Meehl, P. E. (1970). Nuisance variables and the ex post facto design. In M. Radner & S. Winokur (Eds.), *Minnesota Studies in the philosophy of science: Vol. IV. Analyses of theories and methods of physics and psychology* (pp. 373-402). Minneapolis: University of Minnesota Press.

Nuisance Variables and the Ex Post Facto Design

Paul E. Meehl

In a recent important contribution Kahneman (1965) has pointed out a psychometric difficulty in the use of matched groups, analysis of covariance, and partial correlation as methods of holding constant the influence of a variable which we cannot control experimentally. Anyone acquainted with psychological and sociological literature will surely agree with Kahneman's initial sentence, "Spurious correlations and confounding variables present a characteristic and recurrent problem to the social scientist." The particular aspect of this many-faceted problem with which Kahneman deals is the fact of statistical "undercorrection" which arises from imperfect reliability in measuring the variable to be controlled. The literature abounds in examples of failure to recognize this difficulty, and hardly any faculty member goes through an academic year without sitting on several doctoral orals in which the candidate—not to mention his adviser—is blissfully unaware of the magnitude of the error that may be thereby introduced, sometimes of vitiating proportions. The present paper is in no sense to be viewed as a criticism of Kahneman's contribution. However, I am afraid I shall make matters worse by pointing out the copresence of a source of error which is at times equally serious as the one to which Kahneman addresses himself but which usually works in the opposite direction. Furthermore, I have no constructive suggestion to offer, and I am unaware that anybody has presented one. It is my opinion that the high prior probability of a joint (and typically countervailing) influence of the source of error pointed out by Kahneman and the source of error I shall emphasize brings about the circumstance that many traditionally acceptable designs in psychology and sociology are methodologically unsound. To put it most extremely, the so-called ex post facto "experiment" (Chapin, 1955; Greenwood, 1945) is fundamentally defective for many, perhaps most, of the theoretically significant purposes to which it has been put. It is perhaps no exaggeration to say that the net influence of Kahneman's criticisms and my own, if valid, is to make a scientifically sound ex post facto design well-nigh impossible with presently available methods.

Frequently research in biological and social science presents the problem of "spurious association" (a concept which in itself deserves a more thorough philosophical analysis than it has, to my knowledge, been given by either statisticians or social scientists). Typically these are research problems in which the organisms under study are in some way "self-selected" (Greenwood, 1945, pp. 126-129) with respect to an experience, setting, or property which is one of the variables of research interest, or a variable known to be correlated with the latter. That is to say, we have to deal with situations unlike the laboratory experiment in which a randomizing procedure is *externally applied* to a sample of organisms in such fashion that the sources of uncontrolled variance can be said in advance to distribute themselves randomly over experimental treatments. We have to

AUTHOR NOTE: This work was supported in part by the National Institute of Mental Health (Research Grant M-4465) and in part, through my summer appointment as professor in the Minnesota Center for Philosophy of Science, by the Carnegie Corporation of New York. A shorter version appeared as Technical Report PR-69-4 from the Research Laboratories of the Department of Psychiatry, University of Minnesota, 1969.

deal with the case in which we, as investigators, do not select what part of the city a child lives in or what college he goes to or what religion his parents profess, but instead must take the "experiment of nature" (as Adolf Meyer would have called it) as it comes. Such investigations, lying somewhere between anecdotal, clinical, or "naturalistic" impressions and laboratory experiments, attempt to combine the necessity of taking the organisms as they come with such scientific procedures as accurate observation, quantitative assessment of variables, and mathematical analysis of the data. (I do not wish to convey the impression that the *only* reason we proceed thus is the fact that we are physically or ethically unable to manipulate and randomize all of the variables, since a case can also be made, and would be made by many clinicians, social scientists, and ethologists, that observing the phenomena in their "natural setting" may also have distinct qualitative advantages over the artificial situation presented by the laboratory. This, of course, is not to say that the one is any more or less "real" than the other. Anything which happens is real. We merely recognize that a tiger in the laboratory, or a tiger in the zoo, does not live in the same kind of stimulus field, and hence does not maintain the same kind of long-term psychological economy, as one in the Bengal jungle.) Example: If we investigate schizophrenia, with an eye to either its genetic or its environmental determiners, we have to take the schizophrenics as they come. This is because neither our scientific-information nor our ethics permits us to produce schizophrenia experimentally, or to predetermine who is a potential schizophrenic and assign such persons randomly to nonschizophrenogenic family environments. Example: If we are interested in economic behavior of say, incentive-pay problems), we cannot have any assurance that a short-term laboratory microcosm involving learning nonsense syllables and "payment" in extra grade points represents an adequate experimental analogue, let alone an identical kind of psychological situation (only reduced in temporal scale), to the question with which we started.

I make these observations of familiar truths to avoid any possibility of being misunderstood as saying that only laboratory experiments, in which control and randomization can be effectively imposed by the investigator, are intrinsically appropriate or scientific. Such a view is far from my philosophical position. There are good reasons, some practical and some methodological, for studying behavioral phenomena "in the state of nature." These reasons are sometimes so good that even the ex post facto design may be preferable to the laboratory method, and will in many cases be better than leaving an important problem completely unresearched (Campbell, 1969[1]; see also Campbell & Stanley, 1963, where the Chapin-Greenwood ex post facto design is totally rejected, but on the basis of "regression artifacts," a source of bias more akin to Kahneman's problem than to mine). The criticisms I shall advance are aimed at forestalling fallacious inferences of the kind commonly made from such designs, but more importantly, are made with the hope of inducing the mathematically competent and statistically creative among us to work on a problem whose importance is, I am persuaded, greatly underestimated by most social scientists.

There are three distinguishable aspects of what I take to be one core difficulty with the method of statistical matching in non-laboratory designs. Their precise logical relationship is not clear, but they are prima facie distinguishable, so I shall discuss them

---

[1] Note also that in that paper, although Campbell is dealing with more informative situations where time changes are available, he lists "Selection" (= differential recruitment of comparison groups) as a source of bias. As I see it, this rubric would cover two of the three difficulties I am raising for the static case.

separately. I do not thereby prejudge, nor will the sequel premise, that the three are fundamentally different.

For convenience of exposition, and without, I hope, being prejudicial to any issues, I assume in what follows that Kahneman's problem does not exist. That is, I presuppose (counterfactually) that we possess infallible (perfectly reliable and valid) measures of the "nuisance variable" which we intend to "control" by matching, analysis of covariance, or partial correlation. I do not see that it makes any fundamental difference what kind of statistical control we employ. I should imagine that any novel method of control which was "after-the-fact statistical" in character, that is, which relied upon some kind of generation of equivalent samples, or some kind of statistical correction for an alleged nuisance variable's influence, would suffer from the same methodological taint.

The first problem which arises is what I shall label *systematic unmatching*. This is most clearly exhibited by the method of matched pairs, in which we artificially constitute a nonrandom sample of the original population by selecting pairs of subjects who are pairwise equated on the nuisance variable. In such cases we are usually interested in the causal influence of an "input" variable X on an "output" or "consequence" variable Y, and we do not have experimental control of X, so that the organisms are somehow, directly or indirectly, *self-assigned* to "treatments" (levels of X). Here the usual reason why we match or partial out some third variable Z is our methodological suspicion that Z may exert a significant causal influence upon both X and Y and, consequently, that the prima facie association between X and Y (with Z left to vary freely, neither controlled experimentally nor partialed out by some statistical device) would reflect an output difference which is "spurious." My first thesis, in a nutshell, is the following: If one is a psychological determinist, or even a quasi-determinist, he must assume that for any but the most trivial and "unpsychological" examples of input variable X, the naturalistic self-selection of the organisms for treatments or levels of X must itself be determined. Hence, the result of holding constant an identified nuisance variable Z will, in general, be to systematically *unmatch* pair members with respect to some fourth (unidentified) nuisance variable W.

Stated in the abstract this thesis seems pretty hard to avoid, but it may sound like a hairsplitting academic point. So let me concretize it to show how serious a problem it presents for the researcher. Let us suppose we are interested in the "influence" of amount of schooling upon subsequent income. We cannot control who stays in school and who drops out before graduating from high school. Even if we could ethically and politically control it, by stopping some students and continuing others, we would be thereby defining a new type of population psychologically, whose statistics would hardly be generalizable to the "natural population" of our original problem.

We enter the files of students in a certain city school system and we divide them into those who did and those who did not complete the twelfth grade. We find that the high school graduates are earning markedly higher salaries twenty years later, that is, at the time of the investigation. We are not so naive as to take this finding at face value, because we recognize that there might be certain individual-differences variables, located "within the organisms themselves," that would be relatively stable over time and that would, on the one hand, influence income and, on the other hand, also influence the individual's self-selection for values of the input variable, that is, school level attained. An obvious example of such a nuisance variable is intelligence. We realize that the

differences in income might be due (partly) to the fact that the high school graduates were as a group more intelligent than the dropouts, and that this difference in IQ would be (partly) causative of continuance versus noncontinuance of education. So we enter the files for IQ and perform a statistical correction, either by a method such as analysis of covariance which utilizes the total *N*, or by defining subsamples of the original sample in which individuals are matched pairwise for IQ. If such matched groups differ in income, we conclude (fallaciously) that the difference is "not attributable to intelligence."

