Influence of Philosophical Concepts of Causality on Causal Modelling in Statistical Research*

Ursula M. Sondhauss
Collaborative Research Centre
"Reduction of Complexity for Multivariate Data Structures"
Department of Statistics
University of Dortmund

Abstract

In this paper causality is seen from a pluralist point of view: About its physical reality one can not claim to know anything as it is seen to be not a matter of fact but rather a matter of human perception. Different philosophical concepts of causality are presented and their philosophical impact and their limitations are sketched. They are shown to underlie different statistical causal modelling. For Pearl's (1995) causal graphical models, the decision theoretic causal graphical models of Heckerman and Shachter (1995) and the potential-response model of Rubin (1974) it is indicated how the advantages and disadvantages of their underlying causal concepts influence their applicability in certain scientific contexts and their acceptancy by individual researchers.

Keywords: causality, philosophy, causal statistical modelling, causal graphical models, potetial-response model

Introduction

If we try to define causality as some physical entity for which we have sufficient reason to claim its truth, we end in the so called Münchhausen-Trilemma (Albert, 1968): Either we (1) continue the argumentation in some infinite regress or we (2) produce a cycle and hope that no-one will realise it or we (3) call some primitives self-evident so that any reasonable person simply <u>must</u> agree with us. This situation is well known. Most researchers agree that causality is some helpful concept with

---

* Presented at the Workshop "Causal Networks: from Inference to data mining" CaNew'98, held in conjunction with the biennial Iberoamerican Conference on Artificial Intelligence IBERAMIA'98

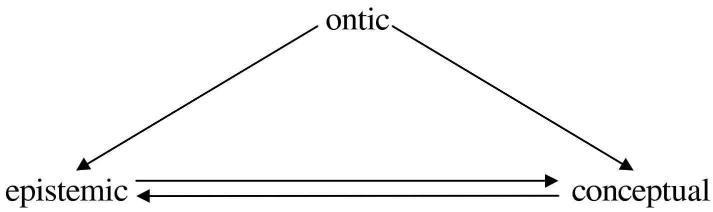
which humans organise their surroundings, but its physical "truth" is not decidable. This could lead to a very open and pragmatic dealing with causal models. Yet to me it seems, when the use of causal models in empirical research is discussed the different believes about the "truth" of causality play an important role - the more important and negative the less explicit people are about their believes. Implicit believes can become implicit assumptions which are like hidden prejudices in social life: they have a lot of influence but are hard to fight since they are not open to criticism. Not believing in the possibility of prejudice-free science I see the plurality of believes as appropriate means to an objective science. Important to that end is consciousness about the epistemological level of different scientific questions and clarity of one's own believes.

In this paper I will discuss the epistemological level of scientific questions concerning causality using a three-level model of cognition that I will introduce in the second section of this paper. In the third section I will present different plausible concepts of causality and in the forth section three causal models in statistical science will be reviewed in the light of these concepts: the causal graphical models of Pearl (1995) and Heckerman and Shachter (1995) and the potential-response model of Rubin (1974). Some conclusions will be drawn in the last section.

Three Levels of Cognition

The presentation of this model follows Koch (1994). A basic assumption of the model is the existence of a reality independent of our perception. It is called ontic level of cognition or ontic world. Ontic reality is seen to produce the experiences that are accessible to our senses (including their expansion by equipment). These experiences constitute the epistemic level of cognition. The conceptual level finally is used to mediate between epistemic and ontic reality: the sensual experiences get structured to build concepts about the ontic reality.

The interrelation between the three levels can be graphically represented as follows:



The individual arrows mean:

ontic→epistemic: the part of the ontic world that is accessible to our senses is observed and understood as interaction between realising subject and ontic world.

ontic→ conceptual: this can be understood as representing the human analogue to computer hardware that consists of the ontogenetic and phylogenetic developed characteristics of human cognition that influence and limit the human urge and capability to organise sensual experiences. This gives an interpretation of Kant's (1724-1804) opinion that causality was a form that human reason prescribed all experiences and thus can never been proven by experience (Kant, 1922).

conceptual → epistemic: ideas of the world influence the perception of it.

epistemic → conceptual: concepts are build on the basis of sensually experienced material.

