Cattell R B 1944 'Parallel proportional profiles' and other principles for determining the choice of factors by rotation. *Psychometrika* **9**: 267–83
Cattell R B 1952 *Factor Analysis*. Harper, New York
Cattell R B 1954 *Culture Fair Intelligence Tests, Scales 1, 2, and 3, Forms A and B, rev. edn*. IPAT, Champaign, IL
Cattell R B 1957 *Personality and Motivation Structure and Measurement*. World Book Co., New York
Cattell R B 1964 Validity and reliability: A proposed more basic set of concepts. *Journal of Educational Psychology* **55**: 1–22
Cattell R B 1971 *Abilities: Their Structure, Growth, and Action*. Houghton Mifflin, Boston
Cattell R B 1972 Real base, true zero factor analysis. *Multivariate Behavioral Research Monographs* **72**(1): 1–162
Cattell R B 1984 The voyage of a laboratory, 1928–1984. *Multivariate Behavioral Research* **19**: 121–74
Child D 1998 Raymond Bernard Cattell (1905–1998). *British Journal of Mathematical and Statistical Psychology* **51**: 353–7
Goldberg L R 1968 Objective Personality and Motivation Tests—theoretical introduction and practical compendium—Cattel R B and Warburton F W. *Contemporary Psychology* **13**: 617–9
Hall C S, Lindzey G 1978 *Theories of Personality*, 3rd edn. Wiley, New York
Horn J L, Cattell R B 1966 Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology* **57**: 253–70
McArdle J J 1984 On the madness in his method: R. B. Cattell's contributions to structural equation modeling. *Multivariate Behavioral Research* **19**: 246–67
McArdle J J, Cattell R B 1994 Structural equation models of factorial invariance in parallel proportional profiles and oblique confactor problems. *Multivariate Behavioral Research* **29**(1): 63–113
Nesselroade J R, Cattell R B (eds.) 1988 *Handbook of Multivariate Experimental Psychology*, 2nd edn. Plenum, New York

J. R. Nesselroade

Causal Counterfactuals in Social Science Research

The term 'counterfactual conditional' is used in logical analyses to refer to any expression of the general form: 'If A were the case then B would be the case.' In this usage, A is usually false or untrue in the world so that A is 'contrary to fact' or counterfactual. Examples abound. 'If kangaroos had no tails, they would topple over' (Lewis 1973b). 'If an hour ago I had taken two aspirins instead of just a glass of water, my headache would now be gone' (Rubin 1978). Perhaps the most obnoxious counterfactuals in any language are those of the form: 'If I were you, I would … .'

Lewis (1973a) observed the connection between counterfactual conditionals and references to causation. He finds these logical constructions in the language used by Hume in his famous discussion of causation. Hume defined causation twice over. He wrote '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*' (Lewis 1973a, italics are Lewis's).

Lewis draws attention to the comparison between the factual first definition where one object *is* followed by another and the counterfactual second definition where, counterfactually, it is supposed that if the first object 'had not been' the second object would not have been either.

It is the connection between counterfactuals and causation that makes them relevant to social science research. From the point of view of some authors, it is difficult, if not impossible, to make any sense of causal statements without using counterfactual language (Lewis 1973a, Holland 1986, Rubin 1978, Robins 1985, 1986). Other authors are concerned that using such language gives an emphasis to unobservable entities that is inappropriate in the analysis of empirical data (Dawid 1997, Shafer 1996). The discussion here accepts counterfactuals in discussions of causation and will explain their role in the estimation of causal effects based on the work of Neyman (1923, 1935), Rubin (1974, 1978), and others.

We begin with a simple observation. Suppose that we find that a student's test performance changes from a score of X to a score of Y after some educational intervention. We might then be tempted to attribute the pretest-posttest change, $Y-X$ to the intervening educational experience, that is, to use the gain score as a measure of the improvement due to the intervention. However, this is social science and not the tightly controlled 'before-after' measurements made in a physics laboratory. There are many other possible explanations of the gain, $Y-X$. Some of the more obvious are: simple maturation, other educational experiences occurring during the relevant time period, and differences in either the tests or the testing conditions at pre- and posttests. Cook and Campbell (1979) provide a classic list of 'threats to internal validity' that address many of the types of alternative explanations for apparent causal effects of interventions. For this reason, it is important to think about the real meaning of the attribution of cause. In this regard, Lewis's discussion of Hume serves us well. From it we see that what is important is what the value of Y *would have been* had the student *not had* the educational experiences that the intervention entailed. Call this score value, Y^*. Thus enter counterfactuals. Y^* is not directly observed for the student, that is, they *did have* the educational intervention of interest, so asking for what their posttest score *would have been* had they *not had it* is asking for information collected under conditions that are contrary to fact. Hence, it is not the difference $Y-X$ that is of interest, but the difference $Y-Y^*$, and the gain score has causal significance relative to the effect of the educational experience only if X can serve as a substitute for the

counterfactual Y^*. In physical-science laboratory experiments such a substitution is often easy to make, but it is rarely believable in many social science applications of any consequence.