I say "fallaciously" because, of course, Kahneman's point applies here. That is, the unknown true intelligence level of an individual lies somewhere between the best estimate we could get by knowing how far he persisted in school, clearly one of the several fallible indicators of brains, and the IQ we find in the files, another fallible indicator of brains. His true intelligence lies somewhere between these, at a position which is sometimes estimable but more often not. But we are passing Kahneman's objection here and assuming counterfactually that the files contain infallible measures of intelligence.

Now if there is in fact a correlation between brains and persistence in school, matching dropouts with completers for infallible IQ surely results in the samples we generate being *un*matched for some *other* determining factor or factors capable of influencing the probability of school continuance. And, on the average, the members of a pair will presumably be more badly unmatched on these other factors (having been matched on IQ) than they would have been if we had let the chips fall where they may. Example: A stupid adolescent who continues through high school may do so because his parents put a very high emphasis upon educational achievement, and a bright one may drop out because his parents do not value such performance. The introjection of parental values is surely one of the major variables reflected in almost any kind of achievement, educational or vocational. This introjection would presumably function as a nuisance variable W which is left uncontrolled by matching on Z (= IQ). More importantly, the matching of groups on variable Z tends, on the average, to increase systematic unmatching on W. Thus, a dropout matched at 125 IQ with a continuer will be an extreme (low) deviate on, say *n Achievement* (McClelland, 1961; McClelland, Atkinson, Clark, & Lowell, 1953); whereas a continuer matched at 90 IQ with a dropout will be an extreme (high) deviate on *n Ach*. Within each pair, large systematic differences in *n Ach* (or any other unmeasured nuisance variable influencing self-selection for school continuance) will be practically guaranteed by the matching procedure. Or again, individual differences in "sociopathic-like" (low-anxiety, defective impulse-control, acting-out tendencies) will surely affect the dropout incidence (Hathaway & Monachesi, 1963). If two boys are equal on an infallible IQ measure but one has graduated from high school and the other one has dropped out, there is a good chance that they differ on this component, which is not one which our file data normally enable us to assess. I hope these examples show that, rather than being a minor blemish on the ex post facto design, the likelihood of systematic unmatching represents a major methodological weakness which is likely to corrode the entire investigative enterprise.

A second difficulty, which I shall call the *unrepresentative subpopulation problem*, is the first one as seen from the population-sampling point of view. If we match pairwise for a nuisance variable, such as a demographic factor that is known or supposed to be sizably correlated with each variable of interest, what we do (willy-nilly) by the matching procedure is to identify samples from subpopulations that differ systematically from the

entire population of interest. If the nuisance correlations are small, the "improvement" achieved statistically will be negligible, that is, the matching was relatively pointless. If it is large, the systematic departure of the resulting subpopulations from the original population in certain parameters will be correspondingly increased. In the extreme case we may be working with samples from a subpopulation which differs very markedly from the population of original interest. This means that our statistical generalization must be carefully confined to the unrepresentative subpopulations specified by the matching operation, and while that can of course be done, it will frequently leave us without an answer to the main question which aroused our research interest in the first place. Example: Suppose we have evidence to indicate that there is a relationship between the incidence of schizophrenia and socioeconomic class. We want to study the properties of a certain psychometric device, such as the Rorschach or MMPI, or some kind of cognitive performance such as abstraction ability or visual perception, in schizophrenics versus manic-depressives. I daresay that almost any competent Ph.D. candidate would take it completely for granted that his design would require a matching for socioeconomic and educational level. He finds the expected sizable difference between his manic-depressives and schizophrenics with regard to socioeconomic level on some suitable measure (e.g., the Hollingshead Index), and in order to "control" for its "spurious influence" (I put these phrases in quotes not ironically but to indicate that one does not have a clear notion precisely how the statistical control is related to the control of causal influence, discussed below as a third difficulty) he does not sample randomly from the entire hospital population of the two diagnostic groups but instead he constitutes a matched sample in which each schizophrenic is paired with a manic-depressive having the same social class index. Depending upon how he goes about this matching, our investigator may or may not be able to specify a statistically definable subpopulation, but let us assume that he can. He then samples randomly from these subpopulations to get the actual group of patients he studies on the output variable (abstraction ability or perceptual speed or Rorschach F+ or whatever it may be). Now it is obvious that this subpopulation is an atypical one, because the matching procedure will practically guarantee that on the average his schizophrenics are of somewhat higher socioeconomic class than the schizophrenic hospital population generally; and, similarly, the manic-depressive subpopulation from which he samples is now a biased subpopulation from the universe of manic-depressives. That is, the schizophrenic group sampled is pulled upward from their population social class value and the manic-depressive group is pulled downward from their population social class value; otherwise, of course, successful matching would not have been achieved. The expected result of such a procedure is a marked reduction in variance, which is the usual empirical finding (see Chapin, 1955, chapters III–V, for several examples). One cannot avoid the consequence that either this degree of departure is large enough to be worth worrying about or it isn't. If it is *not* large enough to worry about, there was no merit to engaging in the matching operation; if it *is* large enough to be worth worrying about, then one has a new problem by virtue of the fact that he is now studying unrepresentative (higher class) schizophrenics and unrepresentative (lower class) manic-depressives. And of course psychologically this is a very serious difficulty. Presumably some schizophrenics, as well as semi-compensated or compensated schizotypes (Meehl, 1962, pp. 827-838, 1964; Rado, 1956, 1960; Rado & Daniels, 1956), either remain in or gravitate to a lower social and educational class because of the general social incompetence associated with

schizotypy (Dunham, 1965), the obvious exceptions being individuals possessed of rare gifts that society rewards in special domains, for instance, esthetic or intellectual talents. By contrast, manic-depressives are "clinically well" (and, except to the very skilled eye, not detectably different from normal persons) between their psychotic episodes; further, there are certain features of the manic-depressive inter-psychotic character structure which are highly rewarded economically in the American culture, such as the social extraversion, the competitive striving, and a special sort of narcissism which these persons possess in spite of their superficial affiliative tendencies. A schizophrenic who remains in the sample after the matching operation is likely to differ from his more typical schizophrenic brethren in dimensions such as achievement motive, ego strength, energy level, frustration tolerance, social skills, perseverance, and goodness knows what all, variables likely to be significant influences with respect to the psychometric or experimental output measure under study. And the same is true, but in the other direction, for the manic-depressives. Additional biasing effects, almost inevitable given the relatively poor reliability of psychiatric diagnosis, will be a heightened proportion of misdiagnosed cases in both directions, and an inflated proportion of so-called "schizo-affective psychoses," who are atypical of *either* a manic-depressive or a schizophrenic population. I do not see how it is possible to make any valid correction for this kind of influence, since we are here talking about numerous unknown nuisance variables that become jointly definitive of unspecifiably deviant subpopulations. But what we *can* say, if we are psychological determinists, is that the two groups of patients under investigation are both unrepresentative of their respective diagnostic categories.

The third component of this problem is so obvious that one would be embarrassed to dilate upon it, except for the fact that a remarkable number of social scientists seem almost oblivious of the point. I shall call it *causal-arrow ambiguity*. While every sophomore learns that a statistical correlation does not inform us about the nature of the causality at work (although, except for sampling errors, it does presumably show *some* kind of causal relation latent to the covariation observed), there has arisen a widespread misconception that we can somehow, in advance, sort nuisance variables into a class which occurs only on the input side of the causal arrow and another class which occurs only on the output side.[2] This is, of course, almost never the case. The usual tendency, found widely among sociologists and quite frequently among psychologists (particularly among those of strong environmentalist persuasion), is to assume *sub silentio* that there is a set of demographic-type variables, such as social class, domicile, education, that always operate as nuisance variables to obscure true relationships or generate "spurious" ones, functioning primarily or exclusively on the *input* side from the standpoint of causal analysis. This automatic assumption is often quite unjustified. Example: We study the relationship between some biological or social input variable, such as ethnic or religious background, upon a psychological output variable, such as IQ or *n Achievement*. We find that Protestants differ from Catholics or that whites differ from blacks. But we find further that the ethnic or religious groups differ in socioeconomic class. We conclude, as

---

[2] For a brief but very clear analysis of the possible ways in which correlations among three variables may arise see Kempthorne (1957, pp. 283-286). In discussing "adjustments" (for nuisance variables) Kempthorne warns (p. 284) that "the adjustment of data should be based on knowledge of how the factor which is being adjusted for actually produces its effect. An arbitrarily chosen adjustment formula may produce bias rather than remove the systematic difference. It is this fact which tends to vitiate the uses of the analysis of covariance recommended in most books on the analysis of experiments."

an immediate inference and almost as a matter of course, that we have to "control" for the socioeconomic class variable, in order to find out what is the "true" relationship between the ethnic or religious variable and the psychological output variable. But of course no such immediate inference is defensible, since on certain alternative hypotheses, such as a heavily genetic view of the determiners of social class, the result of such a "control" is to bring about a spurious reduction of unknown magnitude in what is actually a valid difference.

