's, and posterior P moves correspondingly toward the iid distribution with parameter r. (Two observers with different Q's will converge to the same posterior provided both Q's assign positive mass to every interval).

Note first that de Finetti makes a major concession in his response. Why <u>should</u> people have exchangeable priors? Indeed he concedes too much, for there are plausible observation sequences that would (eventually) convince anyone that the drawings are <u>not</u> from an exchangeable distribution--e.g., readings of day or night at 12-hour intervals, yielding 010101010101... .

How then to proceed? Consider the frequentist approach. The probability $Pr(x_t = 1)$ is taken to be the limiting relative frequency of 1's in a long sequence of observations, and this is to be the same for all t. How about $Pr(x_t = 1, x_{t+1} = 0)$? Presumably, this is the limiting relative frequency of 10's in a long sequence of observations of successive <u>pairs</u>, and this is to be the same for all t. (These pairs will

overlap). Similarly for triples, etc., up to any finite length. Now we ask, for what class of processes are all these conditions fulfilled? That is, while we do not use the frequency definition of probability, we may still raise the question, when is it true (with probability one) that any realization of the process will, for any tuple $i_0, i_1, \ldots, i_n$ of 0's and 1's, have a limiting relative frequency equal to $Pr(x_t = i_0, \ldots, x_{t+n} = i_n)$ for all t? The answer is the ergodic processes, and we may take these very conditions as the definition of ergodicity. To the extent that frequentist concepts make sense, to that extent do ergodic processes pervade the world.

First some technical notes on these processes. (The concept of ergodicity is not quite standardized). The definition extends immediately from two-point state spaces as above to arbitrary state spaces: for any n, for any measurable sets $B_0, \ldots, B_n$, $Pr(x_t \in B_0, \ldots, x_{t+n} \in B_n)$ is (with probability one) the limiting relative frequency with which (n + 1)-tuples lie in $B_0, \ldots, B_n$ respectively. Next, it extends to continuous time, "relative frequency" being replaced by "fraction of time". Next it extends to space (not necessarily isotropically), referring to relative frequency in a volume of space going to infinity, and thence to space-time processes. Finally, ergodicity embraces some rather surprising processes. e.g., the process that assigns probability 1/2 to the two realizations 010101... and 101010... is ergodic (the set of translates of a periodic function with uniformly distributed phase is ergodic).

Frequentists insist that probabilities can be unknown, whereas Bayesians, de Finetti in particular, regard the concept of unknown probability as confused. These positions are now easily reconciled. We may be convinced that a certain process $x_1, x_2, \ldots$ is ergodic but not know which ergodic process it is. We then have an ergodic process-valued random variable (parameter) which itself has a distribution reflecting our cognitive perspective. Our overall distribution is then a probability mixture of ergodic processes. But it is then a general stationary process (i.e., one whose probabilities are translation-invariant).

To spell out this last point: A set of probability distributions is convex if it is closed under mixing. The set of all stationary processes on $x_1, x_2, \ldots$ is convex. As a convex set, its extreme points are precisely the ergodic processes, and any stationary process can be

expressed (uniquely) as a mixture of ergodic processes--that is, as an integral with respect to a distribution over the space of ergodic processes. (These statements are subject to technical qualifications; see Jacobs,1960, Maitra,1977). This set-up is completely parallel to de Finetti's theorem, which states that the convex set of exchangeable processes is constituted by mixing over its extreme points, which are the iid processes. (Note that exchangeable processes are stationary and iid processes are ergodic).

We start, then, with a perspective represented by a stationary process over x_1, x_2, ... As information accumulates (either by direct observation of some x's, or by observation of other random variables dependent on them), the distribution over the mixing parameter sharpens, and in the limit may approach the true underlying ergodic process. It <u>will</u> approach the process if x_t can be observed directly, and the observed relative frequencies yield better and better estimates of the true process, just as the frequentists maintain.

The idea of an ergodic-process-valued variable is not just an artificial construct, but seems to fit many common concepts. Let x_t be the weather on day t in a certain place. Climate is the distribution of weather (not "average weather"), so that climate is a random variable (or parameter) whose range of values is the set of proceses that weather obeys. Living in a place for a while gives one a good idea of the climate, so that one's beliefs focus on a particular ergodic process, but variations in the weather remain uncertain. Similarly, for a geographic province, physiography is the distribution of relief; for a fabric, pattern is the distribution of color, texture the distribution of weave; for a person, character is the distribution of mood and action; and zeitgeist plays the same role for an historical epoch.