Mackie (1985) gave an example of typical causal questions within each of these three levels. Questions on the ontic level ask: *'What is causation in reality, in the objective world: what actually goes on in what we take as typical cases of causation?'*. On the conceptual level we ask: *'What is our present established concept of causation, of what causes and effects are, and of the nature of the relation between them?'* On the epistemic level we want to know: *'What is causation in reality so far as we know it? What can we observe or discover or establish or reasonably believe about what we take as causes and effects?' (Mackie, 1985, S.178).*

Answers to causal questions on the ontic level I see as a matter of decision of the individual person. Conceptual answers should be investigated with respect to their inner consistence and their connection to other related ontic question as for example determinism and indeterminism. Any causal concept should correspond to certain ways of enlarging our epistemic knowledge and thus should correspond to certain epistemic observations.

Some concepts of causality

The foundation for the modern view of causality is the regularity theory of Hume (1711-1776). According to Hume (1995) there are two main characteristics that let people think two events might be causally linked: succession in time and contiguity[2]. These components alone do not succeed to capture the human notion of causality. Hume stated that humans have the idea of some necessary connection between cause and effect. Knowing this necessary connection would allow us to draw some a-priori inference from only the causes to their effects. Because of the problems of induction this necessary connection can never be proven to exist and stays fictious. In real life it is replaced by the observation of a constant conjunction between the cause and the effect. A causal statement of singular events is considered to be true because of some generic causal statement that is inferred from the observation of some or many "similar" singular events. Thus causality is a matter of human perception and not of fact.

From the standpoint of a logical empiricist science should only be about epistemic phenomena and not about some fictious entities. Consequently Russell (1872-1970) stated that causality should not be used in scientific concepts at all (Russell, 1913). The statistician and logical empiricist K. Pearson (1857-1936) wanted to replace the concept of causality by correlation and contingency tables as he regarded nature as symmetric (Pearson, 1911). This strict rejection of causality by logical empiricists was given up for example by the physicist Frank (1884-1966) who also did not

[2] Both these conditions are not unambiguous. Time succession is a questionable part of a definition of causality mainly because time itself is defined by causality (second trap of the Münchhausen-Trilemma). Contiguity is dubious mainly because it can never be proven by empirical verification. I suppose these difficulties to be known and I will not go into details.

accept causality as a concept of some fundamental law of reality but - like Hume - as some concept to structure our surroundings (Frank, 1988).

A definition of causality that is in accordance with this perspective is the strict interventionistic concept: It does not explain the human understanding of causality by the fictious necessary connection but on self-reflection of actions in which we experience ourselves as originators of changes. To say "*p* is a cause of *q*" then means that by doing *p* one can bring about *q* (von Wright, 1971). The strict interventionistic definition of cause is criticised because it is antropomorphistic and because it does not capture the use of the word in daily life. If the concept shall be expanded to cover events that can not really be manipulated then one needs hypothetical manipulations via imagination and analogy. This makes the definition vulnerable to other critisism: the argumentation is circular because one needs causal knowledge for the analogy to establish causal knowledge.

A further development of the interventionistic concept is what I call here the human motivation concept (Ströker, 1992). In this concept causality is not a characteristic of the phenomena but of the person who observes the phenomena. There are two main motivations: Enlarging our epistemic knowledge by searching for strategies to reach a goal and enlarging our conceptual understanding by searching for explanations for phenomena that are not covered by any existing concept.

The so-called counterfactual concept of causality replaces the necessary connection not by constant conjunction or possible manipulation but by some counterfactual statement. Actually Hume himself has brought forward the counterfactual argument directly following his well known definition of cause:

> *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. <u>Or in other words where, if the first had not been, the second never had existed</u>. (Hume, 1995)*

To overcome the problem that the truth of counterfactual statements cannot be decided by classical logic Lewis (1973) invented the idea of possible worlds: Two observed events *c* and *e* are causally dependent if a possible world exists in that *not-c* and *not-e* are true and if this possible world is closest to the real world among all possible worlds in which *not-c* is true. On this basis one decides that if *c* had not been, *e* had not existed.

