UC Berkeley
IURD Reprint Series

Title
The Limits of Statistics in Planning Analysis

Permalink
https://escholarship.org/uc/item/2p5013ck

Author
Neuberg, Leland G.

Publication Date
1977-03-01

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California



LELAND G. NEUBERG

The Limits of Statistics in Planning Analysis

Institute of Urban & Regional Development
University of California, Berkeley

Reprint No. 145

Reprinted from: Quality and Quantity 11
(March 1977): 1-26.

The Institute of Urban and Regional Development serves the faculty and students of the Berkeley Campus conducting research into a wide array of policy-relevant questions. Its foci of inquiry are the processes of urban and regional growth and decline, the consequences that alternative styles of governance have for patterns of development, and the scientific bases for development policy and for policy planning.

Toward these ends, current research is directed to the economics of urbanization in the West and in the Third World; techniques of environmental management; effects of using new energy sources; emergent patterns of suburbanization and central area reconstruction; evaluation of alternative modes of transportation; design of effective means for supplying social services; contemporary issues in housing policy; regulation of urban growth and uses of land; determinants of urban livability and the factors shaping urban life; and potential improvements in methods of analysis, evaluation, and planning.

The Institute's members are drawn from departments throughout the Berkeley Campus, and their research findings are published in their respective professional journals and in commercially distributed books. The Institute prints and distributes many of their research papers; a catalog of publications is available on request.

Quality and Quantity, 11 (1977) 1–26

© Elsevier Scientific Publishing Company, Amsterdam – Printed in The Netherlands

THE LIMITS OF STATISTICS IN PLANNING ANALYSIS *

LELAND GERSON NEUBERG

Postdoctoral Fellow, Institute of Urban and Regional Development and Department of Statistics, University of California at Berkeley

...the adherents of rationality – in particular the planners and designers – are now in the difficult position of having to rethink their premises and to understand their limits

from page 366 of Daniel Bell's *The Coming of Post-Industrial Society*

Introduction: Levels of the World and a Possible Role for Statistics on Each Level

The free will/determinism problem has plagued philosophers at least since Descartes [1]. With the development of genetic theory in biology beginning in the ninteenth century a resolution of the free will/determinism problem was implicit. Mendel's law of dominance suggests that the genetic structure of the two parents determines the probability of a particular character trait in the offspring [2]. Thus the brain's structure and how one thinks and behaves is probabilistically-determinate [3]. Genetic theory soon assumed a (relatively) undisputed position in biology [4]. Yet the free will/determinism problem continued to appear in philosophical discourse [5].

At the turn of this century the development of the theory of quantum mechanics in physics implied a second resolution to the free will/determinism problem. Quantum mechanics suggests that states of the brain (or world) are probabilistically-determinate. That is, given an initial state of the brain (or world) at time t_0, one can state precisely (at

* A course in philosphy from John Rawls and one in mathematical statistics from Erich Lehmann have influenced the orientation of my thinking though only I can be held responsible for what is written here.

least in theory) the probability of any state's obtaining at time $t_1 > t_0$. As in the case of genetics in biology, quantum mechanical theory in physics is no longer subject matter for debate [6]. Yet the free will/determinism problem continues to be debated in philosophy [7].

How can a problem receive at least two resolutions and still remain unresolved? The answer to this puzzle seems to lie in replacing the unitary world previously seen with a multi-leveled world [8]. Thus quantum mechanics resolves the free will/determinism problem on the level of physics but leaves it in existence on other levels – e.g. biological and psychological. (That is, quantum mechanics tells us that the "atomic structure of the brain" is probabilistically-determinate but this tells us nothing a priori about the "properties of the brain's cells" or about the psychological-level entity "mind".) Genetics deals with the problem on the biological level and not, for example, on the physics or psychological level. (That is, genetic theory tells us that the "properties of the brain's cells" are probabilistically-determinate but this tells us nothing a priori about the "atomic structure of the brain" or about the psychological-level entity "mind".) With the distinguishing of various levels of the world the sense in which a philosophical problem can simultaneously be resolved and remain unresolved becomes apparent [9].

Now what can one say about the bearing of knowledge and ways of knowing on one level to knowledge and ways of knowing on another level? Britton Harris (1967, p. 328) suggests one possible relationship:

> The rise of statistical mechanics has opened the door for forms of open systems even in physics and chemistry, and it would appear that the revolutionary advances at the cellular level may lay a basis for unraveling system behavior at all levels.

Thus Harris suggests the possibility of *transferability* of knowledge or way of knowing from one level to another. Yet there seems to be little empirical evidence for and few a priori reasons for believing in automatic transferability. Levels are often different in fundamental ways which may preclude transferability. For example, a biological system differs from a social system in at least two ways: (1) objective death of a biological system has no parallel in a social system and (2) the elements of a social system – social institutions – are not a priori limited in number and function as are the elements of a biological system – cells [10]. These differences may preclude transferability of knowledge or ways of knowing between the biological and social levels.

Basic to both quantum mechanics and genetics is the concept of

probability. Quantum mechanics conceives of matter as probability waves (Richtmyer et al., 1955, pp. 189–192) while genetics describes the probability of an offspring with a given trait (see Webster, 1962, p. 918 and Cramer, 1946, p. 139). The concepts of probability and statistics have been successfully [11] employed elsewhere too in analysis of the physics and biological levels. In statistical mechanics, for example, the postulate of equal a priori probability plays a central role (Huang, 1963, pp. 139–142). In biology stochastic models have become fundamental in the analysis of the propagation of epidemics of diseases, mortality, and population growth [12].

In the so-called empirical social sciences statistical method has become the near dominant way of knowing. That is, propositions of fact most often will not receive serious consideration by social scientists unless an argument based on statistical analysis can be marshalled in their favor. Yet the empirical success (see [11]) of statistical models is more equivocal on the social than on the physics or biological levels. For now, two examples of nonsuccess will suffice: (1) econometric time-series models, for example, must usually be redone every few years to continue explaining with reasonable accuracy and (2) tests of hypotheses about social-level events in seemingly similar contexts often diverge [13].

So, although statistical method is very popular in social-level analysis, empirical evidence of (relatively) unequivocal success on this level occurs less frequently than in either physics or biological-level analysis. In this essay I, in part, attempt a theoretical explanation of this empirical phenomenon. That is, possible a priori limitations on the use of statistics in social-level analysis (including planning) are examined. I begin with an explication of a concept which appears necessary for the employment of statistics on any level – the concept of probability.

Section I: The concept of probability

With respect to the *mathematical* (-axiomatic) properties of probability there seems to be general agreement. Suppose $S = \{s_1, s_2, ...\}$ is a sample space. Then a real-valued function p: $P^S \rightarrow R^1$ (where P^S is the set of subsets of S) is a probability function iff: (1) $p(s_i) > 0$ for all s_i in S; (2) $\Sigma_i\ p(s_i) = 1$; and (3) $p(E) = \Sigma_{si \in E}\ p(s_i)$ for all E in P^S [14]. For our purposes, what is important is that the whole of the mathematical theory of probability can begin with this definition (or set of axioms) alone. One need not associate the sample space in the definition with any particular situation in the world nor, more importantly, specify

how the particular probabilities for each of the sample points are chosen.

Suppose, for example, that the sample space of the definition has two elements, $S = \{H,T\}$. Then there are an infinite number of ways of choosing the probabilities. For if $p(H) = p$ where $0 \leqslant p \leqslant 1$ and $p(T) = 1 - p$ then $p(X)$ is a probability function. For example, p might $= \frac{1}{2}$ so $1 - p = \frac{1}{2}$ so $p(H) = \frac{1}{2}$, $p(T) = \frac{1}{2}$.