Another example is the objection to the use of certain kinds of test items on measures of intelligence, when that objection is put *solely* in terms of the statistical fact that social class differences exist on these items. I cannot enter here into the substantive merits of that controversy, which is extraordinarily complex, and to which no adequate general solution seems to exist at present. No one would deny that if a certain kind of cognitive performance involves a content to which lower class children have inadequate environmental exposure (a notion which would have a high prima facie plausibility even without any research), such an item is not a "good item," assuming we are interested in the assessment of basic capacity variables. But what I do wish to query is the usual assumption among many psychologists and sociologists that *of course* whenever we find that a given kind of test item discriminates social class, it follows rather directly that it is an inappropriate item, such that measures compounded out of such items are to that extent "biased" or "invalid." This immediate inference is fallacious.

That it is fallacious can easily be discerned by considering the statistical consequences of a counter-hypothesis, and noting that they are indistinguishable from those of the conventional one. Suppose, to take the extreme case, that socioeconomic level were *completely* determined by abstract-conceptual intelligence, and that abstract-conceptual intelligence were *completely* determined by the genes; then it would follow as a consequence that high-valid items would discriminate social class perfectly. Analogy: We make a file study of the incidence of positive tuberculin tests in a random sample of patients seen in an outpatient clinic, and discover that test positives occur more frequently among the *lumpenproletariat* than they do among Cadillac drivers. We do not conclude forthwith that the tuberculin test is "invalid" because it is "biased" against the poor! Why not? The reason nobody concludes this is, of course, that we all *already know* how the direction of the causality runs. Similarly, in agricultural experiments we know that an analysis of covariance in which the nuisance variable statistically controlled is, say, a soil characteristic will give us the "right answer," because our well-corroborated causal model *tells us in advance* the direction of the causal arrow. Nobody in his right mind supposes that the yield of corn in August causally determines random table entry or certain properties of the soil present during the preceding spring and summer; therefore we are confident that an analysis of covariance will give us the causal answer in which we are interested as agricultural experimenters. The same is rarely the case when the behavior scientist partials out or matches with respect to a nuisance variable, because the latter may itself be (and, in general, *will be*) a *dependent* variable with respect to a variety of nuisance factors which we can perhaps say something plausible about, but which we do not know how to measure or control. There is no general justification for the routine assumption that demographic and allied variables such as religion, size of community, educational level, ethnic and religious background, and social class should be taken as

always functioning solely on the input side and, therefore, as always appropriately "controlled" by a matching operation or by some similar type of statistical correction.

I would go further than this and suggest that it is not only incorrect to insist that groups must routinely be matched on such demographic or other nuisance variables, but that, for all we know, in some unknown proportion of designs the net effect of such matching is not to improve the validity of the inferences made but is actually to introduce systematic error. I do not wish to maintain that matching makes matters worse more often that it makes them better, but I consider it an open question on the present evidence. If I were advising a doctoral candidate who asked me whether he should control for educational and social class in comparing schizophrenics and manic-depressives with respect to the presence of psychometric thought disorder, I would honestly not know what to tell him. I suspect I would have to tell him that if he didn't match, he would be in danger of flunking his doctoral oral, because most of the members of the committee would be operating on the traditional assumption that he should have done so; but that so far as I myself was concerned, the unrepresentative character of the resulting matched samples would be such that I wouldn't know what he would be entitled to conclude if he got a difference, and even less if he failed to get a difference, on the output variable having followed such a matching procedure.

This line of argument does not conflict with what we teach students in courses on experimental design with regard to the purely *statistical* influence of matching procedures upon design sensitivity or what Fisher calls "precision." It is, of course, true that a matching procedure will (if successful) have the effect of reducing the error term which appears in the denominator of a significance test, and in that sense will give us higher power. But that statistical truism is in no way incompatible with the claim I am making here, namely, that we are thereby defining different subpopulations and consequently that the parameters we are estimating may not be the parameters we were originally interested in.

Perhaps the most succinct (but still general) way of formulating the problem of controlling nuisance variables statistically in a nonexperimental context would be "How are we entitled to interpret the associated counterfactual conditional?" I set aside the superpositivistic approach that purports to eschew any such counterfactual, claiming to confine itself to the observations plus the formalism—a sort of "psychologist's Copenhagen interpretation"—since I have not found any theoretically interesting cases in which this "minimum interpretation" is *consistently* adhered to. And this is hardly surprising, since if one genuinely intends to utilize the statistical formalism *solely* for predictive purposes, there is no rational basis for introducing such statistical "control." That is, it makes no sense to speak of a correlation as "spurious" or "in need of correction" *unless* a possible error in causal-theoretical interpretation is envisaged. Thus the correlation between years of schooling and subsequent salary—I of course neglect the separate problem of ordinary sampling errors—stands on its own feet, and if you want to forecast income from schooling, the "influence" of IQ as a shared statistical component can be neglected (= allowed to operate) at the purely descriptive level. In every instance that I have come across in which the investigator felt it necessary to employ partial correlation, analysis of covariance, or artificially concocted matched samples to "avoid the influence" of an alleged nuisance variable, the rationale of such a procedure lay in his wish to conclude with a causal-theoretical inference or, at least, a counterfactual conditional of some kind.

When a social scientist of methodological bent tries to get clear about the meaning, proof, and truth of those counterfactuals that interpret statistical formalisms purporting to "control the influence" of nuisance variables, he is disappointed to discover that the logicians are still in disagreement about just how to analyze counterfactuals.[3] It appears that the logical (and epistemological) analysis of counterfactuals is a task involving some of the deepest and oldest of philosophical problems (e.g., the modalities, extensional logic's adequacy, substance and property, character of natural laws, identity, the kinds of "contingency" and "necessity," the meaning of 'accidental' in a determinist framework, the theory of proper names and definite descriptions). I had intended to include something that I hoped would be new and constructive at this point of my discussion, but deadline obligations and my status as a philosophical amateur have combined to make me more realistically modest in aims. I hope, however, that what I have to say at present about counterfactuals does not depend on precisely how the logicians ultimately agree to "fix them up." I am encouraged in this hope by the fact that agreement does exist about the important role of the explicandum, and—to a considerable extent—about criteria for a satisfactory explication. One main area of agreement—of direct relevance to the social scientist's problems—is the intimate connection between a counterfactual's legitimacy and the natural-law/accident-universal distinction. One way (the main way, some hold) in which a natural law differs from an accidental universal is that the former legitimates a counterfactual while the latter does not. "If Kosygin had not learned Russian, he would be unable to speak it" is presumably a sound social-science counterfactual, relying on the laws of psycholinguistics. But we cannot rely on the accidental universal "All persons who discuss politics with Meehl speak English" to legitimate a counterfactual "If Kosygin were to discuss politics with Meehl, he would speak English."

As I read the record, there are some counterfactuals we wish to exclude because we doubt that they are meaningful, but we want to assure that criteria adequate to exclude them will not inadvertently forbid other similar-appearing counterfactuals which *do* seem intuitively meaningful, and of great importance in the discourse of science and common life. Take, for example, what may be labeled (nonprejudicially) as 'counter-identicals,' that is, counterfactual statements concerning a named or definitely described individual, where the protasis falsifies one of his properties. In spite of Leibniz, the scientist, lawyer,

---

[3] While the problem of interpreting conditionals is an ancient one (see, e.g., Hurst, 1935; and Mates, 1949), and the issues of present controversy were adumbrated by W. E. Johnson in the 1920s (see Johnson, 1921, chapter III, 1924, chapter I), the current concern over the logical analysis of counterfactuals and their relation to natural laws was precipitated by the papers of Chisholm (1946) and Goodman, (1947). An extensive literature on the subject followed these seminal contributions. See, perhaps best read chronologically, Will (1947), Hampshire (1948), Beardsley (1949), Hiż (1949), Popper (1949), Kneale, (1950), Pears (1950), A. R. Anderson (1951), O'Connor (1951), Weinberg (1951), Storer (1951), Bergmann (1952), J. Anderson (1952), Brown and Watling (1952, 1950–1952), Diggs (1952), Schneider (1953), Watling (1953), A. R. Anderson (1954), Chisholm (1955), Cooley (1957), Sellars (1958), Watling (1957), Downing (1959), Popper (1959b), Kneale (1961), Rescher (1961), Walters (1961), Mackie (1962), Simon and Rescher (1966), Nerlich and Suchting (1967), Popper (1967), and Molnar (1969). Much of this linguistic analysis, while inherently interesting, has little or no value for the social scientist seeking methodological clarification on his scientific use of counterfactuals. I found the papers by Hiż, Popper, Kneale, Sellars, Mackie, Simon and Rescher, and Molnar most illuminating. See also brief or related discussions in Braithwaite (1953, pp. 295-300), Burks and Copi (1950), Burks (1946, 1951, 1955), Carnap (1936-1937, 1966, pp. 196-215), Hempel and Oppenheim (1948), Hempel (1950, 1966, pp. 56-58), Kneale (1949, pp. 70-78), Lewis (1946, pp. 211-233), Nagel (1961, pp. 68-73), Pap (1962, chapters 15 and 16, pp. 273-306; 1958a), Popper (1959a, pp. 420-441), Quine (1959, pp. 15-17), Reichenbach (1947, chapter VIII, pp. 355-404), Sellars (1948), and Broad (1933).

physician, and ordinary man will—I think correctly—insist that many such counter-identicals are meaningful and useful. We surely do not wish to adopt a semantic convention which denies the status of wff to, say, "If defendant had driven his car with ordinary care, plaintiff would not have sustained injury," or "It was fortunate for me that I had a flu shot, since everyone else in the family fell dreadfully ill with flu." Contrast these counter-identicals with this one (example courtesy of Dean Kenneth E. Clark): "If Meehl and I had lived in the sixth century, he would have been an archbishop, and I would have been Merlin's research assistant." Is this counterfactual legitimate? Hard to say, but if so, it will take some doing to unpack satisfactorily. Worse is "If Caesar had been born in 1900, he would have been a fascist." Still worse is "If my maiden aunt were a tram car, she would have wheels."

