Causality in applied statistical research

Jon Hovland

1. Nov. 05

Introduction

"For whom the bell curves" is an international project drawn on the finding that statistics as a mathematical field is developed in interaction with the State (Desrosires 1998, Hacking 1990, Lie et al. 2001, Porter 1995). The project will study in which way governing and statistics is coupled today, and research consequences of this coupling. A sub-project will examine applied statistical method, and derive ontologies implied by the algorithms in use (Sætnan 2005). This paper examines whether there are specific ideas of causality implied in these practices, and possible consequences of this.
This question will be entered through first a slight overview of the meaning of causality, and different understandings of this concept. To identify the statisticians´ use of causality and reduce bias, we will reduce the units investigated from individual to actions, that is; actions of research (based on ideas from Mjøset 2001). It is concluded that statistical knowledge is limited to a very few understandings of causality, which leaves the statistician with some difficult choices. To trace consequences of this, we will look into the domain of the "Numbers as language"-discourse. The final conclusion arrives at a warning of limitation of thought through method-discourses centred on statistics and numbers.

Causality

"[...] how totally pointless, pointless, and downright silly it is to think that one can ever state precisely what it is for one thing to cause another."(K Codell Carter 1991 in Kern 2004: 23)

Simple overviews of causality are impossible, as causality means very different things. As always there is an element of different interpretations of words, however, there is more to it: Differences often seem to narrow down to differences in world views. Therefore this section will be organized in sampled ideas from western (Aristotle, Hume) and eastern (Vaiśeṣika, Confucius) classical Philosophy, Logics, Literature and Medicine. This will all be somewhat confusing, and will be summed up through five general factors that we observe in questions of causality.

Western classical Philosophy

Aristotle and Hume are seen as premise setters on subjects of causality in philosophy. Aristotle distinguished between four forms of causality (Gilson 1984: 1-16). (1) The material cause, what things are made from - paper is the material cause of the print of this paper; (2) the formal cause: the pattern, model, or structure upon which things are made - the formal cause of this paper would be the scientific discourse of paper writing. (3) The efficient cause: What makes something come into existence. I would be the efficient cause of this paper. (4) The final cause: The reason for which the something is put to existence - the potential completion of my Ph. D is the cause of this papers existence. Aristotle found the final cause the most significant.

Hume put things in a stricter way: *"[...]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 had not been, the second had never existed"* (Hume 1772 [2000]: 54). In part one causes are nothing but situations where all c in C [1] and always causes one e of type E. Cause is therefore succession and regularity. The second part ("Or in other words...") rather implies that all e in E presupposes one c in C; something rather different, later in history of sciende to be termed counterfactual analysis (Lewis 1973). Hume found cause and effect the single most useful notions in science, as knowledge of these will *"[...] teach us to control and regulate future events by their causes (ibid.: 53)."*.

Eastern classical Philosophy

I will use two central directions in eastern Philosophy to observe some fundamental differences as to classical western thought. The Indian Nyāya-Vaiśeṣika based on the-

[1] A note on notation: By small letters I mean single phenomena, large letters express typologies of phenomena. In other words: The purchase of my Mercedes 1983 No W123855594831973 September 2004 is phenomena (c), of the type(C) buying car, whilst bad economy (E) may be exampled by not paying housrents before November 2004 (e).

orising *dharma* (or occurrence), and the Chinese Confucianism.

In Nyāya-Vaiśeṣika cause is generally everything that does not obstruct the effect. These will further be recognised as (1) the substantial cause; the substance in which the effect takes place [2]. All other conditions relevant to the effect are grouped into efficient and nonsubstantial causes. The efficient causes are rather similar with Aristotle´ s notion. The nonsubstantial effect is the idea of the substance involved in conditioning the effect (Matilal 1975: 41-45). Translated this would probably mean that causing a sweater to become red due to the colour of the thread, will not hinder the fact that others may produce red sweaters from other red threads.