The foundations' problems in probability theory are *interpretive* ones: (1) how a sample space is to be associated with a situation in the world and (2) how the particular probabilities for the sample are to be chosen. There are two prominent schools of thought on these interpretive questions: the dominant so-called frequentist or objective probabilists and the less prominent so-called personal or subjective probabilists [15]. Cramer (1946, p. 143) has a precise statement of the frequentist interpretation:

> To any event E connected with a random experiment ξ, we should be able to ascribe a number P such that, in a long series of repetitions of ξ ["under similar circumstances"], the frequency of E [= No. of times E occurs ÷ No. of times ξ is repeated] would be approximately equal to P [16].

Thus, in the example introduced earlier, the experiment ξ might be the repeated tossing of a coin with connected events heads (H) and tails (T). Then $p(H) = \frac{1}{2}$ would have the frequentist interpretation: in a long series of tosses the number of times heads turned up ÷ the number of tosses would be approximately equal to $\frac{1}{2}$.

One of three major criticisms of the objectivist viewpoint (at least implicitly) questions the appropriateness of such a concept of probability for social-level analysis. Savage (1972, pp. 61–62) points out:

> In the first place, objectivist views typically attach probability only to very special events. Thus, on no ordinary objectivist view would it be meaningful, let alone true, to say that on the basis of available evidence it is very improbable, though not impossible, that France will become a monarchy within the next decade [17].

Thus the social-level experiment which might result in France's becoming a monarchy in the next decade is essentially unique – not repeatable under similar circumstances – and hence barred from being assigned a probability under Cramer's frequentist interpretation cited earlier [18].

Thus the first limitation of (frequentist) statistics in social-level analysis stems from possible uniqueness of social-level events in contrast to say physics-level or biological-level events. In the latter two cases the a

priori assumption of an experiment repeatable under similar circumstances seems plausible. In fact, in laboratory experiments in particle physics or cellular biology the experimenter often has sufficient control over the situation to insure "similar circumstances" at each repetition. In social-level experiments the experimenter (observer) often has less (no) control over "circumstances" and also social-level "circumstances" are usually less easily defined than physics or biological-level "circumstances". So the assumption of an experiment repeatable under "similar circumstances" is less plausible on the social level.

The critique of uniqueness of social-level events strikes at the soundness of much social-level statistical analysis. The usual situation in a social-statistical study is a population model – inference of a population parameter (point or confidence interval estimation) from a sample statistic. The a priori assumption is that each element of the sample has been drawn at random, under similar circumstances, from the same population. If this assumption is not fulfilled the meaningfulness of the resulting explanation is open to question. One area, for example, where the assumption appears systematically doubtable is in econometric time-series analysis. Here the sample points are, originally, possibly distant from each other in social-level time [19]. It is difficult to conceive a sense in which the elements of such a sample have been "drawn at random, under similar circumstances, from the same population". Hence the uniqueness of social-level events is the basis of an explanation of the explanatory unreliability of econometric time-series models over longer periods of time [20].

The second prominent interpretation of "probability" is the subjective or personal probability interpretation. This interpretation may be specified with two statements: (1) the (subjective) probability of an event's occurring is the denominator ÷ the sum of the numerator and the denominator of the odds a person would give in betting on an event's occurring and; (2) a person would give odds on any event's occurring [21]. Thus, in our coin tossing example the subjective probability interpretation of $p(H) = \frac{1}{2}$ is that a person would give even-money odds in betting that heads turns up in the toss of a coin. A subjective probability interpretation is often (although needn't necessarily be) involved in Bayesian, decision theoretic, and probabilistic game theoretic analyses [22].

Notice that the subjective interpretation shifts the focus of attention from the world to a person speculating about the world. The objective interpretation is a statement describing repeated situations in the world while the subjective is a statement about what a person will do when confronted with a situation in the world. In fact, the proponents of

subjective probability usually don't even claim that subjective probability statements are statements about the world (minus the person):

> Even if valid analyses of the notion (of a "slightly inexact" or "almost true" theory) have been made or are made in the future, for the ordinary theories of science, it is not to be expected that those analyses will be immediately applicable to the theory of personal probability, normatively interpreted; because *that theory is a code of consistency for the person applying it, not a system of predictions about the world around him*. (Savage, 1972, p. 59; italics mine.)

Thus personal probability, for Savage at least, is merely a guide as to how to make *consistent* judgements about the (social-level) world.

The advantage and disadvantages of the subjective interpretation of probability, as compared to the objective, are apparent. (2) of the subjective interpretation overcomes the necessity for repeatability of the objective interpretation so that subjective probabilities attach to *all* social-level events [23]. If Savage's position that personal probability is merely a method of *consistent* judgements is accepted then subjective probability statements tell us virtually nothing about the (objective) social-level world in sharp contrast to frequentist statements. If one in fact wishes to translate subjective statements into objective ones, the personal interpretation gives us no method for doing so – does one use some sort of weighted average, possibly heavy weights to "experts", of subjective probabilities for a statement about the (objective) social-level world for example? Consistency is, after all, a weak criterion – many people can be simultaneously consistent, nowhere in agreement, and everywhere wrong (in an objective sense) for example [24].

Consider again (see [8]) the classical philosophical existence/perception problem. Berkeley's resolution of this problem was (too briefly): essence = perception – i.e. the inference of the existence of "external objects" corresponding to perceptions requires the untestable (ontological) assumption that those objects do indeed exist (see Berkeley, 1957, pp. 3–222). According to Kuhn the dominant philosophical-scientific paradigm since Descartes-Newton builds on that ontological assumption:

> Many readers will surely want to say that what changes with a paradigm is only the scientist's interpretation of observations that themselves are fixed once and for all by the nature of the environment and of the perceptual apparatus. On this view, Priestly and Lavoisier both saw oxygen [= fixed perceptual correspondent of an "external object"], but they interpreted their observations differently; Aristotle and Galileo both saw pendulums [= fixed perceptual correspondents of "external objects"], but they differed in their interpretations of what they both had seen. (Kuhn, 1970, p. 120).

Kuhn finds this traditional scientific-philosophical paradigm wanting but finds a language for talking differently and yet consistently about the situation elusive [25].

The unresolved objective/subjective probability question stems in part from the assumption of fixed perceptual correspondents of external objects. The objectivist viewpoint of probability presupposes there are fixed ("objective") perceptual correspondents of "external objects" [26]. A subjectivist probability statement, on the other hand, is one about a perceiver's estimation and therefore doesn't presuppose the existence of fixed perceptual correspondents of "external objects" [27]. For the objective probabilists a perceived social-level world exists independent of who is looking at it while for the subjective probabilist the perceived social-level world is, at least in part, created by the observer. The emergence of a more satisfactory probability concept awaits the arrival of a more satisfactory resolution of the question: what, if anything, must be fixed, from observer to observer, in the perceived social-level world?

Section II: Causality, statistics, and planning [28]

The concepts of causality and probability seem to have developed historically in close association [29]. Bickel tells us:

> Classical mechanics and other branches of the exact sciences are built around the principle that like causes produce like effects. Given no obstacles, apples fall, pushed objects move, etc. Probability theory provides a model for situations in which like causes can produce one of a number of unlike effects. A coin that is tossed can land head or tails. (Bickel, 1971, p. 1)

Thus, for example, that states of the world are probabilistically-determinate means: if the world is repeatedly set in state S_0 at time t_0 and observed again at time t_1, then the (number of times a given state S_1 is observed at time t_1) ÷ (number of observations at t_1) in a large number of observations seems to approach a fixed number. In one trial, however, cause S_0 can produce one of a number of unlike effects – the possible S_1's [30].