Furthermore, this structure captures at least part of the distinction between information-giving and causation discussed above. Causation refers to probabilistic dependence <u>within</u> an ergodic process. By observation we find out which is the true process, and thus get a better picture of the true causal structure.

<u>8. Toward the Full Prior.</u> It is pleasant to have found the necessary and sufficient conditions for the applicability of frequentist ideas. Nonetheless the task of quantifying cognitive perspectives is just be-

ginning. For one thing, the world does not seem to be drawn from a stationary process. For another, the distribution over the mixing parameter has not been specified: which stationary process is the appropriate one? (By the principle of non-dogmatism, positive mass should be assigned to every nonempty open set in the parameter space in some appropriate topology, but this still leaves too much freedom). Finally there is the problem of the reference class that bedevils all frequency approaches.

To take this last issue first, consider the probability that a given person will die within a year. One looks at relative frequencies: but which ones--all people, or people of the recent past, of the same age, sex, income, or what? The narrower the class, the more relevant, but the shakier the inference. Ultimately everything in the universe is unique (if only in space-time location), which yields a reference class of one, and no data. We want some kind of weighting scheme--the more similar the instance the greater the weight--but what is the appropriate metric?

This is a difficult question with profound ramifications to which I offer not an answer but a method of approach. Consider the simplest form of similarity, which is contiguity in time. Time already comes equipped with a natural metric (the one in which the laws of nature assume their simplest form). Given an ergodic process, we may find the distribution of x_t conditional on other observations; in general, the closer the observation is to t the greater its weight. Contiguity in space, and in space-time, may be treated analogously. Finally, similarity of "quality". One must ask, which qualities, and similarity in what respect? Some quality ranges have a natural ordering (intensity of light and sound, pitch, color, speed, shape, etc.). But basically similarity is established to the extent that qualities are associated in space-time processes: similarity itself derives ultimately from contiguity.

Next we tackle the issue of non-stationarity. There are a number of transformations that take ergodic proceses into other ergodic processes, and that may also take some non-stationary processes into ergodic processes. The basic idea is to "invert" these transformations to generate non-stationary processes from the underlying ergodic ones. Ergodicity is preserved under abstraction: if $x_1, x_2, \ldots$ is ergodic,

so is $f(x_1)$, $f(x_2)$, ...; under clumping: y_1, y_2, ... is ergodic, where $y_t = (x_t, x_{t+1}, \ldots, x_{t+r})$, r fixed. If the x_t's are real or vector-valued, ergodicity is preserved under linear filtering, in particular under moving sums and differences (and under differentiation, if this operation is well-defined). These last seem the most important, and their inverses--cumulative sums and integrals--in general yield non-stationary processes. (The Box-Jenkins, 1970, approach to time-series analysis uses this as its basic method for generating non-stationary processes: the ergodic ARMA processes are cumulatively summed one or more times to yield the ARIMA processes).

The standard paradigm for a process in the natural sciences is initial conditions together with laws of development (usually differential equations, perhaps stochastic) yielding the time-path of the process. The laws ideally are to be autonomous, i.e., independent of particular times and places. The situation here is analogous to that above, with the laws themselves being the state-values of the ergodic process, and the initial conditions specifying the constant of integration. The fact that differentiation tends to "ergodicity" is a clue to why the basic laws of nature are differential equations.

The ideal concept of an ergodic process involves an infinite sequence of variables x_1, x_2, ... In the real world few things go on forever, so we must deal with processes bounded in time (and in space as well): the penny melts, the climate changes, the swatch of fabric is bounded. Note one consequence: if the run is short we may never get a good fix on what the underlying process really was.

A related question is: are there any universal laws? To translate into our language: is there a family of random variables $x_{s,t}$, indexed by all space-time locations, such that $x_{x,t} \in A$ is false for all s, t, where A is a subset of the common range of the x's that is observable? Some systems of induction assign probability zero to universal laws, which appears paradoxical in view of the numerous such laws that science seems to have discovered. Actually there is no paradox. The principle of non-dogmatism requires that the law holding in any bounded space-time region be given positive probability, but not for unbounded regions, since the latter would take forever to verify. (At the same time it is not clear why all universal laws <u>should</u> be given probability zero). As for science, while many laws do hold over vast regions,

there seems little justification for extrapolating to infinity. For example, the "constants" of nature (of Planck, Boltzmann, etc.) may be slowly varying. But what is a scientist to do? Since no one knows as yet how to assess the probabilities of different extrapolations, it is simplest just to assert laws without qualification, and to wait for others to find their limitations. Thus a Popper-type strategy may be justified pragmatically for the time being.