A formal model or language for discussing the problem of estimating causal effects (of the form $Y - Y^*$ rather than $Y - X$) was developed by Neyman (for randomized experiments) and Rubin (for a wide variety of causal studies) and will be called the Neyman/Rubin model for causal effects here.

1. Prospective Causal Studies

The Neyman/Rubin model is most easily understood in the context of a *prospective causal study* that has the general structure specified by this sequence of events.

(a) Subjects or experimental units of study are identified.

(b) Baseline or pretest information about these units is recorded.

(c) The units are either assigned to (in a controlled study) or select themselves to (in studies without the control of assignment) exposure to one of the treatment conditions or interventions of the study.

(d) These units are then subsequently exposed to their assigned or self-selected treatment condition (and each unit is affected by this exposure in a manner that is unrelated to the exposure conditions of the other units).

(e) At an appropriate later time an outcome, endpoint, or posttest measure is recorded for each unit in the study.

The type of study that this five-part schema is intended to cover includes most randomized comparative experiments as well as many types of pretest/posttest quasiexperiments or observational studies. It should be emphasized that, properly interpreted, the Neyman/Rubin model has application to other types of causal studies (see Holland 1988, Holland and Rubin 1988, Robins 1997), but, for our purposes here, prospective causal studies are already sufficiently complicated and inclusive.

The condition mentioned parenthetically in (d), that the exposure conditions of the other units do not affect the outcomes associated with a given unit, is very important, and is clearly an assumption that would not be true in general. For example, in the study of infectious diseases your vaccination will affect my likelihood of contracting polio. Rubin explicitly identifies this assumption calling it the Stable Unit-Treatment Value Assumption, or SUTVA. SUTVA will be assumed throughout this discussion.

2. The Neyman/Rubin Model

The version of the Neyman/Rubin model used here is adapted from that of Holland (1986), and differs from the original versions by Neyman (1935) and Rubin (1978) mainly in its emphasis on population quantities. One of the main benefits of this model is that it identifies certain 'counterfactual conditional expectations' as the location of key assumptions about the inferential structure of any causal study and its resulting data.

The prospective causal study begins with the 'units,' 'subjects,' or 'cases' of the study, and the ith unit is denoted by the subscript 'i.' It may help the reader to imagine that this is a discussion about a very large sample of units. Denote the population of units under study by P. For the most part P will lie quietly in the background without being noticed. The baseline information that is collected or recorded for unit i will be denoted as a *vector* of numerical information, z_i.

There is a 'causal' variable denoting a set of possible 'treatments' or 'exposure' conditions to which each unit in the study could be exposed. For simplicity we assume that these are only two treatment conditions denoted by $x = 1$ (treatment) and $x = 0$ (control). A more complicated version of this would let x be a number representing the 'strength' of the treatment level, but we will just use the dichotomous case here.

An important aspect of causal variables designating such treatments levels or intervention conditions is the assumption that the level of exposure for any unit *could have been different* from what it actually was. This condition excludes 'attributes' of units (such as race, gender, age, or pretest score) as causal variables in the sense that such attributes *cannot* have 'unit-level causal effects' in the sense that we will define below. This idea is discussed more extensively in Holland (1986, 1988), and is mentioned again in another context at the end of the present discussion. Each unit is exposed to one treatment level and the value of x to which i is exposed is denoted by x_i.

Finally we come to the outcomes or dependent variables in the study, and here is where a special notation is needed. We let $Y_i(x)$ denote the (numerical) response that would be recorded for unit i if unit i were exposed to treatment level, x. For each i, $Y_i(x)$ is a function of x. It should be emphasized that $Y_i(x)$ is not directly observed unless $x_i = x$. This is an important point because it is crucial to realize that the $\{Y_i(x)\}$ do not denote *observed data* like z_i and x_i do, but rather the $\{Y_i(x)\}$ are 'potential outcomes' that lie behind the observed values of the outcome variable. For this reason, we denote the potential observations by capital letters to distinguish them from quantities that are directly observable, which we denote by lower case letters.