To begin, consider Hume's definition of cause and effect:

> Similar objects are always conjoined with similar. Of this we have experience. Suitably to this experience, therefore, we may define a cause to be an object followed by another, and where all the objects, similar to the first, are followed by objects similar to the second. Or, in other words, where, if the first object had not been, the second never had existed. (Hume, 1955, p. 87) [31]

Hume's definition may be extended into a definition of probabilistic causality (DPC) – A causes (effect) B iff (always and with necessity) the following three conditions hold: (C1) A is associated with B with some probability ($\neq 0$) ("similar objects are always conjoined with similar"); (C2) A precedes B in time ("an object followed by another"); and (C3) the association ($\neq 0$) of A and B is not "spurious" – i.e. A and B are not in fact both effects of some third cause C ("if the first object had not been, the second never had existed") [32].

The DPC can be tested for in what will be called a controlled, random assignment experiment (CRAE). For example, suppose one wishes to test the hypothesis that smoking causes lung cancer in a population. One might draw $m + n$ subjects at random from the population and assign, again at random, m not to smoke and n to smoke. After a suitable length of time has passed the two groups would be examined for their incidences of lung cancer. Using a binomial model [33] one would assume X = lung cancer is distributed binomial (P_1,m) in the nonsmokers and Y = lung cancer is distributed binomial (P_2,n) in the smokers. The joint distribution of X and Y (independent) would be $\binom{m}{x}\binom{n}{y} P_1^x q_1^{m-x} P_2^y q_2^{n-y}$ and one would be testing H: $P_1 \leqslant P_2$ against K: $P_1 > P_2$. Acceptance of H would mean acceptance of the hypothesis that smoking causes lung cancer in the population, while rejection of H would mean rejection of the hypothesis that smoking causes lung cancer in the population [34].

A CRAE has three features which make it an experimental (and in that sense empirical) test for causality [35]. First, the *random* sample of the population allows one to make a statement about the population based on sample measurements. Second, the *random* sample of the population *and* the *random* assignment to treatment and control groups assure that *if* an association ($\neq 0$) exists (C1) then it is not "spurious" (C3). Third, the experimenter-exercised control in inflicting treatment assures (C2). The (technical) weaknesses of a CRAE are two. First, random sampling and/or random assignment may be difficult to achieve. Second, at a given significance level, for a small sample, the test result is not independent of the particular model (binomial in my example) assumed [36].

The principle difficulty with CRAE's is that the required CRAE may not be employable in human population analysis for ethical (or other) reasons. For example, the infliction of smoking treatment, possibly against their wills, on human beings selected at random from the population is prohibited. There are experiments one could perform to test the hypothesis that smoking causes lung cancer in a human population but they might be less decisive than the CRAE. For example, one might

draw at random a sample of m subjects from the population of nonsmokers and n subjects from the population of smokers and use the same binomial model as in the CRAE. Mathematically, the test would be the same as with the CRAE – H: $P_1 \leqslant P_2$ against K: $P_1 > P_2$. Rejection of H here would mean rejection of the hypothesis that smoking is associated with lung cancer in the population (and hence rejection of the hypothesis that smoking *causes* lung cancer in the population since association is a necessary condition for causality (C1)). Acceptance of H here, however, would merely be acceptance of the hypothesis that smoking and lung cancer are *associated* in the population.

In addition to the ethical constraint (which made the second type of experiment, rather than a CRAE, appropriate) one might face an information constraint. The pre-experiment partition of the (human) population into smokers and nonsmokers required for the second experiment might be impractical. If so, then one might instead draw at random a sample of s subjects from the population and see how many fall into each of the four categories: nonsmokers with cancer (BA), smokers with cancer ($\bar{B}A$), nonsmokers without cancer ($B\bar{A}$), and smokers without cancer ($\bar{B}\bar{A}$). This time one would use a multinomial model and test H: $\Delta \leqslant 1$ against K: $\Delta \leqslant 1$ where $\Delta = P_{\bar{A}B}P_{AE}/P_{A\bar{B}}P_{\bar{A}\bar{B}}$. Rejection of H here would mean rejection of the hypothesis that smoking is associated with lung cancer in the population (and hence rejection of the hypothesis that smoking *causes* lung cancer in the population, again because association is a necessary condition for causality (C1)). Acceptance of H here, however, would again be merely acceptance of the hypothesis that smoking is *associated* with lung cancer in the population.

Notice that the latter two types of "experiment" [37] (in contrast to CRAE's) don't test for conditions C2 and C3. Suppose, for example, that one wished to test the hypothesis that lung cancer causes smoking in the population. At least two plausible theoretical explanations of the phenomenon might be offered. On the psychological-social levels there might be a tendency for persons (knowing they have lung cancer) to conclude there is no sense in further abstaining from smoking. On the biological-psychological levels the pain of lung-cancer might be eased by smoking, inducing (knowing or unknowing) lung cancer sufferers to smoke. Given a human-population ethical constraint, this time against inflicting cancer, (as well as the same information constraint as before), then in order to test that lung cancer causes smoking in the population, one would employ an experiment identical to experiment two (experiment three), to test that smoking causes lung cancer in the population! [38]

The planner is a (would-be) manipulator and controller of social-level

events. The CRAE is therefore a very attractive tool to him. However, on the social level, the employment of CRAE's is limited by constraints of social-order in addition to the ethical and informational constraints we have seen on the bio (medical) level. A planner in the U.S., for example, could not simply select at random $m + n$ banks, nationalize m and leave n alone and test later for a difference in customer service [39]. In his discussion of causality on pages 72–89 Hume's examples are never drawn from what we are in this essay calling the social level of the world. His favorite example is the collision of two billiard balls (physics level); another often-mentioned example is the mind's control over the body (psychological level). The DPC has been successfully tested for in experimental physics and (with less although substantial success) via the CRAE in bio (medicine) and psychology. As was pointed out in the introduction, the a priori assumption of the CRAE's transferability to the social level would be an error.

One is then still left with the problem of social-level causality. There seem to be at least three ways to proceed: (1) weaken the DPC, (2) find another experimental way of testing for the DPC, or (3) find a way of testing for the DPC in which some of (C1), (C2), (C3) are tested for by something other than experiment. The elimination of (C1) leads toward a historical conception of causality, alluded to in [32] of this section and to be taken up again in the conclusion. The reinterpretation of (C2) leads in the direction of feedback conceptions of causality (also alluded to in [32]). The elimination of (C3) leads to a position which might be called "strict empiricism" [40]. The search for an experimental way of testing for the DPC on the social level (2) has led many social scientists, in recent years, to an approach called "causal-modeling". The latter approach leads of necessity, in spite of the confusing views of some of its proponents, to a way of testing for the DPC in which some of (C1), (C2), (C3) are tested for by something other than experiment (3).

Simon (1972) is a seminal paper on the causal-modeling approach. The statistical setting of this approach is a linear-regression model beginning with two variables, say X and Y. The procedure tests for a "causal relation between the two variables" (p. 6) – i.e. whether (C1) and (C3) of our DPC are satisfied. One begins by calculating the correlation coefficient r_{XY}. If r_{XY} is not significantly different from 0 one concludes that no causal relation exists (i.e. (C1) doesn't hold) and is done. If r_{XY} is significantly different from 0 one looks at the partial correlation $r_{XY.Z}$ where Z is a conjectured source of "spuriousness". If $r_{XY.Z}$ is also significantly different from 0 one concludes Z is no source of "spuriousness". If $r_{XY.Z}$ is significantly different from 0, on the basis of added assumptions soon to be discussed, one decides

between Z is an intervening variable (and hence not a source of "spuriousness") and Z is a source of "spuriousness".