If one conceives of an individual as a bundle of properties, there is a difficult problem in unpacking all such counter-identicals. I believe the best way to do it is to begin with a distinction between the actual world and other imagined (hypothetical) worlds belonging to the same world family, where 'world family' designates the infinite set of conceivable worlds sharing nomologicals but differing in particulars (Nerlich & Suchting, 1967, pp. 233-235; Popper, 1959a, p. 430; 1967; Sellars, 1948). Assuming that this can be done satisfactorily (and no one has, to my knowledge, offered a criticism of Sir Karl Popper's 1967 paper attempting to rigorize it), I think we could then offer a translation of counter-identicals in terms of world lines in some unrealized world of our world family, sharing coordinates with the named or described individual's actual world line up to the critical event (e.g., failure to obtain his flu shot as planned), and diverging thereafter. His properties and most of his relations would be identical with those of the actual individual up to that space-time point, but would diverge—perhaps increasingly—thereafter. In stipulating semantic rules for the well-formedness of a counter-identical, there would doubtless be a certain arbitrariness about which of an individual's properties are, so to say, "privileged properties," such that a counterfactual denying them is forbidden. Intuitively one feels that it is essential to the person called 'Caesar' that he be an ancient Roman, but it is not essential to Meehl that he receive a flu shot. Of course a rule excluding "If Caesar had been born in 1900…" is laid down in the interest of avoiding strange and counterintuitive discourse and preventing unprofitable puzzles, and we do not wish to forbid too much. Thus it makes sense to begin a counterfactual with "If an American child born in 1900 had Caesar's complement of genes" (wildly improbable but not, I submit, counter-nomological) but this admissible case need not be forbidden by a rule adequate to forbid the counter-identical beginning with "If Caesar [proper name, denoting an individual who satisfies a certain definite description] had been born in 1900…" My hunch is that a sufficiently tolerant set of exclusion rules could be rigged up, keeping in mind that an adequate logician's translation of legitimate proper-name or definite-description counter-identicals need not—I think will not—show the individual's name recurring on the right-hand side of the equation; just as in Russell's theory of definite descriptions itself, we have learned to accept the fact that an unpacking adequate to avoid paradoxical metaphysics leaves us without 'the present King of France' as a single semantic element on the right-hand side. But the development of these suggestions must wait for another occasion.

Accepting provisionally the world-family concept and the associated distinction between nomologicals and accidental universals, we see that the interesting cases for

social-science methodology would remain problematic even after the cute counter-identical puzzles of logic seminars had been liquidated. This is because the social-science cases of interest are not (by and large) in danger of counter-definitional meaninglessness but, instead, may suffer from counter-nomological falsity or contradictoriness. That is, the problematic counterfactuals of psychology and sociology do not typically find us wondering "What does it *mean*, does it make any *sense*?" but rather, "Is it consistent and true? *Could* [nomologically] the counterfactual hypothesis be satisfied, given the nomologicals presupposed? And, *if* it could, does the counterfactual conclusion *follow* within that nomological system?" (Hiż, 1949). Since our warrant for asserting counterfactuals consists of the nomologicals of our world family, plugging in counterfactual particulars so as to yield a different world of the family, we must avoid unwittingly contradicting ourselves in the antecedent. Consider statements like: "Imagine that these organisms, which in fact have properties P_1, P_2, … P_k, Q_1, Q_2 , … Q_m, had instead possessed properties P_1, P_2, … P_k, Q_1', Q_2', … Q_m'; then …" (Note that this way of talking is ubiquitous in biological and social science—we cannot even understand the notion of a *control group* without admitting such formulations!) To get to the counterfactual conclusion following '… then …' we rely on natural laws. But what if the natural laws relied on forbid the counterfactual antecedent P_1, P_2, … P_k, Q_1', Q_2', … Q_m'? How do we know that these are compossibles, that is, that the counterfactual conjunction is not nomologically forbidden?

I am not talking about what might be a logician's technical problem, that is, the non-existence of a general algorithm for stepwise deciding whether this conjunction would instantialize a counter-theorem. No, the problem is not so esoteric as that. The problem lies in the incompleteness of the social scientist's nomological network. Underlying (derivationally and causally) the known laws of social science are the unknown ones—the "true reasons why" the known laws are the way they are. Furthermore, very odd but true, *some of the laws are, from a philosopher's viewpoint, not nomologicals but accidental universals*. This is because many "laws" of biological and social science are structure-dependent and history-dependent in a special way, so that while their logical form (taken singly) is that of laws of nature, they are not derivable from the fundamental nomologicals (laws of physics). Many "taxonomic" laws are pseudo-nomological, which is one reason why examples like "All crows are black" are unsuitable for most philosophy-of-science discussions. Unfortunately it is not always easy to ascertain when a biological or social-science generalization (taken as true and well evidenced) is really akin to "All silver melts at 960.5° C"—a nomological—and when it is akin to "All the coins in my pocket are silver," an accidental universal. It may be objected that the melting point of silver is also structure-dependent, but this, while true, does not prevent the generalization's being a true nomological, because we can (theoretically) include a characterization of the micro-structure in our "theoretical" definition of the technical term 'silver,' in which case the structure dependence is fully represented in the antecedent. That is, we have "If a substance is silver [= has such-and-such micro-structure], it melts at 960.5° C," a proposition presumably entailed within (complete) physical theory as a consequence of the fundamental nomologicals. Viewed this way, the generalization is a theorem within a formalized physical theory (and, note carefully, would be nontrivially true for all worlds in our world family ever if no silver existed in some of them). In biology, the statement "A mammal dies if deprived of oxygen" is of this sort, since its structure dependence can analogously be represented in an adequate theoretical (anatom-

ical + physiological) definition of 'mammal.' By contrast, the taxonomic generalization "All mammals have paired gill-slits at some stage of their development" is an accidental universal, as is "If a species of animal has a heart, it has kidneys." These taxonomic property correlations are—like Meehl's friends' English-speaking and his silver coins—"historical accidents," reflecting the course of evolution which could have been different given the same fundamental nomologicals but differing initial conditions of the earth.[4]

Most of the statistical "laws" (correlations) investigated in disciplines such as differential psychology, personology, clinical psychology, and sociology are more akin to the accidental universals of taxonomy than to genuine derived nomologicals. The social scientist who works in these fields studies covariations between selected *dispositions* manifested by individuals ("traits," "capacities," "temperamental or cognitive parameters") and also the correlations of these with a variety of *status* variables and *life-history antecedents*. The nomological network and initial conditions that gave rise to these statistical associations are horrendous in number and complexity. They involve factors ranging in kind from genetic drift in the remote past when a certain ethnic group was forming to the child's internalization of religious and political ideologies.

It is hardly necessary to give examples, which abound on every side, but I will provide one extreme case to convey the flavor. Suppose a clinical psychologist working in neurology finds (as he would if he bothered) that the normal siblings of children with Tay-Sachs's disease (infantile form of amaurotic family idiocy) are somewhat less prone to physical aggression than random "control" children, and that they show a pattern of superior verbal and inferior spatial abilities on standardized tests. He might be misled into some pretty fruitless genetic, neurological, or social speculations if he were somehow ignorant of the religio-ethnic category *Jewish*. As it happens, we can provide a plausible explanation of these strange correlations, but we have to rely on several different sorts of information from very different disciplines. The fact that Tay-Sachs's disease is almost (not quite) confined to Jews presumably arises from some ancient accident of genetic drift under migration (this mutation can hardly have any reproductive advantage), combined with the cultural fact of a zealous religiously based avoidance of miscegenation. The lesser physical aggression of Jewish children is cultural, partly based upon traditional contempt for violence ("The goyim use their fists as a substitute for brains," as one of my Jewish patients put it) and the Jews' centuries-old persecuted minority status which renders physical counteraggression a poor tactic. There are data showing a rather pronounced verbal/spatial disparity among Jews (Lesser, Fifer, & Clark, 1965; Lesser & Stodolsky, 1967) so that the "Jewish factor" also underlies the association between this trait relation and Tay-Sachs's disease in a sibling. The differential ability pattern for Jews itself remains to be explained, however. Easy cultural explanations are available (e.g., Talmudic value of words) but one cannot entirely exclude a genetic contribution as partially responsible. In any case, our present-day trait correlations are the end result of the confluence of factors ranging from random genetic mutations and drift to the "historical accident" that a Middle East tribe of gifted nomads invented ethical monotheism five or six thousand years ago!