Note that bounded processes. coupled to each other by regime changes, may, form a broader point of view be considered one overall process. Wars and revolutions generally signal regime changes, but from the panoramic perspective of a Sorokin or a Richardson they are routine events in an overall ergodic process. One's cognitive perspective should include the distribution over possible regime changes, their type, timing and location.

Finally, we come to the distribution of the mixing parameter. "Objectivist Bayesians" such as Laplace, Jeffreys and Jaynes--who are close in spirit to the viewpoint of this paper--look for symmetry or invariance principles to fix the prior probability assignment. The idea is to find a natural way of expressing blank ignorance.

The standard objection is that there is no such way; a distribution uniform over x is non-uniform over y = f(x) if f is non-linear, and y is as good a parametric representation as x. Laplace and Jeffreys are open to this objection. In addition Jeffreys makes his invariant prior depend on the likelihood function, a highly objectionable procedure. For suppose we now consider another variable which depends on the parameter; then the likelihood function changes and so does the prior; but a (marginal) distribution should not depend on which other variables are in our stock. The procedure of Jaynes, 1968, escapes these objections by making the distribution depend on the physical content of the parameter. Specifically, the distribution should be invariant under rigid motions in space-time. (Compare the notion that calendar time or geographic location should not appear explicitly in natural laws). Thus a distribution over time should be uniform, and if that makes t^3 nonuniform, so be it! Note that gambling devices such as roulette wheels, dice and cards have this physical symmetry (with the minimal asymmetry needed to distinguish the various outcomes). Also,

with less intuitive force, distributions should perhaps be invariant under scale changes.

The case in hand involves the space of all ergodic processes. We want to find the distribution that yields the "centroid" of the convex set of stationary processes. Whether sense can be made of this concept remains to be seen.

There is, however, one defect in this approach. We are never in a state of complete ignorance concerning the parameter. Similar situations have arisen in the past, and we have some experience about the kinds of ergodic processes (or their non-stationary generated processes) that have been realized. The parameter, in short, is one of a family. If we now apply frequentist methods here, we are committed to an ergodic process over these parameters, which recreates the same situation at the higher level.

Thus we get a hierarchy of processes. What perhaps closes things up is the requirement of scale or level invariance: a priori the world should look the same at any level, so that wholes and parts have self-similar distributions. This idea remains to be carried out. Note that experience will in general break this prior symmetry, so that a posteriori the atomic, human, and cosmic levels can be quite heterogeneous.

9. Human Epistemology. Of all the foibles of "human epistemology" listed above, the most fundamental is the limited information-processing capacity of the human mind. (Simon has pursued this theme for decades; see Simon, 1982). Detailed, accurate and timely information about the world is useful, of course, but the resources we devote to these tasks must compete with other uses. Thus complexity has a cost, and must be traded off against these other desiderata. A program, called the post-Bayesian approach, has been launched to refound all of statistical inference on the inaccuracy-complexity tradeoff (Faden and Rausser, 1976). For example, hypothesis-testing is viewed as a contest between a simpler but less accurate working model (acceptance) and a more accurate but more complex alternative (rejection). This viewpoint leads to a radical alteration of the usual test criteria.

One consequence of these limitations has already been mentioned: the limited stock of random variables that we possess. Adding to this stock has a cost that must be balanced against the benefits of finer perception, understanding and actions that become possible. It pays to invest heavily in detailed concepts regarding our local environment and our occupation. (Eskimo languages have a rich vocabulary involving snow, Arabic similarly for camel-culture, according to Max-Müller).

Another consequence is vagueness, both in meaning and probability assessment. Most of our concepts are identified by words. When we form new concepts it is often economical to name them by old words, changing the meaning of the latter, rather than coin new terms. Thus arise multiple meanings, doubts and misunderstandings in interpersonal communication and even within ourselves with our faulty memories.