The first weakness of the procedure is that, in the absence of an as yet unmentioned assumption, it doesn't test for (C3). This assumption is: there is no other variable besides Z that might be a source of "spuriousness" and Z is not causally related to X and/or Y nonlinearly. One might use Simon's procedure on one conjectured source of "spuriousness" at a time in hopes of avoiding the additional assumption. The difficulty is that the set of variables that could be sources of "spuriousness" is conceivably infinite so that the extended procedure is possibly interminable. The random assignment feature of a CRAE eliminates all possible sources of "spuriousness" simultaneously (even a conceivably infinite set of sources) and so is stronger than any extended version of the causal-modeling approach [41].

The second weakness of Simon's approach is the assumptions required to decide between Z is an intervening variable and Z is a source of "spuriousness". The assumptions are of two kinds: (1) it is assumed that certain variables are not directly causally dependent on other variables and (2) it is assumed that the error terms are uncorrelated. On the nature of the assumptions Simon tells us:

> Again it must be emphasized that these assumptions are "a priori" only in the sense that they are not derived from the statistical data from which the correlations among x, y, and z are computed. The assumptions are clearly empirical. (Simon, 1972, p. 10).

Thus the assumptions are "a priori" in the sense that they are not guaranteed by the experiment at hand. The sense in which they are "empirical", however, is less clear.

First let us consider the second type of assumption – that the error terms are uncorrelated. This assumption is required to assure that the number of equations equals the number of unknowns which allows inference-estimation of the (three) unknown population coefficients. The assumption could be tested for in another experiment of the type under discussion but only if one *assumed* that the three population coefficients were 0 – i.e. only if one assumed what the initial experiment was designed to test. As for the first type of assumption – that certain variables are not directly causally dependent on other variables – Simon himself tells us that an experiment of the type under discussion can't establish this (decisively): "we do not known whether the causal arrow should run from x to y or from y to x" (p. 6). Hence, given that there is at least a subrealm of social-level reality where

CRAE's are barred – where only experiments in which at least two variables are not controlled by the experimenter – are possible, *the added assumptions* of Simon's procedure of testing for the DPC *can't be (decisively) established experimentally.*

Given that a (decisive) experimental test of the added assumptions in the causal-modeling approach can't be made, one can't say a priori that those assumptions are empirical. Consider the first of Simon's first type of assumption: "the age of a person does *not* depend upon either his candy consumption or marital status" (p. 9) [42]. If a causal-modeling type of procedure, possibly in the experiment at hand, demonstrates that the correlations of age and candy consumption and of age and marital status are not significantly different from 0 (rather unlikely in the latter case) then the assumption is empirical. If not then the assumption is a priori. Simon's assertion that "the assumptions are clearly empirical" is at best misleading and at worst false [43].

In the subrealm of the social-level immune to controlled experiment only theory can be the basis for the experimentally untestable pieces of the DPC. A theory of A causes B may be defined as follows: (T) is a theory of A causes B iff it is a sequence of propositions, some empirical and some (but at least one) a priori, which explain how and why A leads to B. Thus the role of theory in (at least a part of) the social sciences may be fundamentally different from its role in the strongly experimental sciences. In the latter its role is to be tested experimentally while in the former its role is, in addition, to provide explanation which is experimentally untestable. Of course this perhaps unique role for theory in the social sciences is not without problems.

Consider the biological-level smoking–cancer problem again. We saw earlier that human population experiments can't distinguish between a theory that cancer causes smoking and that smoking causes cancer. In animal experiments, however, one can test the first theory (and maybe even the second, offering smoke-filled maze-rooms). Then adding the results of the animal experiments to the results of the human population sampling experiments, with the added (plausible but experimentally untestable) assumption that human biological systems are similar to animal (probably rat) biological systems with respect to this matter, one has a quasi-experimental test for the theory (or both theories – they could both be sound).

However on the human social level, (plausible) recourse to similarity of animal social systems must be less frequent. The social theorist's explanation of social-level phenomena stems from a (perhaps relatively unmalleable) social-level worldview. Kuhn tells us:

> I was struck by the number and extent of the overt disagreements between

> social scientists about the nature of legitimate scientific problems and methods . . . somehow, the practice of astronomy, physics, chemistry, or biology normally fails to evoke controversies over fundamentals that today often seem endemic among, say, psychologists or sociologists. (Kuhn, 1970, p. viii)

Kuhn offers no explanation for this contrast inasmuch as he focuses primarily on the natural sciences. Our analysis suggests that the contrast is due, at least in part, to the (permanent) absence of an experimental paradigm for testing for causality in social-level phenomena.

III: The Role of Assumption About the Population in Social-Statistical Inference

The usual goal of social-statistical analysis is inference. That is, one seeks to make a statement about a population based on information about a sample. One usually restricts attention to a certain class of procedures determined by the estimation or testing problem under consideration – e.g. invariant, unbiased, or invariant-unbiased procedures [44]. From within the restricted class of procedures one which is optimal in a sense depending on the type of problem is chosen – e.g. a (uniformly) most powerful test. To begin the selection sequence one must make an assumption about the distribution of the property (or properties) under study *in the population.*

The most common assumption is that the property (or properties) under study is (are) distributed according to one of a family of distributions $\{P_\theta\}$ (on R^n) where θ indexes the family and $\Omega =$ *the set of possible* θ*'s, is a "nice" subset of* R^k. Such a model is called a *parametric* one. Typically, the family of distributions is assumed to be one of a class of families called exponential families [45]. Thus, for example, in our lung cancer-smoking examples earlier we assumed cancer was distributed binomially with probability P_i in the population so the (exponential) family of distributions was the binomial one, the parameter space Ω was equal to [0, 1]. Or in the linear model assumed in the discussion of causel-modeling one frequently assumes X is distributed normal (μ_1, σ_1) and Y is distributed normal (μ_2, σ_2) so the (exponential) family of distributions would be the normal one, the parameter space Ω would be equal to $\{(a,b)/a \in R^1, b \in R^1, b > 0\}$.

Now what is the basis of the (parametric) assumption about the property (or properties) in the population? In the lung cancer-smoking examples the basis of the binomial family assumption is our theory of what cancer is – i.e. we assume lung cancer is something which one

either has or doesn't have. The binomial family assumption is no sounder than the theory -- e.g. does a person who has had lung cancer and been "cured", have or not have it? [46] Huber gives a humorous account of the history of the normality assumption. It appears that, at least in the case of measurement error in estimation problems, normality was assumed because it implied that the selection sequence outlined earlier would result in employing the sample arithmetic mean as estimator (see Huber, 1972, pp. 1042–1045). Huber suggests that the assumption could have been given a quasi-empirical basis:

> It seems to me that this kind of discussion borders on dogmatism; a more rational action would have been to look at actual error distributions in large samples, to check whether they were compatible with a normal distribution and, if not, to develop a different theory of estimation. (Huber, 1972, p. 1043)

Yet, as he later states: "By 1960 one had recognized that one never has a very accurate knowledge of the true underlying distribution" (p. 1045). Hence, Simon possibly to the contrary, the basis of the (parametric) assumption about the property (or properties) in the population is a priori rather than empirical.

At least three approaches to inference have been developed which may be seen as involving modification of the usual parmetric assumption about the underlying distribution. That assumption is, that the underlying distribution belongs to a family of distributions indexed by a "nice" subset of R^k but that nothing is known about the value of the index (= parameter) for the population. In Bayesian analysis one strengthens the parametric assumption by adding that the parameter is itself a random variable with a known (prior) distribution. Robust procedures are those with aspects which are insensitive to (at least) small departures from the parametric assumption. Hence they are a way of weakening the assumption. Nonparametric or distribution-free procedures which replace the parametric assumption with a much weaker one are available for some situations.