---

[4] I can still recall vividly my astonishment, during the first conversation I ever had with Professor Carnap in the middle 1950s, when—in response to my objection to his tentative definition of 'derived nomological' that it would render many "laws" of biological and social science as accidental universals—he replied calmly, "But of course they are; it is, however, quite harmless to call them laws, for most purposes."

The puzzling Tay-Sachs correlations are rendered easily explicable by the clear-cut character of the clinical entity (pathognomic signs, early appearance, regular course) and its simple mode of inheritance (Mendelian recessive of complete penetrance). When we deal with nonpathological traits or trait clusters involving only moderate correlations among continuous variables ("loose-knit syndromes") the causal unscrambling job is much harder. Consider, for example, the association between socioeconomic level, child-rearing practices, and impulse control (inhibition of overt aggression, ability to postpone gratification, frustration tolerance). Social learning doubtless plays the major role in producing these correlations, but it would require environmentalist dogmatism to rule out the possibility of some contribution of polygenic "temperament" factors. There may be inherited dispositions that act through several distinct causal chains, converging upon the same correlational result. Basic CNS parameters affecting one's capacity to inhibit, one's rage readiness, anxiety proneness, delay tolerance, social dominance, and so forth, could contribute by concurrently influencing (1) the educational and vocational level attained by the parents, (2) the social models they provide for the child, (3) the child's genetic disposition to respond to social controls, (4) the parental reactions to the child's modes of responding, (5) the over-all gratification/frustration level in the home, and so forth.

We now know that such "temperamental" traits as aggressiveness, social dominance, anxiety susceptibility, liking for alcohol, exploratory tendency, rate of recovery of sex drive after copulation, and general activity level are partially gene-determined in the mouse; that the Basenji dog breed differs markedly from the beagle hound in its capacity to develop a canine "conscience" through affectionate socializing experiences with humans; and that in the human species, a sizable genetic component of variation ("herit-ability") obtains for several personality traits, including general intelligence, several "special abilities" (e.g., dexterity, mechanical, spatial, verbal), pattern of vocational interests, self-control, anxiety proneness, impatience, social introversion, the phenomen-ology of emotional experience, and the needs for autonomy, affiliation, aggression, and self-exhibition. (I have recently seen a manuscript reporting unexceptionable research findings to the effect that Chinese neonates are more "placid" than Caucasians when tested under standard conditions during their first 72 hours after delivery!) The weight of presently available evidence and the rapid rate at which more of the same is accumulating is such that any rational social scientist should view as a *wide-open research problem* the role of genetic variations in determining inter-trait, trait-history, and trait-status correla-tions. (See Bloch, 1969; Freedman, 1958; Gottesman, 1963; Lagerspetz, 1964; Lindzey, Winston, & Manosevitz, 1961; McGill & Blight 1963; Scarr, 1966, 1968, 1969; Shields, 1962; Slater & Shields, 1969. On behavior genetics generally, see Eckland, 1967; Fuller & Thompson, 1960; Glass, 1968; Hirsch, 1962, 1967; Manosevitz, Lindzey, & Thiessen, 1969; McClearn, 1962; McClearn & Meredith, 1966.)

I stress the genetic factors partly, in all frankness, to combat the environmentalist brainwashing which most of my philosopher readers will have received from their under-graduate social-science classes; but mainly because the commonest error in handling nuisance variables of the "status" sort (e.g., income, education, locale, marriage) is the error of suppressing statistically components of variance that, being genetic, ought not to be thus arbitrarily relegated to the "spurious influence" category.[5]

---

[5] See Burks and Kelley (1928). Professor Jane Loevinger, upon reading a draft of the present chapter, called my attention to this 42-year-old contribution, which I confess never to have read. My sole justification for retaining

Since socio-psychological correlations are the outcome of so complex a causal situation, the formulation of legitimate counterfactuals is extraordinarily difficult. It should be noted that this complexity obtains not merely because of the sheer *number* of relevant factors so commonly mentioned, but also because in the life histories of a group of subjects there are numerous possibilities of *correlated initial and boundary conditions* (e.g., an upper class subject has heard better grammar and may also possess family-name leverage at college admission), *subject-selected learning experience* (e.g., if you never give studying a try you can't discover that getting A's can be fun), *social feedback loops* (e.g., aggressive personal style elicits counteraggression by social objects, which may further increase the subject's own aggression), *autocatalytic processes* (e.g., poor performance yields situational anxiety as a by-product, which further accelerates performance decline), and *critical junctures in* "*divergent*" *causality* (e.g., atypical carbohydrate breakfast → mid-morning hypoglycemia → temper outburst at boss → failure to get expected promotion → last straw for ambitious wife → divorce scandal → alcoholism → suicide).[6]

---

here those portions that essentially repeat the old Burks-Kelley arguments is that the social-science literature shows that many of my brethren must never have read them either. I have sometimes wondered whether it is only in the inexact sciences that rather simple methodological truths have to be noticed afresh after the passage of an "academic generation" or two. Does this strange phenomenon occur also in physics and chemistry? In psychology one is uncomfortably aware of the truth of Gide's remark, "It has all been said before, but you must say it again, since nobody listens." For an often-ignored job fifteen years after Burks and Kelley, see Loevinger (1943). An excellent methodological discussion of genetic factors in relation to social class—the nuisance variable most often "controlled for" in social-science research—is Gottesman (1968). I have found remarkably little explicit discussion of the causal-arrow ambiguity problem in writings by professional statisticians, the most helpful exception being Kempthorne (1957). Presumably this is because they (a) take it as perfectly obvious, (b) think in terms of agricultural research, where background knowledge usually excludes one of two causal directions, and (c) deal with experimental manipulations rather than "passive observation" of cross-sectional relations presented by experiments of nature, such as we perforce study in differential psychology and sociology. When the causal-arrow ambiguity problem is briefly considered in connection with analysis of covariance by statisticians, their concern is over the possible influence of "treatments" upon the "concomitant [= nuisance] variable," such that a regression-based adjustment of output means would lead to an underestimate of treatment effects. This is not quite the same as our present problem, although closely related. See, for example, Bartlett, (1936); Cochran (1957); Scheffé (1959, pp. 198-199); and Ostle (1963, pp. 456-457). What is more surprising is the lack of explicit discussion in expositions of analysis of covariance written by psychologist-statisticians for social-science readership. For a brief, clear, and persuasive refutation of the still-prevalent notion that statistical weights in multivariate predicttion systems somehow quantify "[causal] influence," see Guttman (1941). This 29-year-old SSRC bulletin is insufficiently known and still very much worth study.

6 Roughly, in divergent causal chains, small initial-condition fluctuations determine very different remote outcomes; in convergent situations, small fluctuations "average out" so that whether any one individual initial event is E or ~E has a negligible effect on the system's direction of movement. See Langmuir (1943), London (1946), Meehl (1954a/1996, pp. 37-67). Langmuir's distinction is of course implicit in numerous historical and fictional treatments of the theme "small causes, great effects." A familiar example is speculation about whether World War I would have broken out if the obstetrician who delivered Wilhelm II had been more skillful, as a result of which the Kaiser would have been spared his withered arm, hence would have felt less need to overcompensate, and so forth. "For want of a nail the shoe was lost; for want of a shoe the horse was lost; for want of a horse the rider was lost; for want of a rider the battle was lost; for want of a victory the kingdom was lost." Machiavelli (*The Prince*) points out that all of Cesare Borgia's careful planning went for nothing because he could not have foreseen that he would be lying desperately ill at the very moment the Papacy was vacated by the death of his father (Alexander VI). For fictional emphasis on the critical role of minor, quasi-random fluctuations, see Sterne, *Tristram Shandy*; Tolstoy, *War and Peace*; O'Hara, *Appointment in Samarra*; and J. H. Wallis, *Once off Guard*. See also London (1952); Jordan (1955, pp. 108-113); Platt (1966, pp. 174-177); Pirenne and Marriott (1959); and Ratliff (1962, pp. 442-445). As an extreme, dramatic, but perfectly possible example of Jordan's *Verstärkung* or Langmuir's divergence, suppose Adolf Hitler to have been lost on a dark night while serving as message-runner in World War I. A quantum-indeterminate event in his retinal receptors is amplified neurally as a result of which

The correlational statistics relating trait, status, and history variables within a defined social group depend causally upon the "accidental universals" (more precisely, the "accidental joint frequency distributions") that happen to prevail in that society, given its gene pool, geographic setting, economic system, class structure, political institutions, legal forms, and so on. In attempting to formulate quantitative counterfactuals on the basis of these statistics, we implicitly assume that imagined alterations in selected particulars would be nomologically possible without an entailed disturbance in the statistical structure (the numerical claims of the counterfactual being based upon that structure's parameters). As of this writing it remains unclear to me when, if ever, this assumption is warranted, although it does seem that some situations make it more plausible than others. The trouble is that, while I cannot produce any clear criteria, I have the impression that the "safest" cases are those in which well-confirmed theoretical knowledge already exists. (In agricultural experiments we can be confident about the *causal* status of soil heterogeneity as a nuisance variable; hence calculating what Fisher labels "adjusted yields" in an analysis of covariance leads fairly directly to a legitimate counterfactual concerning the output averages.) If I am essentially correct in this impression, the social scientist's position is discouraging because he wants typically to rely upon his quantitative counterfactuals as a basis for causal theorizing rather than the other way round.