The connection between the potential outcomes and the actually observed outcomes is then given by the equation

$$y_i = Y_i(x_i) \tag{1}$$

where y_i is the observed outcome or value of the dependent variable for unit i. The idea behind Equation (1) is that to get from the potential outcomes

$\{Y_i(x)\}$ to an observed outcome we must select the value of x in $Y_i(x)$ to be the value to which i is actually exposed, that is, $x = x_i$, and then we obtain y_i from the potential observations, $\{Y_i(x)\}$, via Equation (1).

The *observed data* for each unit i is the vector (z_i, x_i, y_i). The potential outcomes, $\{Y_i(x)\}$, are never observed for *all* values of x for a fixed unit, i, but only for the specific x-value to which i is actually exposed, x_i. It is sometimes said that the $\{Y_i(x)\}$ are 'counterfactual' because they are not actual observations. Here they are called *potential observations* because they could have been observed had x_i been different than it was.

It is helpful to use a notation such as, $E(y)$, to mean the average value of y_i across the (large number of) units in P. Furthermore, an expression such as

$$E(y \mid x = a) \tag{2}$$

will mean the average value of y_i across all of the (large number of) units in P for which $x_i = a$. The use of the expectation notation is to make certain quantities clearer in their meaning, and may be justified from either a frequentist or Bayesian point of view. It should be noted that within the expectation notation, the subscript i, denoting the unit, is suppressed because, within the scope of the $E(\ )$ operator, i is averaged over.

3. Using the Neyman/Rubin Model

An important fact about the Neyman/Rubin model is the Fundamental Problem of Causal Inference (Holland 1986), which is: It is impossible in principle to *observe* $Y_i(x)$ for more than one value of x for any one unit, i. Any procedure that claims to have avoided the Fundamental Problem of Causal Inference can always be shown to be based on untestable assumptions. Sometimes such assumptions are plausible, and sometimes they are not.

A basic definition that we are now in a position to make is that of a 'unit-level causal effect.' Because we are restricting attention to the simple case of two treatment levels, $x = 0/1$ we may restrict attention to these differences

$$\begin{aligned} Y_i(1) - Y_i(0) = &\text{the casual effect of } x = 1 \\ &\text{relative to } x = 0 \text{ for unit } i \end{aligned} \tag{3}$$

In the Neyman/Rubin notation, the unit-level causal effects are the *basic* quantities of interest in causal inference. However, the Fundamental Problem of Causal Inference is now immediately seen to *be fundamental* because it implies that unit-level causal effects are, themselves, *never directly observable*. Thus, we are always reduced to making assumptions that allow us to make some sort of conclusion or inference about these causal effects. This is the place where the Neyman/Rubin model might appear to be impractical for applied research, that is, because its most basic parameters, the unit-level causal effects, are not directly observable. Furthermore, this is exactly the place where the potential exposability of all of the levels of x to any unit is seen to be crucial to the foundations of the theory. The definition of causal effect requires this assumption so that the difference, $Y_i(1) - Y_i(0)$, is meaningful. It is the Fundamental Problem of Causal Inference and this definition of causal effect that makes causal inference both more interesting and more difficult than the simple computation of correlational and associational measures. The unit-level causal effects mimic the comparison between a 'factual' and a 'counterfactual' identified in the quote mentioned earlier from Hume by Lewis.

It is now time to introduce the notion of an Average Causal Effect (ACE). In general, an ACE is any average of unit-level causal effects. The most general ACE has the form

$$\text{ACE} = E[Y(1) - Y(0) \mid A] \tag{4}$$

where A denotes some collection of units defined in terms of either z_i or x_i or both. In Equation (4) we again suppress the subscript i because it is being averaged over. As an example of an A in Equation (4) we might use $A =$ 'all the units in the study,' in which case the ACE is the average causal effect over all of P. But other cases might be of interest, for example, $A =$ 'all units where i is male and for whom $x_i = 1$.' In this case the ACE is for the males in treatment group 1. Here we restrict our attention to the ACE that is called the 'effect of the treatment on the treated' in which A denotes all of the units for which $x_i = 1$, that is

$$\begin{aligned} \text{ACE} &= E[Y(1) - Y(0) \mid x = 1] \\ &= E[Y(1) \mid x = 1] - E[Y(0) \mid x = 1] \end{aligned} \tag{5}$$