In Bayesian analysis one assumes (1) an underlying conditional (posterior) parametric distribution $p(X/\theta)$ for the property X under study in the population and (2) a (prior) distribution $\pi(\theta)$ for the parameter. Hence a Bayesian model is specified by the single joint distribution of X and θ : $p(X,\theta) = p(X/\theta)\ \pi(\theta)$. Frequentists insist that $\pi(\theta)$ be determinable through the usual repeated experiment (although it needn't itself be a parametric distribution). Suppose, for example, a wholesaler has for years received weekly large shipments of a rocket component and his customers have always reported defectives so he has frequencies π_i

that there are i defectives. If he small samples n from the most recent lot and assumes that $p(X/i)$ is distributed $b(n,i/N)$, where N is the size of the recent shipment, and that $p(X,i) = p(X/i)\ \pi_i = b(n,i/N)\ \pi_i$ then he would be employing a Bayesian model. For a given confidence level, the length of the interval estimate for i would be $\leqslant$ the length gotten ignoring past shipments and using an ordinary binomial model [47].

One can employ Bayesian analysis in social-level econometric (e.g. linear regression) problems. For example, one might have values of population coefficients routinely calculated in yearly cross-sectional population studies. Then one could use the frequencies of ranges of values as prior distributions in a current-year cross-sectional sampling study assuming variables are conditionally normally distributed. The difficulty with such an approach is that on top of the usual, possibly unreliable, parametric assumption (in this case that variables are (conditionally) normally distributed) one has the, possibly unreliable, assumed prior distributions [48].

One approach to weakening the usual parametric assumption is the employment of a robust procedure. Classically, "robustness of validity" (i.e. that the actual confidence level be close to or on the safe side of the level in the event the parametric assumption holds) was the sought-after feature of a procedure. Since 1960 there has been considerable concern for "robustness of performance" (stability of power or of the length of the confidence interval under (at least small departures) from the parametric assumption) which usually implies "robustness of validity". (see Huber, 1972, pp. 1045–1046)

In large-sample analysis the employment of Studentized statistics often assures "robustness of validity" (and sometimes "robustness of performance"). For example, consider Student's t-statistic for a one sample problem:

$$t = \frac{\sqrt{n}(X_1 + \ldots + X_n)/n}{\sqrt{\sum_{i=1}^{n}(X_i - X)^2/(n-1)}}$$

where it is assumed X_i is distributed normal $(0,\sigma)$. Then for large n, the central limit theorem implies the numerator is distributed $N(0,\sigma)$, the strong law of large numbers implies the denominator $\simeq \sigma$ (with probability 1), and so Slutsky's theorem implies t is distributed $N(0,1)$ *independent of the assumed normality of the* X_i*'s.* Thus, at least for large samples, the significance level of the t-test will be approximately the same with or without the assumption of normality of the underlying distribution.

Many common tests, however, do *not* possess the "robustness of validity" property in the large sample case. For example, the F-distribution employed in testing linear hypotheses depends on the fourth moment of the underlying distribution in the large sample case. Hence if the normality assumption doesn't hold the significance level can depart considerably from what it would be in the event the normality assumption did hold. One can Studentize the F-statistic by dividing it by a fourth-order sample moment and large-sample tests based on the resulting statistic will have the "robustness of validity" property. This procedure, however, would involve the possibly rather tedious calculation of a fourth-order sample moment and is therefore not commonly employed. Robustness properties in *small* sample procedures is one large area of current research in theoretical statistics [49].

For testing problems, the development of distribution-free or nonparametric procedures beginning in the late 1940's has been a second method of weakening the usual parametric assumption. With a nonparametric assumption restricting attention to unbiased procedures leads one to permutation tests while restricting attention to invariant procedures leads one to rank tests. Permutation tests, unfortunately, become rapidly difficult to perform as sample-size increases. Rank tests are discrete and so have a disadvantage of only a few possible significance levels for the outcome.

Consider, for example, the two sample problem where $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ are samples from distributions with unknown but continuous densities $f(x)$ and $f(y - \Delta)$ and one requires an *unbiased* test of the hypothesis H: $\Delta = 0$ against the alternative K: $\Delta > 0$. This is the nonparametric version of the usual parametric problem where $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ are assumed from populations $N(\epsilon,\sigma)$ and $N(\eta,\sigma)$ and one requires an *unbiased* test of the hypothesis H: $\eta - \epsilon = 0$ against the alternative K: $\eta - \epsilon > 0$. In the nonparametric case the most powerful permutation test (which depends on f and Δ and hence is not *uniformly* most powerful) requires the selection of $k = \alpha\ (m + n)!$ points from among $(m + n)!$ possible points – a task which for even moderate size m and n (say m and $n \simeq 10$) is quite tedious (see pp. 183–189 of Lehman (1959) for details).

On the other hand, suppose $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ are samples from distributions with unknown but continuous densities $f(x)$ and $f(y - \Delta)$ and one requires an *invariant* test of the hypothesis H: $\Delta = 0$ against the alternative K: $\Delta > 0$. As before, this is the nonparametric version of the usual parametric problem where $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ are assumed from populations $N(\epsilon,\sigma)$ and $N(\eta,\sigma)$ and one requires an *invariant* test of the hypothesis H: $\eta - \epsilon = 0$ against the alternative K:

$\eta - \epsilon > 0$. In the nonparametric case the Wilcoxian two-sample rank test, which is most powerful against *some* alternative (but not *uniformly* most powerful), requires the finding of the ranks of n (or m) of the sample points in the combined sample of $(m + n)$ points – a task which is not difficult. However, there are only $(m + n)$ possible significance levels for the outcome which is a disadvantage, especially severe with small samples (see pp. 232–240 of Lehman (1959) for details).

So, a possibly unreliable and often difficult to test parametric assumption about the underlying distribution has been the custom in social-statistical inference. The assumption can be strengthened, as for example in Bayesian analysis, but at a cost of further possible unreliability. Robust procedures have been one form of weakening the usual parametric assumption often at a cost of introducing tedious calculations. Small-sample robust procedures continue to be an open research area. Nonparametric or distribution free procedures, primarily in testing problems, have proved a second form of weakening the usual parametric assumption. Again the cost is sometimes tedious calculations and also, in the case of rank tests, discrete significance levels.

Conclusion: History as a Complement to Statistics in Social-level Empirical Research

Three classical philosophical problems [50] and four problems in social-level analysis play a role in imposing limits on the use of statistics in social-level empirical research and in planning. The three philosophical problems are: (1) the free will/determinism problem, (2) the existence/perception problem, and (3) the problem of causality. The four problems in social-level analysis are: (1) possible uniqueness of social-level events, (2) an infinite a priori set of relevant social-level variables, (3) constraints on social-level experimentation, and (4) assumptions about the underlying distribution in social-statistical inference.

The free will/determinism problem suggests distinguishing levels of the world which in turn formulate our problem: might there not exist limits for the statistical way-of-knowing on some levels which don't exist on others? The existence/perception problem and the possible uniqueness of social-level events combine to make the usual (objectivist) definition of probability problematic on the social level. The infinite a priori set of relevent social-level variables and constraints on social-level experimentation join to make an experimentally testable concept of social-level causality impossible. Assumptions about the

underlying distribution in social-statistical inference further restrict the utility of statistics.

The limits for statistics in social-level analysis can be overcome while remaining within the statistical way-of-knowing, but only partially. Subjective or personal probability overcomes the difficulties of objective probability but subjective probability is in turn limited, if arbitrariness is to be avoided, to a code of consistency for the person applying it. Statistically sophisticated causal-modeling approaches can partially (but only incompletely) test for causality experimentally. Assumptions about the underlying distribution can sometimes be weakened but often not without creating nearly insurmountable calculative difficulties.