To concretize the discussion, consider again the example of treating a student's IQ as a nuisance variable in a research study which aims to ascertain the relationship between educational level attained and subsequent adult income. Since the textual interpretation of the counterfactual corresponding either to an analysis of covariance or to the now largely abandoned partial correlation presents an identical problem, I shall use partial correlation because the statistics of the situation is easier to discuss. The working formula for a partial correlation, being expressed in terms of algebraic operations (taking products and differences) upon the three zero-order correlations, obscures what really underlies the process of "partialing out" a nuisance variable such as IQ. In deriving the partial correlation formula, what do we do? Let x = educational level attained, y = adult income, and z = IQ. In the algebra underlying the final partial correlation formula, which purports to tell us "what the correlation between income and schooling *would be*, except for the influence of IQ [as a nuisance variable]," designated by the partial correlation coefficient $r_{xy \cdot z}$, what we do algebraically in the derivation is to construct a set of residuals constituting a difference variable u, obtained by regressing the first variable of interest x upon the nuisance variable z; we then consider the set of residuals constituting a constructed variable v obtained by regressing the other variable of interest, y, upon the nuisance variable z; and then we correlate these residuals. The resulting coefficient of correlation r_{uv} is called the partial correlation between x and y with z held constant (= $r_{xy \cdot z}$). Since it turns out in the algebra that the magnitude of this new coefficient is computable directly from the zero-order correlations without actually going through the steps of computing all of these residuals u_i and v_i on the individual subjects, the cookbook user of partial correlation is not, so to speak, forced by the working formalism (unless he refreshes himself on the derivation) to look the counterfactual problem squarely in the face when asking himself how this final derived number is to be textually interpreted.

---

he turns his head toward a faint light source just as an enemy sniper fires slightly off target. Six million Jews would have escaped liquidation thereby!

Let us examine one of those residuals as it appertains to an individual subject of our research investigation. Plugging in the value of his IQ in the best fitting *x*-on-*z* regression equation (I assume linearity as a condition for the Pearson *r* to be an adequate descriptive statistic), we "estimate" how far he should go in school. Similarly, plugging his IQ into the regression equation of *y*-on-*z*, we estimate how much money he should be earning at age 35. We then find that he didn't go precisely as far in school as our regression equation would "predict," nor does he earn exactly as much money at age 35 as the other regression equation would "predict." That is, there is a discrepancy u_i between what we would expect him to earn and what he actually earns, and a discrepancy v_i between how far we would expect him to go in school and how far he actually went in school. It is these two discrepancy values u_i, v_i which are correlated over the entire group of individuals. The question of interest is, how is each of these to be interpreted as applied to him? Can we say, for example, "If this subject had had a higher IQ by so-and-so many points, then he would have proceeded farther in school, by such-and-such many grades"? Does the regression line of schooling upon IQ legitimate such a counterfactual? I do not assert dogmatically that it does *not*; but it seems to me evident that there is considerable doubt about whether it *does*. Do we mean, for example, "If everything else that happened to him was exactly as it in fact was, but his IQ had been so-and-so many points higher, then he would have gone such-and-such many more grades in school"? Is that the intended translation? If it is, is it a valid counterfactual legitimated by the regression equation? I for one do not know, and I doubt that anybody else knows either. It might very well be a counter-nomological, since it might require a violation of some laws of social psychology for his parents, teachers, and peers to treat him exactly as they in fact did, given that his IQ was significantly changed from what it in fact was. It does not, for example, require any far-out speculating to be fairly certain that a child with an IQ = 140 living in a somewhat anti-intellectual proletarian family would be reacted to rather differently by his siblings, and by his high-school dropout father, than would a child, similar genetically in all other respects, but with an IQ of 95! It won't do to solve this by main force, simply saying, "Well, we are going to insist upon translating the counterfactual so as to ensure that everything else happens to him exactly as it did, given his actual IQ." An easy way to exclude this heavy-handed approach is to point out that there is a necessary quantitative interdependence between such factors as social reinforcement and the behavior of the individual under study. To say that we are going to assume that everything else is just as it was in this type of situation is rather like saying that we are going to assume that everything about a pigeon's reinforcement schedule in a Skinner box could be "just as it was," while concurrently assuming that the pigeon responded at twice as high a rate. Under such circumstances, either you have to decide that the pigeon will end up receiving a larger number of total reinforcements, or—if it is insisted that the total pellets delivered are to be held constant in the counterfactual—then there must be an alteration of the reinforcement schedule. You can't have it both ways. Furthermore, in the human case matters become very complicated because of the fact that humans can talk to themselves about the schedules they're being put on by their social environment. If we insist, say, that the proportion of times a school teacher says "right" versus "wrong" in the child's second-grade school experiences be held constant, then giving him another 30 points in IQ will require that the teacher say "wrong" on quite a few occasions when the child "knows that he is in fact right." Obviously this will have a profound effect on his attitudes

regarding work, achievement, payoff, elders as representatives of the larger society, and so on.

I do not of course mean to argue that there cannot be *any* counterfactuals involving "corrections for nuisance variables" that are (a) meaningful and (b) true. My point is that it is frequently—I incline to say typically—difficult to decide about their meaningfulness, and even more difficult to decide about their truth. One can rarely interpret counterfactually a residual about a regression line or plane with confidence that he knows what the counterfactual means and that it is a valid consequence of the relevant nomologicals.

Part of the trouble here is as discussed above, that the statistical system under study is a resultant of the influence—interactive and frequently mutual, that is, involving feedback—of a large number of variables, known and unknown, and we happen to have selected three of them for study, none of them having been experimentally manipulated by us. From the standpoint of the statistician aiming at a safe (minimum) interpretation, a partial correlation coefficient between variables x and y with z held constant is nothing but the zero-order correlation obtained when we regress x upon y within a narrow z-slice, provided that relationship is invariant over z-slices (rarely tested!). That is, we define a plane located in the three-variable box which is parallel to the xy-plane and located z units out on the z-axis. The locations of the person points in this box are the end result of a multitude of causal factors, varying all the way from a single mutated gene that renders particular individuals mentally deficient to the interpretative vagueness of certain legal language in the Civil Rights Act. There is nothing about the formalism for characterizing the distribution of person points confined to a given z-slice—a process which is of course unobjectionable when given the statistician's minimum interpretation—that enables us to formulate a counterfactual without having to worry about what the whole box would look like if the world were different in certain important ways, biologically and socially, from the way it in fact is. It is easy to see this by considering what a very strong counterfactual, textually interpreting a partial correlation, would read like. We often speak of the partial correlation as telling us what the "true correlation would be if the nuisance variable were held constant." Suppose we attempt the counterfactual "If there were no IQ differences in the population, then the correlation between years of schooling and subsequent income would be = $r_{xy \cdot z}$." This strong counterfactual is clearly impermissible on two counts.

First, the antecedent is (effectively) counter-nomological in genetics, given the probabilistic mechanisms of gene assortment. (If this objection were to be rebutted by pointing out that the laws of genetics are themselves—strictly speaking—"accidental universals," structure-dependent outcomes of our world's cosmic history, one rejoinder would be that for the social scientist, operating at *his* level of explanation, the laws of biology can be taken as nomologicals.) Second, even if we allow the antecedent, we surely cannot assume that the statistical structure would be as it is if human beings all had the same g-factor. (For expository simplicity I have treated IQ as g-factor, which is of course a gross distortion. The IQ is a fallible measure of g-factor, and g-factor is itself the result of polygenic hereditary components interacting with life-history parameters. Needless to say, this oversimplification only weakens my argument.) In fact such a supposition would almost certainly be erroneous. The whole educational system would probably have evolved very differently. Teachers' attitudes and beliefs about students would be radically different from what they are. Employers' interpretations of the school record at job entry would be quite unlike what they are in our world. It would be pointless for me to compile

a long list of "social-facts-that-would-be-otherwise" in documenting something so obvious as the theses: *If a major source of achievement-related individual differences were removed, society would be considerably changed; and the statistical structure relating trait, history, and status variables would be so materially different that quantitative counterfactuals based upon the received structure's parameters are all invalid*.