To avoid irrelevant causal conditions to be taken account of, the Nyāya will consider conditionality and unconditionality. Conditioned relations are equivalent to spurious effects as we know them from statistical language: The fact that both my colleague and me do science studies and use gray machines, is conditioned otherwise, that is; through corporate agreements, and is not causally relevant.

Confucianism as presented in its *Book of Changes* is a presentation of causes and effects, a complex study of "if-then" situations. From this we would conclude a quite humerian notion of cause. However, this is again based on an organic understanding of the world, in which multiple causation is self-evident. That is, everything is a cause of the effect, and others than the one cause identified or sought in the purpose of an end, may be the most powerful (Wu 1975: 13-22). From this one will conclude that this is a pragmatic notion of causation, where one will identify causes to organize knowledge, but without any attempt of providing laws, as these are not accountable.

Logic

Causality is obviously a challenging subject to philosophers of logic: It is absolutely central in our ways of organizing the world around us, on the other hand it is not accomplishable - how to define causality logically. Lewis (1973: 556) has derived Hume´ s definition like this:

"[...]let C be the proposition that c exists (or occurs) and let E be the proposition that e exists. Then c causes e, according to a typical regularity analysis ,iff [3] *(1) C and E are true; and (2) for some nonempty set ε of true law-propositions and some set φ of true propositions of particular fact, ε and φ jointly imply $C \supset E$, although ε*

[2] Similar to Aristotle´ s material cause, although one would in this case also include nonmaterial substances such as mind and spirit.

[3] "If, and only if"

and φ jointly do not imply E and φ alone does not imply $C \supset E$."

In English: By stating that c causes e, we imply that categories C and E exist. Putting that implication aside we further imply that these are regulated by law(s) ε and conditioned by fact(s) φ (translating to sociology, we would term this 'context') that together insists on a causal relationship between C and E. These laws ε and facts (context) φ to not cause E, and the facts (context) alone do not create the causal relationship. This creates problems for the logician, as examples will be non-realistic[4]. In the following we will have a look at attemptions to solve this problem.

1. All effects have causes, we just cannot see them all: *"In nature nothing remains constant. Everything is in a perpetual state of transformation, motion, and change. However, we discover that nothing simply surges up out of nothing without having antecedents that existed before. Likewise, nothing ever disappears without a trace, in the sense that it gives rise to absolutely nothing existing at later times. This general characteristic of our world can be expressed in terms of a principle which summarizes an enormous domain of different kinds of experience and which has never yet been contradicted in any observation or experiment, scientific or otherwise; namely, everything comes from other things and gives rise to other things"* (Bohm 1999: 1).

2. Rejecting of overdetermination, or never allowing more than one cause. In doing one assumes that all other factors than the cause will be pre-empted from causation. This way one can determine a one-way causality (Strand 2001: 16). This is then done by finding the regular preceding event. This is termed singularism (Beebee 2003: 258). *Ceteris paribus* is the expression for "all else held constant". This means that one will assume causality under the theoretical condition of holding all other factors at a constant level, and thereafter observe possible relationships (Viner 1917: 241). This latter is the philosophy of experimental science, and these two concepts combined create a non-cyclic notion of causality.

3. The above explanations result in scientific determinism. As laws of causality are not accomplishable without determination, and direction is not possible to determine positively, one should rather suggest causation´ s *"[...]complete extrusion from*

[4]An attempt of a realistic example: The contemporary bird virus (c) will cause a pandemic (e). Viruses (C) and pandemics (E) exist. ε: 1. Animal viruses with certain qualities do under certain conditions spread to humans. 2. Once spread to humans, certain viruses cause pandemics. φ: 1. H5N1 has the ability to transform to a human virus. 2. In areas with bird virus spread, the conditions to transform are present. 3. H5N1 is a virus of a kind that will spread quickly among humans. From the definition, φ does not explain the causality, and ε and φ do not in themselves cause E, C is still the cause. However, is this entirely true?

the philosophical vocabulary", (Russell 1917, in Matilal 1973: 45)One can only speak of coexistence and correlation.