What appears to be required is a nonstatistical empirical methodology whose strengths are objectivist statistics' weaknesses. The planner-analyst could then employ the two methodologies eclectically, using one method were the other is weak and vice-versa. Historical method seems to be an approach with strengths complementing the strengths of objectivist statistics. Historians have a long record of successfully coping with the objective/subjective, possible uniqueness of social-level events, infinite a priori set of relevant social-level variables, and non-experimental causality problems.

The historian begins with an at least potentially shared set of documents describing events. The extent to which he carefully takes into account available evidence through citations is one measure of his objectivity. His task is to recreate events on the basis of the shared documents. The source of the recreation, however, is subjective: his own worldview. The pattern of events he describes is based on the documents but recreated through his worldview. As historical analysis develops different schools of historians emerge. Each school shares (at least in part) a worldview – the different schools *see* social-level events differently. The reader of history has a chance to recreate objectivity by examining the various schools' views, weighing their arguments carefully in turn and formulating his own synthesized account. Of course the reader's own worldview, while being shaped by what he reads, is in turn shaping what he sees when he reads (and in fact his worldview is selecting what he reads). So historical analysis involves a highly complex integration of subjective and objective elements.

The historian's own judgement, fed by his extensive reading, narrows the a priori infinite set of social-level variables in a given situation to a few important ones or discovers the essential uniqueness in a given situation. Again, sources of objectivity are (1) the original documents and (2) a comparative analysis of historians' accounts. The causal web the

historian weaves presupposes there is no experimental test for social-level causality. The historian doesn't respect the objectivist statistician's repeatability criterion either. The *source* of the causal connection he stitches is again his own worldview although its *basis* is again the shared documents. Objectivity can again be recreated by the reader's sifting, weighing, and comparing the many overlapping historian-woven causal webs, synthesizing his own, and thus destroying objectivity. If planners can learn to integrate historical and statistical method their ability to aid in shaping social-level events would be greatly enhanced [51].

Notes

1 Descartes (1955) discusses free will on pp. 174–179 and again on pp. 233–236. Hume (1955) writes "of liberty and necessity" on pp. 90–112.

2 Mendel's law of dominance is stated on p. 918 of Webster, 1962.

3 A precise meaning for "probabilistically-determinate" is deferred until section two of the paper – after the explication of probability in section one.

4 Of course there continues to be a good deal of controversy about the *scope* of genetic theory – winess the recent extended controversy on race, class, and I.Q. initiated by Herrnstein (1971). Implicit in this controversy is the question: is "I.Q." in fact a "character trait" and, therefore, does genetic theory apply to its determination? Yet genetic theory itself has become a virtually undisputed element of every high school biology curriculum.

5 For example Peirce (1957, pp. 170–188) examines "the doctrine of necessity (determinism)". He traces the dispute as far back as Democritus and Epicursus.

6 An account of the breakdown of classical physics and the emergence of quantum mechanics is given by Richtmyer, Kennard, and Lauritsen (1955, pp. 106–133).

7 For example Sartre (1956) discusses free will/determinism on pp. 433–556.

8 In speaking of "replacing the unitary world previously seen. . ." I am skirting, for now, a second classical philosophical problem: the existence/perception problem. The earliest treatment of this problem with which I am familiar is on pages 3–222 of Berkeley (1957). This problem has also persisted in the philosophical literature to the present. For example, the first two parts of Sartre (1956) are devoted to this question. Chapter 10 of Kuhn (1956) is a treatment of this problem in the context of his discussion of scientific revolutions. I return, briefly, to the existence/perception problem at the end of section one of the paper.

9 A recent work which employs carefully levels-of-the-world distinctions in critiquing notions of artificial reason is Dreyfus (1972). My attention is restricted *primarily* to the physics, biological, and social levels not because I mean to imply there are no other levels but rather because I have some familiarity with the study of these levels. The a priori assumption that the phenomena of one level are determined on another level is the fallacy of reductivism.

10 Klaus Offe in a recent lecture attributed this example to Jurgen Habermas.

11 By "successful" in this essay I shall mean "good at explaining and predicting phenomena" of the particular level in question.

12 See pp. 45–72 of Chiang (1968) on population growth models and also pp. 189–289 on mortality analysis. These processes, with respect to human animals, are not strictly biological but have important social dimensions. As one might expect then, stochastic models of such processes often prove less successful than stochastic models of physics-level phenomena for reasons to be stipulated in sections one and two of the essay.

13 An explanation for these phenomena will be presented in section one of the essay.

14 In my definition I have used, following Feller (1950, p. 22) a discrete sample space in order to subsequently use reasonably elementary examples. The definition generalizes to the continuous case without altering subsequent arguments. My definition amalgamates Cramer's (1946) two axioms on pp. 152–153 leaving out, for now, the concept of a random variable. It is essentially Wonnacott and Wonnacott's (1969, p. 49) axioms.

15 A third viewpoint, that of the *necessary* probabilists, is outlined by Savage (1972, p. 3). Hume (1955) ties one's belief to a conceptualized repeated experiment on p. 71 of the section "Of Probability" and so should probably be classified as an objective probabilist. The necessary probability viewpoint is touched on again in [32].

16 Cramer's text is a classic in mathematical statistics and also has the best discussion (pp. 141–144) of the frequentist interpretation with which I am familiar.

17 Savage's book is the classic exposition of probability theory from a personalist viewpoint.

18 Savage points out "Many who hold objectivistic views admit that such everyday statements (as "France will become a monarchy within the next decade") may have a meaning, but they insist, depending on the extremity of their positions, that that meaning is not relevant to mathematical concepts of probability or even to science generally." (p. 62). So the difficulty for frequentist statistics in social-level analysis stemming from the possible uniqueness of social-level events is the same possible incompatibility of science and planning discussed by Rittel and Webber (1973, pp. 164–165).

19 The basis of the uniqueness of social-level events may be the special quality of social-level (= historical) time. Lukacs (1971) argues on pp. 89–90 that contemplative modes of thought fostered by capitalist social arrangements, in which he includes statistics (p. 98), often treat, improperly in his view, the time variable in the same fashion as they treat space variables (if they incorporate the time variable at all). Savage [on p. 10] points out how the concept "event" (used in both frequentist and personalistic expositions of probability) makes no formal reference to time.

20 Uniqueness also explains why tests of hypotheses about social-level events in seemingly similar contexts often diverge and is also the source of various aggregation problems in macroeconomic analyses.

21 These are dematematized versions of Savage's two postulates (P4 and P5) on p. 31. Wonnacott and Wonnacott (1969) treat subjective probability on p. 50. Notice (2) doesn't mean the person is not indifferent between possible outcomes (such indifference would be represented by the person's giving even odds on all possible events) but rather that the person is not indifferent between giving odds and not giving odds.

22 A minimum-mathematics treatment of game theory and its limitations is found

in Mansfield (1970, pp. 313–318). A mathematically elementary treatment of game theory may be found in Wonnacott and Wonnacott (1969, pp. 340–349). Decision theory (including game theory and Bayesian analysis) is treated on pp. 312–349. Bickel (1971, pp. 298–341) treats decision theory (including Bayes and minimax procedures) in a mathematically less elementary fashion. He also is a source (p. 155) of some remarks on the employment of Bayesian models with subjective prior distributions (section 3.4 of Bickel, pp. 154–166).

23 If (2) is dropped from the subjective probability interpretation then the class of social-level events to which subjective probabilities attach may be properly contained in the class of all social-level events and the advantage over objective probability is lost. Subjective and objective interpretations in this paragraph are to be understood as social-level interpretations only. Hence Savage on p. 62 cites a possible circularity disadvantage of the objective interpretation which is certainly major and if it applies on the social level applies a priori on all levels. There is also a problem with the meaning of "seems to approach a fixed number" (see section two, p. 9) since there are never more than a finite number of trials. This objective interpretation disadvantage is also germane to all levels.