This paper was criticized by two sociologist reviewers on the plausible but specious grounds that the matched-case method has been replaced in sociological research by the use of multivariate designs. Aside from the fact that current social-science generalizations and theory rely in part upon earlier investigations employing matching, and the fact that matching has by no means been completely replaced by multivariate analysis in social-science research, I must emphasize that these critics do not see the main point I am making. The core difficulty is not eliminated when we substitute multivariate analysis for case-matching, as should be obvious to anyone who understands the mathematics underlying the derivation of multivariate estimates. Thus, for example, in the analysis of covariance, the "influence" of a nuisance variable is sought to be removed algebraically, by calculating an F-test on the means of the output variable of interest upon residuals obtained when this output variable of interest has first been regressed upon the nuisance variable and the output means "adjusted" accordingly. As in the older partial correlation formula, what we are actually doing in the analysis of covariance may be obscured (to the "cookbook user" of statistical formulas) by the fact that computational method bypasses the actual calculation of these individual case residuals about the nuisance variables' regression line. It cannot be overemphasized in the present context that analysis of covariance as a method of control by statistics rather than by experimental manipulation *suffers from precisely the same inherent methodological vice in the social sciences as does the method of matched groups*. In the matched-group method, the investigator physically constitutes a nonrepresentative "artificial" subpopulation for study. In multivariate analysis, he concocts statistically, by the making of certain algebraic "corrections," a virtual or idealized sample, the members of which are fictional persons assigned fictional scores, to wit, the scores the investigator, algebraically infers they *would* have had on the output variable of interest if the alleged causal influence of the nuisance variable were removed. The empirical meaning of this "virtual," fictional, idealized, inferred-score population is totally dependent upon our giving a correct interpretation to the presupposed causal counterfactual (Simon & Rescher,[7] 1966). One might even maintain—although I do not wish to press the point—that modern multivariate analysis is *farther* removed from physical reality than the old matched-group procedure, because the latter at least deals with an actual physical subpopulation, a set of real scores obtained by existent individuals, atypical though they may be; whereas the multivariate method, by its very nature, deals with a fictional or "virtual" score distribution whose elements were generated computationally by the investigator.

As I said above, it is not clear what exactly is the relationship between the three aspects of the problem which I have christened "systematic un-matching," "unrepresentative subpopulations," and "causal-arrow ambiguity." But it seems to me that taken together, and combined with the problem (operating in the other direction) discussed by Kahneman, they force us to the conclusion that a large portion of current research in the

---

[7] This is a very illuminating article, the best I have seen on the subject, and includes a formal proof that *no* statistical manipulations performed on static data can resolve the causal-arrow-ambiguity problem.

behavioral sciences, while meeting the conventionally accepted standards of adequate design, must be viewed as methodologically unsound; and, more specifically, I suggest that the ex post facto design is in most instances so radically defective in its logical structure that it is in principle incapable of answering the kinds of theoretical questions which typically give rise to its use.

References

Anderson, A.R. (1951). A note on subjunctive and counterfactual conditionals. *Analysis*, *12*, 35-38.

Anderson, A.R. (1954). [Reviews of articles by Schneider, Diggs, Storer, Bergmann, and Brown and Watling.] *Journal of Symbolic Logic*, *19*, 68-71.

Anderson, J. (1952). Hypotheticals. *Australasian Journal of Philosophy*, *30*, 1-16.

Bartlett, M.S. (1936). A note on the analysis of covariance. *Journal of Agricultural Science*, *26*, 488-491.

Beardsley, E.L. (1949). 'Non-accidental' and counterfactual sentences. *Mind*, *46*, 573-591.

Bergmann, G. (1952). Comments on Storer's definition of 'soluble.' *Analysis*, *12*, 44-48.

Bloch, A.-M.A. (1969). Remembrance of feelings past: A study of Phenomenological Genetics. *Journal of Abnormal Psychology*, *74*, 340-347.

Braithwaite, R.B. (1953). *Scientific explanation*. Cambridge, UK: Cambridge University Press.

Broad, C.D. (1933). The "nature" of a continuant. In his, *Examination of McTaggart's Philosophy*, Vol. I (pp. 264-278). Cambridge, UK: Cambridge University Press.

Brown, R., & Watling, J. (1950–1952). Hypothetical statements and phenomenalism. *Synthese*, *8*, 355-366.

Brown, R., & Watling, J. (1952). Counterfactual conditionals. *Mind*, *61*, 222-233.

Burks, A.W. (1946). Laws of nature and reasonableness of regret. *Mind*, 55, 1-3.

Burks, A.W. (1951). The logic of causal propositions. *Mind*, *60*, 363-382.

Burks, A.W. (1955). Dispositonal statements. *Philosophy of Science*, *22*, 175-193.

Burks, A.W. & Copi, I.M. (1950). Lewis Carroll's Barber Shop Paradox. *Mind*, *59*, 219-222.

Burks, B., & Kelley, T.L. (1928). Statistical hazards in nature-nurture investigations. *Twenty-Seventh Yearbook of the National Society for the Study of Education, Nature and Nurture, Part I: Their Influence upon Intelligence* (pp. 9-38). Bloomington: University of Indiana Press.

Campbell, D.T. (1969). Reforms as experiments. *American Psychologist*, *24*, 409-429.

Campbell, D.T., & Stanley, J.C. (1963). Experimental and quasi-experimental designs for research on teaching. In N.L. Gage (Ed.), *Handbook of Research on Teaching* (pp. 171-246). Chicago, IL: Rand McNally.

Carnap, R. (1936–1937). Testability and meaning. *Philosophy of Science*, *3*, 420-471; *4*, 2-40. Re-printed with corrigenda and additional bibliography, New Haven, CT: Yale University Graduate Philosophy Club, 1950.

Carnap, R. (1966). *Philosophical foundations of physics*. New York: Basic Books.

Chapin, F.S. (1955). *Experimental designs in sociological research*. New York: Harper.

Chisholm, R. (1946). The contrary-to-fact conditional. *Mind*, 55, 289-307.

Chisholm, R. (1955). Law statements and counterfactual inference. *Analysis*, *15*, 97-105.

Cochran, W.G. (1957). Analysis of covariance: Its nature and uses. *Biometrics*, *13*, 261-281.

Cooley, J.C. (1957). Professor Goodman's 'Fact, Fiction, and Forecast.' *Journal of Philosophy*, *54*, 293-311.

Diggs, B.J. (1952). Counterfactual conditionals. *Mind*, *61*, 513-527.

Downing, P.B. (1959). Subjunctive conditionals, time order, and causation. *Proceedings of the Aristotelian Society*, *59*, 129-140.

Dunham, W.H. (1965). *Community and schizophrenia*. Detroit: Wayne State University Press.

Eckland, B.K. (1967). Genetics and sociology: A reconsideration. *American Sociological Review*, *32*, 173-194.

Freedman, D.G. (1958). Constitutional and environmental interactions in rearing of four breeds of dogs. *Science*, *127*, 585-586.

Fuller, J.L., & Thompson, W.R. (1960). *Behavior genetics*. New York: Wiley

Glass, D.C. (Ed.). (1968). *Biology and behavior*: Genetics. New York: Rockefeller University Press.

Goodman, N. (1947). The problem of counter-factual conditionals. *Journal of Philosophy*, *44*, 113-128.

Gottesman, I.I. (1963). Heritability of personality: A demonstration. *Psychological Monographs*, *77*(9, Whole No. 572).

Gottesman, I.I. (1968). Biogenetics of race and class. In M. Deutsch, I. Katz, and A. R. Jensen (Eds.), *Social class, race, and psychological development* (pp. 11–51). New York: Holt, Rinehart & Winston.

Greenwood, E. (1945). *Experimental sociology: A study in method*. New York: King's Crown Press.

Guttman, L. (1941). An outline of the statistical theory of prediction. In P. Horst, (Ed.), *The Prediction of personal adjustment*. Social Science Research Council Bulletin, *48*, 286-292.

Hampshire, S. (1948). Subjunctive Conditionals. *Analysis*, *9*, 9-14.

Hathaway, S.R., & Monachesi, E.D. (1963). *Adolescent personality and behavior: MMPI profiles of normal, delinquent, dropout, and other outcomes*. Minneapolis, MN: University of Minnesota Press.

Hempel, C.G. (1950). Problems and changes in the empiricist criterion of meaning. *Révue Internationale de Philosophie*, *4*, 41-63.

Hempel, C.G. (1966). *Philosophy of natural science*. Engle-wood Cliffs, NJ: Prentice-Hall.

Hempel, C.G., & Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy of Science*, *15*, 135-175. Reprinted in C.G. Hempel, *Aspects of scientific explanation*, New York: Free Press, 1965.

Hirsch, J. (1962). Individual differences in behavior and their genetic basis. In E. Bliss (Ed.), *Roots of behavior* (pp. 3-23). New York: Hafner.

Hirsch, J. (Ed.). (1967). *Behavior-genetic analysis*. New York: McGraw-Hill.

Hiż, H. (1949). On the inferential sense of contrary-to-fact conditionals, *Journal of Philosophy*, *48*, 586-587.

Hurst, M. (1935). Implication in the Fourth Century B.C. *Mind*, *44*, 484-495.

Johnson, W.E. (1921). *Logic, Part I*. Cambridge: Cambridge University Press.