4. Given that these laws are not accomplishable (causal dependence is impossible), one must reject the idea of c causing e, however not the notion of e being caused by c. Then we may state that e is the cause of e, even though e may in other instances be caused by something else: (d)(Lewis 1973: 563). This is to abolish prediction in causing, but not its power of explanation. Sociologists would be likely to recognize Foucault´ s notion of power in this understanding of causality, where everything seem to be in drift, but causal connections may be done in retrospect - evrything happens for some reason (Sedgwick 2002: 76).

Generally, these solutions to logical problems of causation single out smaller parts of the idea of causation, to thereby provide an acceptable, but highly limiting idea of causes. As we will see, these are not in tune with Literature informed on contemporary forensic psychology and medicine. The next section involves cause seen from forensic medicine, a field where strict rules of causality as they are observed in the text above is no longer respected, however conculsions are yet stated with certainty.

Literature

According to Kern´ s overview of Literature that uses its contemporary forensic medical evaluations from 1830 up to today, we find accounts of cause that does not match those of Philosophy. He defines cause quite broad and circularly as *"[...] a metaphysical (or ontological) concept that refers to actual cause-and-effect relations between events or to dynamic interactions and processes in the world; it is also an epistemological concept that refers to knowledge involved in answering "why" questions about those interactions and processes"* (Kern 2004: 23). He observes two features of the causality in the literature he studies. (1) There is an increasing specificity to the causes that are provided (ibid.: 7). (2) Human causes may be distinguished as motives (inner impulses towards action), intentions (object-oriented plans for fulfilment of motive), purposes of the actions directed towards the goal and reasons as rational grounds for the behaviour (ibid.: 23). He classifies the findings within these concepts as causes of ancestry, childhood, language, sexuality, emotion, mind, society and ideas (ibid.). We will not look further into these groups of reasons. However, it is clear that these notions of cause, do not allow themselves to be limited by logical problems of cause, as those described in above sections.

Medicine

Finally, we will include Freud´ s classes of causes, not because of its relevance to Medicine, but because it is a broad and vivid illustration of cause as seen from a psychoanalysis point of view. His statements as referred to here, are reply to the criticism of his psychoanalysis, a critique that states that since anxious states appear immediately after mental shock, it is unlikely that otherwise would be the cause of the anxious state. Freud replies that one should consult four concepts concerning understanding diseases: The releasing cause; simply the last factor to appear, immediately before the manifestation of the effect. Predisposition; factors in whose absence the effect would never come about. Specific cause; one that is never absent when the effect actually takes place, and at the sufficient intensity will bring about the effect. Contributory causes; such that are not necessary, but co-operate with the predispositions and specific causes (Freud in Lyon 1967: 19).

Summary

The above descriptions provide a slight overview of the different approaches towards causality. They seem disconnected, but after a second read-through we find that they can be described by five factors that one will have to relate to; temporal distance, logical distance, singularity, necessity, and substance. The temporal distance relates to in which degree one allows distance in time between cause and effect. The logical distance refers to the lengthiness of reasoning there is allowed in describing causality. Singularity refers to the complexity allowed in causal description. Necessity refers to whether everything is caused, or only some phenomena. Finally, substance refers to the degrees of materiality required in causes and effects; whether mind can be a cause.
Reduced temporal distance makes determination easier, but increases the danger of spuriousity. Reduced logical distance involves similar dilemmas, but also; increased logical distance enhances bias. Singularity provides better grounding for prediction, but again, there are dangers of spuriousity. Necessity of causality is matter of freedom of will vs. scientific laws: If everything is caused (counterfactual analysis omissed), then everything is determined. The substance axis from material to non-material concerns to what degree one allows invisible elements in the causal chain, the dilemma is one of bias vs. theoretical limitation.