Lewis is very cautious in saying how the closest world can be found - especially how the distance of world may be measured. This is not a necessary part of the axiomatic of his logic in general but it gains importance when one wants to use it for causal analysis. Actually - hidden in the decision about the closest world the generic causal statements still play an important role to decide on the truth of singular causal statements. Any aspects of causality like contiguity, time sucession, possibility of manipulation etc. can be used and at least some must be used to construct closest worlds (e. g. Mackie, 1985).

When the counterfactual causal analysis is used for generic statements about classes of "similar" events *C* that are causes of events in class *E* if *E* follows *C* and *Not-E* follows *Not-C* then this means that *C* is a necessary and sufficient condition for *E*. This is consistent with an pre- and post-deterministic worldview: Similar causes lead to the same effects and similar effects have similar causes. Many people agree with the first statement but would rather think that similar effects might have different causes thus denying post-determinism. The non-post-deterministic variant of the counterfactual definition of causality is the so-called INUS-condition for causes of Mackie (1974). According to this condition some *C* is said to be a cause of *E* if *"it is an* insufficient *but* non-redundant *part of an* unnecessary *but* sufficient *condition..."* for *E*. The counterfactual statement is expressed by the requirement of non-redundancy: if not some other complex condition caused *e* then *c* was necessary for e.

All the preceding concepts have to deal with epistemic indeterminism: In our experiences there is (almost sure) no perfect constant conjunction. To explain this from a pre-deterministic point of view we can invoke the so-called ceteris-paribus-clause: any causal statement is valid with the added phrase "other things being equal". If *c* is not followed by *e* though we say *C* was a cause of *E* then this is due to some things having been different from those singular events where the generic rule is true.

With that explanation the probabilistic definition of causality by Suppes (1970) can be used from a deterministic point of view. But Suppes' concept is also open to an indeterministic worldview as causes do not have to determine their effects but only need to increase their probability. But by this one can not avoid the necessity to take into account other circumstances: the increase must occur whatever the other circumstances are. This can be seen to be some inversed ceteris-paribus-clause. And to differentiate between a cause and its effect the probabilistic definition of cause and effect can not do without time-succession.

Pearl's interventionistic concept

For Pearl (1997) the shaded existence of causality in statistics has two main reasons: the logical empiricists rejection of the concept as such e.g. by the famous statistician Pearson plus the lack of a mathematical formalism to distinguish causal from equational relationships. Pearl's aim is to supply the language of probability with a symbolic representation of causal statements so that causal reasoning can be done in an open and formal way in the statistical research.

The core of the causal graphical model of Pearl is a finite set of variables $\{X_1,..., X_n\}$ that represent some entities of interest. The assumptions about their causal relationships are encoded in a directed acyclic graph in the well-known way (e. g. Pearl, 1988). The conceptual understanding of causal relationships among those variables is a probabilistic version of the expanded interventionistic definition: The causal effect of a variable X on a variable Y is defined as the influence on the probability for $Y=y$ when by an atomic external intervention one fixes $X=x$. Formally it is defined as the function from the possible values $x \in \mathrm{Val}(X)$ to the space of the probability distributions of Y when X is forced to have a certain value $x \in \mathrm{Val}(X)$. The probability distribution of Y when $X=x$ is set, is denoted by $P_{Y|\check{x}}$. This is in contrast to the probability distribution $P_{Y|x}$ of Y when X was observed to have a certain value $x \in \mathrm{Val}(X)$. The calculation of $P_{Y|\check{x}}$ is possible through an underlying pre-deterministic model of the causal relationships among the variables: Each child-parent family $\{X_i, pa_i\}$, $i=1,...,n,$ in the graph represents a deterministic function $X_i=\mathrm{f}_i(pa_i, \varepsilon_i)$, $i=1,...,n,$ where the ε_i, $i=1,...,n,$ are mutually independent, arbitrarily distributed random disturbances, representing the explanation for factual indeterminism between X_i and its direct causes pa_i, $i=1,...,n$. Then $P_{X_j|\check{x}_i}$, $i,j \in \{1,...,n\}$, $i \neq j$, can calculated in a submodel where the mechanism $X_i=\mathrm{f}_i(pa_i, \varepsilon_i)$ is replaced by $X_i=x_i$ and in all other equations $X_k=\mathrm{f}_k(pa_k, \varepsilon_k)$, $k=1,...,n$, $k \neq i$, X_i is replaced by x_i.