Up to now, we have simply defined the basic structure of the data collection design as well as the causal connection between the potential outcomes and the causal variable x, that is, Equation (3). We have not yet identified the connection between any quantities that could be estimated with data and the causal parameters given by either the unit-level causal effects in Equation (3) or the average causal effects in Equation (5). This leads us to the '*prima facie* Average Causal Effects,' (FACEs). The FACEs are what can be estimated from the data. To parallel the ACE in Equation (5) we examine the FACE which is simply the difference between the mean of the outcome variable observed in each treatment group, that is

$$\text{FACE} = E[y \mid x = 1] - E[y \mid x = 0] \tag{6}$$

If we substitute the definition of y in terms of the potential observations, $\{Y_i(x)\}$, that is given in Equation (1) into Equation (6) we obtain

$$\text{FACE} = E[Y(1)\,|\,x = 1] - E[Y(0)\,|\,x = 0] \quad (7)$$

Finally if we combine Equation (5) and Equation (7) we obtain the following basic formula that relates the ACE to the FACE

$$\text{FACE} = \text{ACE} + \text{BIAS} \quad (8)$$

where

$$\text{BIAS} = E[Y(0)\,|\,x = 1] - E[Y(0)\,|\,x = 0] \quad (9)$$

The BIAS term contains two parts, one *factual*, that is, $E[Y(0)\,|\,x = 0] = E[y\,|\,x = 0]$, and the other *counterfactual*, that is, $E[Y(0)\,|\,x = 1]$. The factual part is just the mean of y_i for those units with $x_i = 0$. The counterfactual part is the mean of $Y_i(0)$ for those units for whom $x_i = 1$. $E[Y(0)\,|\,x = 1]$ is a quantity for which there can never be any data because the conditioning event makes the quantity being averaged over a counterfactual. Thus, it is a counterfactual conditional expectation. The value of such counterfactual parameters is that they pinpoint exactly where assumptions must be made that allows causal inference to take place using empirical data. When BIAS = 0, we have FACE = ACE and the empirical FACE equals the causal ACE.

An important condition that insures that BIAS = 0 is the construction of x_i by random assignment which forces x_i and $Y_i(0)$ to be statistically independent of each other as functions of i over P (Holland 1986). When this independence holds, $E[Y(0)\,|\,x = 1] = E[Y(0)\,|\,x = 0]$ and BIAS = 0.

4. Empty Counterfactuals

There is an unsatisfactory and rather misleading use of counterfactuals that sometimes arises in social science research. It occurs when the counterfactual condition, that is, the 'if A were the case' part could never occur in any real sense. Such empty counterfactuals arise when a nonmanipulable factor in a causal study is described as having an 'effect' on some outcome. Examples easily come about in casual causal talk. The effect of gender on salary suggests considering 'what her salary would have been had she been a man.' The effect of test performance on future employment suggests 'the job he would have had had he scored higher on the test.' The effect of English language proficiency on a math test in English suggests 'the mathematics score a non-English speaker would have received had he or she been an English speaker.' These empty counterfactuals arise when the value of a variable for a factor in a study could not have been other than the value that it was. The interesting and useful counterfactuals arise in those cases when the variable could have had a different value that it did for the individuals in a study, at least in principle. Judgment as to when a counterfactual is empty or not is not always easy and may require careful thought in many cases. Consider the examples 'if I were you, I would, …' Lewis's kangaroo and Rubin's aspirin in the opening paragraph. These represent very different kinds of counterfactuals from this perspective. The first is as empty as they get, the emptiness of the second depends on how the kangaroo might not have a tail (our imagination vs. an axe), while Rubin's aspirin could be taken or not.

See also: Causation (Theories and Models): Conceptions in the Social Sciences; Counterfactual Reasoning: Public Policy Aspects; Counterfactual Reasoning, Qualitative: Philosophical Aspects; Counterfactual Reasoning, Quantitative: Philosophical Aspects; Internal Validity; Quasi-Experimental Designs

Bibliography

Cook T D, Campbell D T 1979 *Quasi-experimentation: Design and Analysis Issues for Field Settings*. Houghton Mifflin, Boston

Dawid A P 1997 Causal Inference without Counterfactuals. Research Report No. 188, Department of Statistical Science, University College, London

Holland P W 1986 Statistics and causal inference. *Journal of the American Statistical Association* **81**: 945–70

Holland P W 1988 Causal inference, path analysis and recursive structural equations models. In: Clogg C (ed.) *Sociological Methodology*. American Sociological Association, Washington, DC, pp. 449–84

Holland P W, Rubin D B 1988 Causal inference in retrospective studies. *Evaluation Review* **12**: 203–31