Actions of research and causality

The assumptions about causality in the previous are to some degree connected to general attitudes towards actions of research, and what science is. We will now start

looking into how ideals of research is connected to different limitations on the axis of the different factors of causality. Mjøset refers to three different values concerning acts of research in different notions of theory (2001: 15641). They are referred to as (1) "Committed to the deductive-nomological ideal", (2) "Accepts that sciences studying human actions are distinct from generalizing natural sciences", and (3) "Accepts that ethical fundamentals matter for social science theory". For the sake of the argument, I will not relate these attitudes towards disciplines nor individuals, rather to actions of research.
There is plenty reason to ponder the relationships between all these research act qualities to different notions of causality. However, this specific project concerns statistics, and these methods will be found in group (1); deductive-nomological ideals[5]. The strict frame of numerological-deductive research acts will also apply strict boundaries as to logic and causality. As we see from the dilemmas that the different factors of causality as presented above represent, demands of generalisability and logical deduction is likely to produce one out of two solutions: Give up on causality, or chase the holy grail by developing ever more complex methods in the belief that one will some day provide a method of studies of correlation and variance that will infer absolute causation. We will return to this idea in the conclusion.

Numbers, Acts of research and Power

There is a concurrence between numerological-deductive research acts and the power of numbers. Relating to cumulative theory-building towards more generalized and abstract laws is at its most famous articulated through $E = mc^2$. Numbers are also abstractions, easily attainable in their presentation, but not in their making, and also positive and final. There are two issues to discuss relating to this: The first danger of this is seen through the descriptions of causality. This class of science confronts an importunate dilemma in the logic leap from concurrence through connexion to causality. There are three solutions to this: (1a) Causality is not scientific. We may assume correlation and connections, not causality (represented by Bertram Russell, above citation). (1b) We may observe causality between single phenomena c and e, but not generally between C and E (represented by Nancy Cartwright, 1983). (2) Given that all events are part of cause and effect patterns, and causes match their effects perfectly once all qualities and laws are known, all events are determined (Bohm and Viner, above citations).

This does obviously not reflect the conclusions of all scientists working with statis-

[5]There is all possible reason to believe they will relate to group (2), but this is not the fundament for statistical methods.

tics, which brings us on to second issue. "Statistics is a science that is not exactly mathematics. It *uses* mathematics, in the form of probability, but its procedures are the inverse ones of fitting probability distributions to real-world data" (Cooke 2005: 511). It co-exists and communicates other ways of research actions. It has to, because although mathematics are used, one has to make assumptions and simplifications to do inverse probability.

Both of these troubles relate to causality. The first in its very limiting understanding of causality, the second will be framed through the same dilemma, but still have the opportunity to stretch towards other understandings. In both cases, I will emphasise that the power of conviction is strengthened with the "Numbers as language" discourse, which in turn strengthens the notions of causality implied within these disciplines. Also, in the latter case, whatever notion of causality that is implied might be more importunate than otherwise. Concerning causality this may have two consequences: (1) An interpretive understanding of causality implied in discussing causality in a broader term is hidden from the recipients of scientific knowledge. (2) Understanding of causality is generally shifted towards those suggested as 1a, 1b and 2 above. The power of these consequences is possibly strengthened by general ideas about numbers, or "Numbers as language".

The Language of Numbers

Do numbers make a language? The question of numbers as language may well be written off by ridicule; you may order 2 coffee, but not the substance coffee by way of numbers[6]. However, if we look into what a language actually is, things get complicated. This section is dedicated to pros and cons of naming numbers as language. The conclusion "not language" is followed up by suggestions of "Numbers as language"-discourse.