24 Even a relatively weak criterion such as consitency may elude the subjective probabilist. Hence the definition of regret adopted by Savage on p. 163 is shown to be possibly inconsistent with respect to transitivity, for example, by de Neufville and Stafford (1971, pp. 129–131).

25 Kuhn tries to distinguish "look at" and "see", e.g. "Did these men really *see* different things when *looking* at the same sorts of objects? Is there any legitimate sense in which we can say that they pursued their research in different worlds?" (p. 120). I also have trouble talking consistently with respect to this problem – e.g. in the introduction I talk about "replacing the unitary world previously seen . . ." (assuming one can actually perceive a different world) while in assessing the advantage/disadvantages of objective/subjective viewpoints in this section I talk of "an (objective) world" (assuming there are fixed perceptual correspondents of "external objects").

26 Actually, possibly more weakly, one could say the objectivist viewpoint presupposes a shared set of perceptions.

27 Although subjective probabilists often talk as if there were a shared set of perceptions their viewpoint doesn't need such an assumption.

28 Throughout this section and the next "probability" will mean "objectivist probability" unless otherwise specified. The concept of "probabilistic causality" developed in this section is therefore a limited one.

29 Hume (1955), for example, discusses probability on pp. 69–72 and then causality on pp. 72–89.

30 We have seen earlier that the physics and biological levels are probabilistically-determinate. One might assume that consistency implies that the social level also be probabilistically-determinate. However this need be so only if the levels are in some sense not independent worlds, which a priori is not clear. Of course the social level might in fact be probabilistically-determinate. To demonstrate this would require a social-level theory – e.g. that social institutions constrain people's actions making them more or less likely, etc. Probabilistic-determinism of the social level can't be deduced from probabilistic-determinism of the physics or biological level without an added (reductivist) assumption.

31 Hume adds: "The appearance of a cause always conveys the mind, by a custom-

ary transition, to the idea of the effect. Of this we have experience. We may, therefore, suitably to this experience, form another definition of cause and call it an object followed by another, and whose appearance always conveys the thought to that other. But though both these definitions be drawn from circumstances foreign to the cause, we cannot remedy this inconvenience or attain any more perfect definition which may point out that circumstance in the cause which gives it a connection with its effect. We have no idea of this connection, nor even any distinct notion what it is we desire to know when we endeavor at a conception of it. We say, for instance, that the vibration of this string is the cause of this particular sound. But what do we mean by that affirmation? We mean that this vibration is followed by this sound, and that all similar vibrations have been followed by similar sounds; or, that this vibration is followed by this sound, and that, upon the appearance of one, the mind anticipates the senses and forms immediately an idea of the other. We may consider the relation of cause and effect in either of these two lights; but beyond these we have no idea of it."

Hence although Hume notices that "the mind anticipates the senses" in his second definition, touching upon a possible a priori aspect of the cuase-effect concept, he still wishes to claim that his definition is only "suitably to this experience [of "similar objects are always conjoined with similar"] " – i.e. that the proposition 'A causes B' is empirical. Kant (1958) is critical of the latter view of Hume's (p. 27) but suggests that his earlier observation anticipates his own view that the proposition 'A causes B' is "synthetic a priori" (see [46]). Kant also suggests that Hume's view, presumably that "we have no idea of this connection (between the cause and its effect), nor even any distinct notion what it is we desire to know when we endeavor at a conception of it", implies that the *theoretical* natural sciences are not possible – and yet they exist. In fact Hume's formulation of the vibrating string causes sound example (see above) seems to leave no room for the sound-wave theoretical explanation one learns in freshman physics courses.

This discussion should indicate some of the difficulties involved in establishing a philosophically clear concept of cause and effect. Luckily, the task of this section is a lesser one – namely, demonstrating that within a subrealm of social-level reality where there are certain constraints on experimentation, a decisive experimental test for (probabilistic) causality is impossible.

32 There are a number of difficulties with the definition both in its own right and as to whether it reflects Hume's meaning. Points to be made are: (a) Recall Savage points out that the objectivist view prohibits many common language statements – e.g. statements about the probability of France's becoming a monarchy in the next decade. Thus the "with some probability ($\neq 0$)" in (C1) excludes statements like: "the U.S. policy of containment was a partial cause of Soviet aggresiveness in the cold-war period". This difficulty can be overcome with a historical concept of causality to be discussed in the conclusion; (b) If (C2) is interpreted as a particular A precedes a particular B then we can have feedback phenomena as A at t_1 causes B at t_2 causes A at t_3 ($t_1 < t_2 < t_3$); (c) (C3), the "nonspuriousness" condition, appears to make the DPC circular – i.e. a necessary condition that A be a cause of B is that it be a "nonspurious" cause of B. Also it is not clear that (C3) captures the meaning of Hume's "if the first object had not been, the second never had existed". The latter could be read as

making the DPC a definition of A is a necessary cause of B rather than a definition of A is a sufficient cause of B – the latter being *our* meaning in the DPC. The necessary cause concept runs counter to one's common sense notion of cause: smoking may lead to cancer but we don't expect, in the absence of smoking, to find no cancer – i.e. that A causes B doesn't imply that C ($\neq A$) doesn't cause B too. The logical consequences of adopting the necessary cause (rather than our sufficient cause) definition are explored in Burks (1951); (d) Our DPC is, essentially, the definition employed in Simon (1972): "Hence correlation (C1) is proof of causation in the two-variable case if we are willing to make the assumptions of time precedence (C2) and noncorrelation of the error terms (C3)" (p. 10); (e) The added "(always and with necessity)" makes the DPC synthetic a priori in Kant's sense. That is, for example, the statement 'A causes B' is a statement about the future as well as the past and present and hence can't be empirical. The source of the "always and with necessity" is the analyst's mind – his theory. Thus 'A causes B' can be shown to be false empirically but can't be shown to be true empirically. Any so-called "empirical law" is a synthetic a priori proposition. The synthetic a priori conception of causality is the basis of the necessary probabilists' viewpoint alluded to in [15]. (See [34b] and [43] for related material.)

33 That is, we assume that there are only two possibilities: a population member either has (with unknown probability p) or doesn't have (with unknown probability $(1 - p)$) lung cancer. The role of such assumptions is discussed at length in Section three.

34 (a) The mathematical details of the three examples presented here are found on pp. 140–146 of Lehmann (1959). I put a few in the first example to indicate that the mathematical-statistical models in the first and second examples are identical. Thus the source of stronger claims with respect to causality in the first example is not the mathematical-statistical model but the control exercised by the experimenter; (b) The meaning of accepting a hypothesis should be made clear. Our view is that of Popper (1961) (and Kant – see [32e]): one can't prove a hypothesis, only disprove it. Hence "accept H" means "fail to disprove H".

35 There are a number of types of experiment which retain some but not all of these three features of a CRAE. For example, our second two examples retain only the random sample feature. A study of a volunteer sample of convicts who were going to take up smoking comparing lung cancer incidence before and after would retain only the experimenter control feature (which assures C2).

36 This problem will be discussed in Section three.

37 The label "experiment" for these types of procedures is retained because in statistics they are known as "random experiments". In the laboratory science sense one might not want to call them "experiments".

38 In fact, both theories might be sound.

39 Even assuming a generally accepted measure of customer-service could be found.

40 If it weren't for the bothersome "if the first object had not been, the second never had existed" (and "the mind anticipates the senses") one could read Hume as a "strict empiricist". However, even such selfdeclared empiricists as Simon assume there is possible "spuriousness" to assure against.

41 Notice that the problem of a possibly infinite set of sources of "spuriousness"

is precisely property 6 of wicked problems described by Rittel and Webber (1973) on p. 164.