As for vagueness in probability assessment, there is a cost to fixing and storing each successive decimal place of probability, just as with any other form of measurement. There are just two sources of quantitative probabilities--space-time symmetries and physical measures (relative frequencies are a special case of the latter), and in any real-world assessment one must weigh the relevance of evidence from a diversity of sources, correct for selection bias (see below) and worry about ambiguous meanings. It is not surprising that few of our judgments are sharp.

There have been attempts to model formally both these sources of vagueness. For vague probabilities one deals with upper and lower bounds to the assessment--more generally with sets of probability distributions (e.g., Smith, 1961). These approaches appear to miss an essential point--by taking thought and attending carefully one can sharpen assessments (Winkler, 1967). Someone who realizes that there are no ideal lines and points in the world might try to redo geometry using thick lines and blobs instead, but the resulting complexity would probably far outweigh the gain in accuracy.

As for ambiguous meanings, a vast literature concerning "fuzzy sets" has come into being (Kickert, 1978). These may be given a probabilistic foundation. Let a concept have an ambiguous meaning in a realm X, so that we are not sure to which subset of X it actually applies. We then have a set-valued random variable, and the probability assigned

to a given class of subsets of X represents our degree of belief that the correct meaning of the concept is in that class. Now for $x \in X$, define $f(x) = Pr\{A | x \in A\}$, the probability that x lies in the scope of the concept. $0 < f(x) < 1$, and it seems to correspond exactly to the notion of the "grade of membership" of x in the concept. I see little justification for operations that cannot be derived from this interpretation. Fuzzy sets should be replaced by random sets.

We now come to actual distortions of the assessment process. The most common is the assignment (in effect) of probability one to propositions not justifying such faith. Thus there are "accepted" theories and "facts" that everybody knows. Most of the time we accept our perceptions as veridical, ruling out of court the possibility of hallucination. In effect we truncate the range of some random variables. Something like this is a necessity of our constitution, to avoid being swamped by a mass of possibilities. A person in a continual state of cartesian doubt would be unable to function. The strategy may be analyzed in a complexity-inaccuracy tradeoff context. First, if a proposition has probability very close to zero, it may be set to zero with little distortion and possibly with much simplification. But even a proposition with low probability may be accepted as a working hypothesis if "close" to being true; for example, we may round off a measurement and act as if that number were completely accurate. Many random variables have a value that would make life simple (maybe $x = 0$), some with more complexity, and perhaps a "miscellaneous" or "none-of-the-above" value that would leave us bewildered. There is a strong temptation to accept the first, or at least to reject the last.

How does one reconcile this universal practice with the perspectival view of this paper? In using working hypotheses we are in effect assessing probabilities conditional on that hypothesis. One should think of all convictions, however firmly held, as working hypotheses. The danger is that, with strong convictions, one (rationally) does not investigate the possibility that they may be false (never read the opposition newspaper, etc.) so that one becomes locked into a dogmatic position.

Psychologists have uncovered forms of probability assessment that partake of the character of illusions (Tversky and Kahneman, 1974). Should one modify the principles of inferences to allow for this?

Well, once an illusion is recognized as such we correct for it. Indeed, illusions are sometimes corrected in the very process of perception, as in the phenomena of size, shape, and color constancy. It would, of course, be interesting to know why these illusions arise in the first place. Presumably an explanation arising from the complexity-inaccuracy tradeoff exists.

Closely related to these illusions is the phenomenon of selection bias. When fishing with a coarse net one should not infer the nonexistence of small fish. In general, every observational stance determines the probability of observing entities of a certain type if they are present. The tendency to see the blatant and ignore the inconspicuous, to judge from appearances, has to be corrected for. In the case of testimony, the source of most of our knowledge, we have to assess the reliability of the source, motives for lying, etc. That we do an imperfect job of correction is attested to by the multiplicity of schools of thought, ideologies and sects.

10. Social Versus Natural Sciences. The principles we have outlined are universally applicable, so there is no fundamental conflict between Geistes-vs. Naturwissenschaften. On the other hand to get a fix on human behavior requires insight into the complex "intentional" structure of beliefs and motives that is not open to inspection and only imperfectly conveyed by speech. It is true that we have the advantage (first noted by Vico) of being more sensitive to the nature of other people than to rocks (say), but this "verstehen" advantage is more than cancelled by the covert character of these variables. Thus in general in the human sciences runs are shorter, general laws more limited, and Markov processes less appropriate (because of the presence of "memory" in a broad sense) than in the natural sciences. Also, processes in the human sciences (and in biology) tend to involve learning, adaptation and a trend toward optimization, in contrast to the physical sciences. But these differences subsume under the same overall framework.