Pearl's model is widely applicable and uses highly efficient algorithms to guide causal analyses. Problems occur for users of the causal model of Pearl when there are variables where even hypothetically it is hard to imagine any possible atomic intervention. Modeling them as so-called latent variables would mean to ignore the information one has about their observed values - taking them in as "core" variables might arise doubts in the correctness of the model.

From an expanded interventionists point of view this problem can (almost) anytime be circumvented. If the entity itself can not be hypothetically manipulated then may be what we really want to model is not this entity but something else. As an example Pearls (1995, rejoinder) says gender in a study about gender discrimination might not be the entity of interest but rather the perception of the gender of a person by others.

Potential-Response Model

The two main developers of the potential-response model Rubin (1974) and Holland (1986) combine a strict interventionistic conceptual view of the "cause of an effect" with a counterfactual way to establish "effects of causes" on the epistemic level. The definition of cause is restricted to an experimental milieu - with a universe $\boldsymbol{U}$ or population of "units" $u \in \boldsymbol{U}$ as basic objects of study and a finite set of variables $\{X, Y, ...\}$ which are real-valued functions that are defined for each unit in $\boldsymbol{U}$. Causes in this model are equivalent to a set $\boldsymbol{B}$ of at least two different treatments (e. g. $\boldsymbol{B} = \{b,b'\}$) that the experimenter assigns - not necessarily at random - to the units. Only by this those objects can be seen as causes that are potentially exposable to the units in an experiment and all other characteristics of units are mere attributes. Any causal effect is defined relative to the effect of another treatment: If the Variable Y is chosen as the variable of interest, then $Y_b(u) - Y_{b'}(u)$ is the effect of b in comparison to the effect of b'. This definition is counterfactual: It is impossible to give both treatments to a unit at the same time - thus one of the statements either $Y_b(u)$ or $Y_{b'}(u)$ is necessarily counterfactual. This is called by Holland (1986) the "fundamental problem of causal inference".

To handle the problem one needs some additional assumptions: For example, when the units are said to be "similar" or homogeneous then $Y_b(u) - Y_{b'}(u) = Y_b(u) - Y_{b'}(u'), u' \neq u, u' \in \boldsymbol{U}$. Or when there is no causal transience and the effects are stable over time then $Y_b(u)$ and $Y_{b'}(u)$ can be measured at different times. In general there will be assumptions about the joint distribution of (Y_B, B) on the universe $\boldsymbol{U}$ that helps to estimate $Y_b(u) - Y_{b'}(u)$ or $\mathrm{E}_U(Y_b(u)- Y_{b'}(u))$.

Some distrust in the potential response model might arise in researchers "*because it has been only partially formalized and, more significant, because it rests on esoteric and seemingly metaphysical vocabulary of counterfactuals*" (Pearl, 1997). As a clarifying formalism the Lewis logic can be used: the tupels $\{(Y_b, b), b \in \boldsymbol{B}\}$ can be seen as a set of possible worlds so that per definition when (Y_b, b) represents the true world the closest one is $(Y_{b'}, b')$ when the causal effect of b is defined relative to b'. With that transformation in mind the counterfactual analysis of the potential-response model can use the axiomatic of the Lewis logic to have rules to handle the counterfactual statements. Another possibility would be to see the joint distribution of $(Y_B, \boldsymbol{B})$ as marginal distribution of a causal network and to use the causal semantic of Pearl (1995) or Heckerman and Shachter (1995) to guide the inference. Galles and Pearl (1997) show that the approaches with Pearls or Lewis framework differ only (according to the valid counterfactual statements in the respective semantics) when the causal model is allowed to be non-recursive: then the causal framework poses more restrictions on the set of valid statements than does the Lewis axiomatic.