Johnson, W.E. (1924). *Logic, Part III, The Logical Foundations of Science*. Cambridge: Cambridge University Press.

Jordan, P. (1955). *Science and the course of history*. New Haven, CT: Yale University Press.

Kahneman, D. (1965). Control of spurious association and the reliability of the controlled variable. *Psychological Bulletin*, *64*, 326-329.

Kempthorne, O. (1957). *An introduction to genetic statistics*. New York: Wiley.

Kneale, W. (1949). *Probability and induction*. Oxford: Clarendon Press. Pages 92-110 reprinted as "Induction, explanation, and transcendent hypotheses" in H. Feigl and M. Brodbeck (Eds.), *Readings in the philosophy of science* (pp. 353-367). New York: Appleton-Century-Crofts, 1953.

Kneale, W. (1950). Natural laws and contary-to-fact conditionals. *Analysis*, *10*, 121-125.

Kneale, W. (1961). Universality and necessity. *British Journal for the Philosophy of Science*, *12*, 89-102.

Lagerspetz, K. (1964). Studies on the aggressive behavior of mice, *Annales Academiae Scientarum Fennicae*, *131*, 1-131.

Langmuir, I, (1943). Science, common sense and decency. *Science*, *97*, 1-7.

Lesser, G.S., & Stodolsky, S. (1967). Learning patterns in the disadvantaged. *Harvard Educational Review*, *37*, 546-593.

Lesser, G.S., Fifer, G., & Clark, D.H. (1965). Mental abilities of children from different social-class and cultural groups. *Monographs of the Society for Research in Child Development*, *30*, 1-115.

Lewis, C.I. (1946). *An analysis of knowledge and valuation*. LaSalle, IL: Open Court.

Lindzey, G., Winston, E., & Manosevitz, M. (1961). Social dominance in inbred mouse strains. *Nature*, *191*, 474-476.

Loevinger, J. (1943). On the proportional. contributions in nature and in nurture to differences in intelligence. *Psychological Bulletin*, *40*, 725-756.

London, I.D. (1946). Some consequences for history and psychology of Langmuir's concept of convergence and divergence of phenomena. *Psychological Review*, *53*, 170-188.

London, I.D. (1952). Quantum biology and psychology. *Journal of General Psychology*, *46*, 123-149.

Mackie, J.L. (1962). Counterfactuals and causal laws. In R.J. Butler (Ed.), *Analytical philosophy* (pp. 66-80). New York: Barnes and Noble.

Manosevitz, M., Lindzey, G., & Thiessen, D.D. (Eds.). (1969). *Behavioral genetics: Method and theory*. New York: Appleton-Century-Crofts.

Mates, B. (1949). Diodorean implication. *Philosophical Review*, *58*, 234-242.

McClearn, G.E. (1962). The inheritance of behavior. In L. Postman (Ed.), *Psychology in the making* (pp. 144-252). New York: Knopf.

McClearn, G.E., & Meredith, W. (1966). Behavioral genetics. In P.R. Farnsworth, O. McNemar, & Q. McNemar (Eds.), *Annual Review of Psychology*, *17*, 515-550.

McClelland, D.C. (1961). *The achieving society*. Princeton, NJ: Van Nostrand Reinhold.

McClelland, D.C., Atkinson, J.W., Clark, R.A., & Lowell, E.L. (1953). *The achievement motive* (New York: Appleton-Century-Crofts.

McGill, T.E., & Blight, W.C. (1963). Effects of genotype on the recovery of sex drive in the male mouse. *Journal of Comparative and Physiological Psychology*, *56*, 887-888.

Meehl, P.E. (1954a/1996). *Clinical versus statistical prediction: a theoretical analysis and a review of the evidence*. Minneapolis, MN: University of Minnesota Press. Reprinted with new Preface, 1996, by Jason Aronson, Northvale, NJ.

Meehl, P.E. (1962) Schizotaxia, schizotypy, schizophrenia. *American Psychologist*, *17*, 827-838.

Meehl, P.E. (1964). *Manual for use with checklist of schizotypic signs* (Report No. PR-73-5). Minneapolis, MN: University of Minnesota, Research Laboratories of the Department of Psychiatry.

Molnar, G. (1969). Kneale's argument revisited. *Philosophical Review*, *78*, 79-89.

Nagel, E. (1961). *The structure of science*. New York: Harcourt, Brace & World.

Nerlich, G.C., & Suchting, W.A. (1967). Popper on Law and Natural Necessity. *British Journal for the Philosophy of Science*, *18*, 233-235.

O'Connor, D.J. (1951). The analysis of conditional sentences. *Mind*, *60*, 351-362.

Ostle, B. (1963). *Statistics in research* (2nd ed. Revised). Ames, IA: Iowa State University Press.

Pap, A. (1958a). Disposition concepts and extensional logic. In H. Feigl, M. Scriven, & G. Maxwell (Eds.), *Minnesota studies in philosophy of science: Vol. 2. Concepts, theories, and the mind-body problem* (pp. 196-224) Minneapolis, MN: University of Minnesota Press.

Pap, A. (1962). *An introduction to the philosophy of science*. New York: Free Press.

Pears, D. (1950). Hypotheticals. *Analysis*, *10*, 49-63.

Pirenne, M.H., & Marriott, F.H.C. (1959). The quantum theory of light and the psychophysiology of vision. In S. Koch (Ed.), *Psychology, a study of a science. Vol. 1: Sensory, perceptual, and physiological formulations* (pp. 288-361). New York: McGraw-Hill.

Platt, J.R. (1966). *The step to man*. New York: Wiley.

Popper, K.R. (1949). A note on natural laws and so-called 'contrary-to-fact conditionals.' *Mind*, *58*, 62-66.

Popper, K.R. (1959a). *The logic of scientific discovery*. New York: Basic Books. (Original work published 1934)

Popper, K.R. (1959b). On Subjunctive Conditionals with Impossible Antecedents. *Mind*, *68*, 518-520.

Popper, K.R. (1967). A revised definition of natural necessity. *British Journal for the Philosophy of Science*, *18*, 316-321.

Quine, W.V. (1959). *Methods of logic*. New York: Holt.

Rado, S. (1956). *Psychoanalysis of behavior*. New York: Grune & Stratton.

Rado, S. (1960). Theory and therapy: The theory of schizotypal organization and its application to the treatment of decompensated schizotypal behavior. In S.C. Scher & R.H. Davis (Eds.), *The outpatient treatment of schizophrenia* (pp. 87-101). New York: Grune & Stratton.

Rado, S., & Daniels, G. (1956). *Changing concepts of psychoanalytic medicine*. New York: Grune & Stratton.

Ratliff, F. (1962). Some interrelations among physics, physiology, and psychology in the study of vision. In S. Koch (Ed.), *Psychology, a study of a science. Vol. 4: Biologically oriented fields* (pp. 417-482). New York: McGraw-Hill.

Reichenbach, H. (1947). *Elements of symbolic logic*. New York: Free Press.

Rescher, N. (1961). Belief-contravening suppositions. *Philosophical Review*, *70*, 176-196.

Scarr, S. (1966). Genetic factors in activity motivation. *Child Development*, *37*, 663-673.

Scarr, S. (1968). Environmental bias in twin studies. *Eugenics Quarterly*, *15*, 34-40.

Scarr, S. (1969). Social introversion-extraversion as a heritable response. *Child Development*, *40*, 823-832.

Scheffé, H. (1959). *The analysis of variance*. New York: Wiley.

Schneider, E.F. (1953). Recent discussion of subjunctive conditionals. *Review of Metaphysics*, *6*, 623-649.

Sellars, W.S. (1948). Concepts as involving laws and inconceivable without them. *Philosophy of Science*, *15*, 287-315.

Sellars, W.S. (1958). Counterfactuals, dispositions, and the causal modalities. In H. Feigl, M. Scriven, & G. Maxwell (Eds.), *Minnesota studies in philosophy of science: Vol. 2. Concepts, theories, and the mind-body problem* (pp. 225-308). Minneapolis, MN: University of Minnesota Press.

Shields, J. (1962). *Monozygotic twins brought up apart and brought up together*. London: Oxford University Press.

Simon, H.A., & Rescher, N. (1966). Cause and counterfactual. *Philosophy of Science*, *33*, 323-340.

Slater, E., & Shields, J. (1969). Genetical aspects of anxiety. In M.H. Loder, (Ed.), *Studies in anxiety* (pp. 62-71), Special Publication no. 3, *British Journal of Psychiatry*, Ashford, Kent.

Storer, T. (1951). On defining 'soluble.' *Analysis*, *11*, 134-137.

Walters, R.W. (1961). The problem of counterfactuals. *Australasian Journal of Philosophy*, *39*, 30-46.

Watling, J.L. (1953). Propositions asserting causal connection. *Analysis*, *14*, 31-37.

Watling, J.L. (1957). The problem of contrary-to-fact conditionals. *Analysis*, *17*, 73-80.

Weinberg, J.R. (1951). Contrary-to-fact conditionals, *Journal of Philosophy*, *48*, 517-528.

Will, F.L. (1947). The contrary-to-fact conditional. *Mind*, *56*, 236-249.