11. The Principle of Nondogmatism. The principle states that anything observable should get positive probability. We have mentioned this repeatedly, but a few points need clarification. What about sampling from a continuous distribution? The usual description is actually an idealization: we don't observe x, but a small interval around x, given by the limits of our measuring instruments or our eyesight. The dart

hits an area of the target, not a point. What of a random variable having uncountable range? Most of the alternatives must get probability zero. But again we can observe one of only a finite number of alternatives in a finite time and each of these should get positive probability. As $t \to \infty$ then of course probabilities close down toward zero and alternative paths are uncountable, so observability requires finite time. Finally, Kolmogorov's paradox: conditioning on an event of probability zero. The argument above indicates that this never happens in finite time. But Kolmogorov's theory of conditional expectations gives the right limiting answer as $t \to \infty$, as may be proved by a martingale argument.

12. Conclusion. The ideal realization of this program would be a single universal distribution to apply to any stock of random variables, conditioned on a person's life experiences to give his or her cognitive perspective. The distribution would have a built-in structure that does justice to frequentist insights, and also the space-time symmetry required to justify classical insights.

The program has a long way to go, if indeed it can be realized. I think most of the difficulties in carrying it out have been mentioned, and none of them seem insuperable. If the program cannot be carried out it would be important to know that too. (Perhaps there is an irreducible surd inherited from our evolutionary history). "A man's reach should exceed his grasp or what's a heaven for?"

References

M. Beckmann, A continuous model of transportation, Econometrica, 20 (October, 1952): 643-60.

M. Beckmann, The partial equilibrium of a continuous space market, Weltwirtsch. Archiv, 71 (1953): 73-87.

G.E.P. Box & G.M. Jenkins, Time Series Analysis: Forecasting & Control (Holden Day, San Francisco, 2nd ed, 1976).

B. de Finetti, Foresight: its logical laws, its subjective sources, 1937, translated in H. G. Kyburg and H. E. Smokler, Studies in Subjective Probability (Wiley, New York, 1964).

A. M. Faden, Economics of Space and Time: The Measure-Theoretic Foundations of Social Science (Iowa State University Press, Ames, Iowa 1977).

A. M. Faden & G. C. Rausser, Econometric Policy Model Construction: the post-Bayesian approach, Annals Econ. Soc. Meas., 5 (1976): 349-62.

T. Fine, Theories of Probability (Academic Press, New York, 1973).

I. Hacking, The Emergence of Probability (Cambridge University Press, New York, 1975).

Konrad Jacobs, Neuere Methoden und Ergebnisse der Ergodentheorie (Springer, Berlin, 1960).

E. T. Jaynes, Prior probabilities, IEEE Trans. on Systems Science and Cybernetics, SSC-4, Sept. 1968, 227-241.

H. Jeffreys, Theory of Probability (Clarendon Press, Oxford, 3rd ed, 1961).

L. Kantorovitch, On the translocation of masses, Comptes Rendus (Doklady) de l'Academie des Sciences de l'URSS 37 no. 7-8 (1942): 199-201.

W. J. M. Kickert, Fuzzy Theories on Decision-Making (Nijhoff, The Hague, 1978).

A. Maitra, Integral representations of invariant measures, Trans. Am. Math. Soc., 229 (1977): 207-225.

G. Polya, Mathematics and Plausible Reasoning (Princeton University Press, Princeton, 1954).

K. Popper, Conjectures and Refutations; the Growth of Scientific Knowledge (Basic Books, New York, 1982).

H. A. Simon, Models of Bounded Rationality (MIT Press, Cambridge, 1982), 2 vols.

C. A. B. Smith, Consistency in statistical inference and decision, J. Royal Stat. Soc., B 23 (1961): 1-25.

A. Tversky & D. Kahneman, Judgment under uncertainty: heuristics and biases, Science, 183 (1974): 1124-1131.

R. L. Winkler, The assessment of prior distributions in Bayesian analysis, J. Am. Stat. Assn., 62 (1967): 776-800.