An unambiguous advantage of this model is its adequacy and adaptability for different not necessarily randomised experimental settings - including observational studies. The clear definition of what causes are can be considered as positive or negative. The latter because for many investigations especially in the social context it can be seen as being too restrictive. The model cannot be used to answer causal questions in the sense of Ströker's second motivation where one wants to find explanations for certain observed effects.

Decision Theoretic Causal Model

The definition of causality in the model of Heckerman and Shachter (1995) can be seen as Mackie's INUS-condition transposed in a decision theoretic framework. The model consists of a finite set of decisions $\boldsymbol{D}=\{D_1,...,D_m\}$ on certain actions, possible (relevant) states of the world $\boldsymbol{s} \in \boldsymbol{S}$, and a finite set $\boldsymbol{U}$ of consequences of interest represented by variables $\boldsymbol{X}$, $\boldsymbol{Y}$,... . The worldview is pre-deterministic: Given $\boldsymbol{d} = (d_1,...,d_m) \in \boldsymbol{D}$ and $\boldsymbol{s} \in \boldsymbol{S}$ the value of $\boldsymbol{X}(\boldsymbol{d}, \boldsymbol{s}) = \boldsymbol{x}$ is fixed.

Causality is defined via a notion of limited (un)responsiveness. The consequences $\boldsymbol{X} \subseteq \boldsymbol{U}$ are said to be unresponsive to $\boldsymbol{D}$ in states limited by $\boldsymbol{Y} \subseteq \boldsymbol{D} \cup \boldsymbol{U}$ if we believe that for all states of the world $\boldsymbol{s} \in \boldsymbol{S}$ and decisions $\boldsymbol{d_1}, \boldsymbol{d_2} \in \boldsymbol{D}$ whenever $\boldsymbol{Y}(\boldsymbol{s}, \boldsymbol{d_1}) = \boldsymbol{Y}(\boldsymbol{s}, \boldsymbol{d_2})$ is true then $\boldsymbol{X}(\boldsymbol{s}, \boldsymbol{d_1}) = \boldsymbol{X}(\boldsymbol{s}, \boldsymbol{d_2})$ is true as well. A change in the decisions that does not change $\boldsymbol{Y}$ can not change $\boldsymbol{X}$. Then the variables in $\boldsymbol{C} \subset \boldsymbol{D} \cup \boldsymbol{U}$ are said to be causes of a variable $X \in \boldsymbol{U}$ when $\boldsymbol{C}$ is a minimal set such that X is unresponsive to $\boldsymbol{D}$ in states limited by $\boldsymbol{C}$.

This definition matches Mackie's INUS-condition: Let $Y \in \boldsymbol{C}$ be a variable of interest. Then Y is **i**nsufficient for X as long as $Y \neq \boldsymbol{C}$. And Y is **n**on-redundant. Both statements are true because otherwise $\boldsymbol{C}$ would not be a minimal set. The set $\boldsymbol{C}$ is **u**nnecessary for X as the definition does not claim that $\boldsymbol{C}$ was unique in $\boldsymbol{D} \cup \boldsymbol{U}$ and also because there might be $\boldsymbol{s} \in \boldsymbol{S}$ so that X was unresponsive to $\boldsymbol{D}$. The set $\boldsymbol{C}$ is **s**ufficient for X in the sense that for all $\boldsymbol{c} \in \boldsymbol{C}$ the outcome of X is only dependent of $\boldsymbol{s}$ and no longer of $\boldsymbol{d}$. Thus sufficiency means that nothing more can be done to control or influence X.