42 As Kant (1958) suggests on p. 25: "There can be no doubt that all our knowledge begins with experience". Thus if this is the basis of Simon's claim that the added assumptions are empirical he is no doubt correct but then the claim is somewhat vacuous since it would be true for any assumptions. This seems, in any case, not to be Simon's sense because he adds, curiously: "This is a genuine empirical assumption, since the variable 'chronological age' really stands, in these equations, as a surrogate for physiological and sociological age." (p. 9) The meaning of this assertion is not completely clear but its implication seems to be that Simon sees that the added assumption needn't be empirical. Simon seems to be (perhaps unknowingly) avoiding Kant's notion of a synthetic a priori proposition.

43 Kant (1958) (on pp. 25–38) draws two important distinctions. The first is between a priori and a posteriori (or empirical) knowledge. Propostions are a priori iff they are necessary (as opposed to contingent) and universal (as opposed to particular). That is a proposition is a priori iff one can't imagine a world in which the proposition (or its negation) is ever false. The second distinction is between analytic and synthetic propositions. An (affirmative) judgment is analytic iff the concept involved in the predicate is (covertly) contained in the concept involve in the subject. The category analytic a posteriori is empty but Kant claims, controversially, that there are synthetic a priori propositions. He classes so-called "empirical laws" in this category, e.g.: "that in all changes of the material world the quantity of matter remains unchanged; and that in all communication of motion, action and reaction must always be equal" (p. 34). One might argue that since the advent of Einsteinian (and the breakdown of Newtonian) physics it has been easy to imagine a world in which these two propositions are (sometimes) false and hence that they are therefore not a priori. Kuhn, however might reply that the concepts labeled "matter" and "motion" mean something different in Newtonian and Einsteinian physics (pp. 101–102 of Kuhn (1970)). Hence one could still claim the propositions are a priori.

44 A more extensive list of the historically desired properties of procedures is: admissibility, efficiency, consistency, invariance, unbiasedness, and combinations thereof. Lindgren (1962) discusses these properties on pp. 202–205.

45 See Bickel (1971), p. 136 for the definition of a parametric model and pp. 146–154 on exponential families.

46 In the case of a binomial model, the statement 'a or not a is the case at time t' is analytic a priori and hence always true. The question for this model, however, is whether it doesn't often tend to oversimplify the situation. For example, it is often employed in quality control tests of light-bulb shipments ignoring the possibility of weak but working filaments which will soon go out. Also, the "or" in the binomial model is exclusive. Thus, for example, the model can't take account of the frequent social-level phenomenon of an attitude conflict within one person. For example, A tends to favor the status quo because of certain benefits he receives and the difficulty of achieving a successful revolution. On the other hand he would prefer to see a revolution take place if the difficulties could be overcome. Is A then for or against the status quo?

47 Wonnacott and Wonnacott (1969) present a comparison of Bayesian and clas-

sical estimation on pp. 324–331 and a critique of Bayesian methods on pp. 331–333.

48 Bickel (1971) on p. 155 suggests that the neo-Bayesians, including Savage, feel one must *always* think of the true value of the parameter as being the realization of a random variable whose distribution reflects the belief of the experimenter. With such a subjective prior distribution though, Bayesian analysis loses its predictive pretentions (see Section two).

49 See Huber (1972), pp. 1045–1048 on the definition of robustness and pp. 1053–1054 on Studentizing as well as the rest of the article for a review of the robustness research.

50 My treatment (or better – nontreatment) of these problems was not meant to be in their philosophical contexts (such a treatment would require three full, thorough, separate essays) but rather in their relationship to statistics and planning.

51 One account of what the historian does is found in Williams (1973). Williams and others who revised the liberal historians' account of the cold-war, for example, aided in shaping the worldwide anti-war movement in the '60's which eventually played a role in forcing a U.S. military withdrawal from Vietnam.

References

Berkeley, G. (1957). In M.W. Calkins, ed., *Berkeley Essay, Principles, Dialogues with Selections from Other Writings.* New York: Charles Scribner and Sons.

Bickel, P.J. (1971). *Mathematical Statistics Part I Preliminary Edition.* San Francisco: Holden-Day.

Burks, A.W. (1951). "The Logic of Causal Propositions," *Mind* LX (239, July): 363–382.

Chiang, C.L. (1968). *Introduction to Stochastic Processes in Biostatistics.* New York: John Wiley and Sons.

Cramer, H. (1946). *Mathematical Methods of Statistics.* Princeton: Princeton University Press.

Descartes, R. (1955). *Philosophical Works of Descartes,* Vol. I, translated by S. Haldane and G.R.T. Ross. New York: Dover Publications.

Dreyfus, H.L. (1972). *What Computers Can't Do A Critique of Artificial Reason,* New York: Harper and Row.

Feller, W. (1950) *An Introduction to Probability Theory and Its Applications,* vol. I. New York: John Wiley and Sons.

Harris, B. (1967) "The limits of Science and Humanism in Planning." *Journal of the American Institute of Planners* xxxiii: 324–335.

Herrnstein, R. (1971) "I.Q." *Atlantic Monthly* 228: 43–64.

Huang, K. (1963) *Statistical Mechanics.* New York: John Wiley and Sons.

Huber, J. (1972) "Robert Statistics: A Review." *Annals of Mathematical Statistics* 43, (No. 4): 1041–1967.

Hume, D. (1955) In C.W. Hendel, ed., *An Inquiry Concerning Human Understanding.* New York: The Liberal Arts Press.

Kant, I. (1958). *Critique of Pure Reason,* abridged and translated by N.K. Smith. New York: The Modern Library.

Kuhn, S. (1970). *The structure of Scientific Revolutions,* Second Edition, Enlarged. Chicago: The University of Chicago Press.

Lehmann, E.L. (1959). *Testing Statistical Hypotheses.* New York: John Wiley and Sons.

Lindgren, B.W. (1962). *Statistical Theory.* New York: The Macmillan Company.

Lukacs, G. (1971). *History and Class Consciousness Studies in Marxist Dialectics,* translated by R. Livingstone. Cambridge: The MIT Press.

Mansfield, E. (1970). *Microeconomic Theory and Applications.* New York: W.W. Norton and Co.

Neufville, R. de, andStafford, J.H. (1971). *Systems Analysis for Engineers and Managers.* New York: McGraw-Hill Book Co.

Peirce, S. (1957). In V. Thomas, ed., *Essays in the Philosophy of Science.* New York: The Liberal Arts Press.

Popper, K. (1961). *The Logic of Scientific Discovery.* New York: Science Editions.

Richtmyer, F.K. Kennard, E.H. and Lauritsen, T. (1955). *Introduction to Modern Physics.* New York: McGraw-Hill Book Co.

Rittel, H.W.J. and Webber, M. (1972). "Dilemmas in a General Theory of Planning," *Policy Sciences* 4: 155–169.

Sartre, J.-P. (1956). *Being and Nothingness An Essay on Phenomenological Ontology,* trans. by H.E. Barnes. New York: Philosophical Library.

Savage, L.J. (1972). *The Foundations of Statistics,* Second Revised Edition. New York: Dover Publications.

Simon, H.A. (1972). "Spurious Correlation: A Causal Interpretation," in H.M. Blalock, Jr., ed., *Causal Models in the Social Sciences.* Chicago: Aldine Publishing Company.

Webster's New World Dictionary of the American Language College Edition. (1962). Cleveland and New York: The World Publishing Company.

Williams, W.A. (1973). "Confessions of an Intransigent Revisionist," *Socialist Revolution,* Number 17; Vol. 3, No. 5; 87–98. (Delivered to the AHA in San Francisco, December 29, 1973).

Wonnacott, T.H., and Wonnacott, J. (1969). *Introductory Statistics.* New York: John Wiley and Sons.