Lewis D K 1973a Causation. *Journal of Philosophy* **70**: 556–67

Lewis D K 1973b *Counterfactuals*. Harvard University Press, Cambridge, MA

Neyman J 1923 Sur les applications de la theorie des probabilites aux experiences agricoles: Essai des principes. *Roczniki Nauk Rolniczki* **10**: 1–51 (in Polish: English trans. Dabrowska D, Speed T 1991 *Statistical Science* **5**: 463–80)

Neyman J 1935 Statistical problems in agricultural experimentation. *Supplement of the Journal of the Royal Statistical Society* **2**: 107–80

Robins J M 1985 A new theory of causality in observational survival studies—Application of the healthy worker effect. *Biometrics* **41**: 311

Robins J M 1986 A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical Modelling* **7**: 1393–1512

Robins J M 1997 Causal inference from complex longitudinal data. In: Berkane M (ed.) *Latent Variable Modeling with Applications to Causality*. Springer-Verlag, New York, pp. 69–117

Rubin D B 1974 Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* **66**: 688–701

Rubin D B 1978 Bayesian inference for casual effects: The role of randomization. *Annals of Statistics* **6**: 34–58

Shafer G 1996 *The Art of Causal Conjecture*. MIT Press, Cambridge, MA

P. W. Holland

Causal Inference and Statistical Fallacies

1. Generalities

The pairing of causality and fallacies may seem idiosyncratic. In fact it nicely captures the point that many statistical fallacies, i.e., plausible-seeming arguments that give the wrong conclusion, hinge on the overinterpretation or misinterpretation of statistical associations as implying more than they properly do. The article begins by discussing three main views of causality, briefly indicating the scope for fallacious arguments and then at the end returns to discuss some fallacies in slightly more detail. See *Graphical Models: Overview*.

The very long history in the philosophical literature of discussions of causality is largely irrelevant for these purposes. It typically regards a cause as necessary and sufficient for an effect: all smokers get lung cancer, all lung cancer patients smoke. Here the concern is with situations with multiple causes, even if one is predominant, and where explicit or implicit statistical or probabilistic considerations are needed.

2. Notions of Causality

2.1 Causality as Stable Association

Suppose that a study, or preferably several different but related studies, shows that two features, C and R, of the individuals (people, firms, communities, households, etc.) under investigation are associated. That is, if we take, to be explicit, positive monotone association, individuals with high values of C tend to have high values of R and vice versa. For example C and R might be test scores at a given age in arithmetic and language, or level of crime and unemployment rate in a community.

Under what circumstances might one reasonably conclude that C is a cause of a response R, or at least make some step in the direction of that conclusion? And what would such a statement of causality mean?

2.1.1 Symmetric and directed relations. Association is a symmetric relation between two, or possibly more, features. Causality is not symmetric. That is, if C is associated with R then R is associated with C, but if C is a cause of R then R is not a cause of C. Thus the first task, given any two features C and R, is to distinguish the cases where:

(a) C and R are to be regarded as in some sense on an equal footing and treated in a conceptually symmetric way in any interpretation.

(b) One of the variables, say C, is to be regarded as explanatory to the other variable, R, regarded as a response. That is, if there is a relation, it is regarded asymmetrically.

Often significance tests for the existence of association and of dependency are identical. The distinction being studied here is a substantive one of interpretation. Failure to observe this distinction leads to the fallacy of the overinterpreted association.

2.1.2 Graphical representation. A useful graphical representation shows two variables X_1 and X_2, regarded on an equal footing, if associated, as connected by an undirected edge, whereas two variables such that C is explanatory to R, if connected, are done so by a directed edge. See Fig. 1a and Fig. 1b.

There are two possible bases for the distinction between explanatory and response variables. One is that features referring to an earlier time point are explanatory to features referring to a later time point. The second is a subject-matter working hypothesis based for example on theory or on empirical data from other kinds of investigation. Thus the weight of a child at one year is a response to maternal smoking behavior during pregnancy. In such situations the relevant time is not the time when the observation is made but the time to which the features refer, although of course observations recorded retrospectively are especially subject to recall biases.

As an example of the second type of explanatory-response relation, suppose that data are collected on diabetic patients assessing their knowledge of the disease and of their success in managing their disease, as measured by glucose control. These data may well refer to the same time point and it is not inconceivable that, for example, patients with poor glucose control are thereby encouraged to learn more about their disease. Nevertheless, as a working hypothesis, one might interpret the data assuming that knowledge, C, is explanatory to glucose control, R, considered as a response. This is represented in simple graphical form in Fig. 1b by the directed edge from C to R.

International Encyclopedia of the Social & Behavioral Sciences ISBN: 0-08-043076-7