An advantage of this definition of causality is the possibility that variables that can not be manipulated (attributes in Holland's notion) can be called causes without being unclear what a cause is and without adventurous hypothetical manipulations. With this model causal questions with both motivations in Ströker's sense can be analysed.

Conclusions

All three causal models presented here have advantages and disadvantages. Their main drawbacks are related to their underlying conceptual understanding of causality. As no-one can claim any conceptual model to be the true ontic one the decision to use a causal model at all or which causal model to use depends first of all on the adequacy of the model for the scientific question at hand: Some might be difficult to model in the decision theoretic or experimental milieu so that Pearl's model is the only one that can support the causal analysis. For other studies finding hypothetical atomic interventions for the entities of interest might render the whole analysis dubious. In the second place the decision on using causal models is dependent on the conceptual preferences of the researcher: Whoever wants science to be restricted to the epistemic level will not use of causal models at all. The requirements of a strict interventionist are only met by the potential-response model though the counterfactual argumentation can still be regarded as being awkward. For those who want the scientific use of causality to match the intuitive use in daily life the model of Pearl may be the favorite. A model where entities that can not be manipulated can be called causes and that still has a distinctive definition of what a cause is, is provided by the decision theoretic concept of Heckerman and Shachter (1995).

References

Albert, H. (1968). *Traktat über kritische Vernunft.* J. C. B. Mohr (Paul Siebeck), Tübingen.

Frank, P. (1988). *Das Kausalgesetz und seine Grenzen.* Suhrkamp, Frankfurt.

Galles, D. and Pearl, J. (1997). An Axiomatic Characterization of Causal Counterfactuals, *Technical Report (R-250).* UCLA Cognitive Systems, Los Angeles.

Heckerman, D. and Shachter, R. (1995). Decision Theoretic Foundation for Causal Reasoning, *Journal of Artificial Intelligence Research,* 3, 405-430.

Holland, P. W. (1986). Statistics and Causal Inference (with discussion), *Journal of the American Statistical Association,* Vol. 81, 946-970.

Hume, D. (1995). *Enquiry concerning human understanding.* In James Fieser (ed): *The Writings of David Hume,* Internet Release.

Kant, I. (1922). Metaphysische Anfangsgründe der Naturwissenschaft, *Werke,* Band 4, Cassirer, Berlin.

Koch, G. (1994). *Kausalität, Determinismus und Zufall in der wissenschaftlichen Naturbeschreibung.* In the series "*Erfahrung und Denken*", Band 75, Duncker & Humblot, Berlin.

Lewis, D. (1973). Causation, *Journal of Philosophy,* 70, 556-567.

Mackie, J. L. (1974). *The Cement of the Universe.* Clarendon Press, Oxford.

Mackie, J. L. (1985). Logic and Knowledge, Selected Papers, *Causation in Concept, Knowledge and Reality,* Vol. 1, 178-191.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligence Systems.* Morgan Kaufmann, San Mateo, CA.

Pearl, J. (1995). Causal Diagrams for Empirical Research (with discussion), *Biometrika,* 82, 669-710.

Pearl, J. (1997). The New Challenge: From a Century of Statistics to an Age of Causation, *Computing Science and Statistics,* 29, 415-423.

Pearson, K. (1911). *Grammar of Science (3rd ed.).* A. & C. Black, London.

Rubin, D. B. (1974). Estimating Causal Effects of Treatments in Randomized and Non-Randomized Studies, *Journal of Educational Psychology,* 66, 688-701.

Russell, B. (1913). On the Notion of Cause, *Proceedings of the Aristotelian Society, New Series,* Vol. 13, 1-26.

Ströker, E. (1992). Warum-Fragen. Schwierigkeiten mit einem Modell für kausale Erklärungen. In R. Bubner (ed.), *Kausalität,* Neue Hefte für Philosophie, 32/33, Vandenhoeck & Ruprecht, Göttingen, 105-129.

Suppes, P. (1970). *A Probabilistic Theory of Causality.* North-Holland, Amsterdam.

von Wright, G. H. (1971). *Explanation and Understanding.* Cornell University Press, Ithaca, New York.