"[...]the symbols on a page are just a representation of mathematics. When read by a competent performer (in this case, someone trained in mathematics), the symbols on the printed page come alive - the mathematics lives and breathes on the mind of the reader like some abstract symphony." (Devlin 1998: 5)

Following grammar science, we will be looking for characters and signs, language modules in morphology (forms derived from the same word), phonology (the systematic uses of sounds) and syntax (how words are put together to sentences. Further-

[6]You may translate verbal language to binary language, but this is irrelevant. This is not useful for practical purposes, and is the equivalent of stating that a picture may tell what a letter may, by taking a photo of the letter. The verbal language is the premise provider.

more, one will look other qualities in language, which would be semantics (studies of meaning) and connotations[7].
We definitely find characters and symbols in numbers. Furthermore, we will find representations of the modules. Syntax is represented through techniques, rules and formulas of derivations and calculations. Morphology is represented through different techniques of annotation [8] Succeeding Chomsky revolutionary work in the 50´ s, one has come far in turning linguistics into mathematics. The reason for the progress in this work should, however, maybe not be ascribed to its substantial meaning, but where the money is. Programming- and artificial intelligence industry is dependent on mathematical interpretations of language. We are able to observe this through the fact that breaking-edge computational linguistics do meet hindrances that generally are solved through procedures of probability-calculations [9]. This is not surprising knowing that number-language (if should be labelled so) involves gradation, but lacks instruments of nuance, that is semantics, connotations and phonology. This lack of nuance-providers is why numbers are not fruitful to discuss as a language.

Still there is an idea of "Numbers as language". We have learned it in school (maths is the universal language), we observe it in "Far Side Gallery"-cartoons, where the scientists make jokes by way of formulas on the black-board, or in movies like "Good Will Hunting", where the geniuses communicate through a sometimes duel-like correspondence of mathematical expressions. If we inquire a discourse of "Numbers as Language", we would probably find that this is seen as a language that can reach higher abstractions, and reach further than ordinary language, that it is a language that facilitates logic and reason. Furthermore that the guardians of this language have high social esteem as administrators of truth, similar to the status of those skilled in writing would be in the age of runes on Norway, or the hierographers of ancient Egypt. Numbers will also be a system of preserving reality and truth, and therefore should be highly respected

We have shown how causality is restricted in statistical research, and how this restriction is derived from the clash between deductive-numerological ideals and causal logical challenges, and on the other hand the power of numbers, here explained through the suggestion of a "numbers as language"-discourse. Further research will involve thorough enquieries of the "numbers as language"-discourse, and destillations of what is not mathematics in statistics, and how this can be evidenced as proofs of substance implied in algorithms. There is also reason to further investigate the five factors of

[7]This is based on the definitions of language in Bye et. al. (2003: 17-47).

[8]Take 'x' as in 'unknown. In which way it is unknown, what it relates to, or essential qualities of x, changes in annotations like $\bar{x}$, $\acute{x}$, $\underline{x}$, $\breve{x}$, X, $\mathbf{X}$, $\check{x}$, $\dot{x}$, $\hat{x}$, $\vec{x}$, x^T, x_1 or x_i^k.

[9]Review any volume of *Computational Linguistics*, e.g. O´ Donovan et. al. 2005

cause as they are suggested, and whether they continue to prove useful.

References

- Bohm (1999) David: *Causality and Chance in Modern Physics*, Philadelphia: University of Pennsylvania Press.
- Bye (2003) Patrik, Trond Trosterud og Øystein Vangsnes: *Språk og språkvitskap: Ei innføring i lingvistikk*, Oslo: Det Norske Samlaget.
- Cartwright (1983) Nancy: *How the laws of physics lie*, Oxford: Clarendon Press.
- Cooke (2005) Roger: *The History of Mathematics: A Brief Course*, 2nd ed, US: John Wiley & Sons, Inc.
- Desrosiéres (1998) Alain: *The Politics of Large Numbers*, UK: Harvard University Press.
- Devlin (1998) Keith: *The Language of Mathematics: Making the invisible visible*, New York: W. H. Freeman and Company.
- Gilson (1984) Étienne: *From Aristotle to Darwin and back again: a journey in final causality, species, and evolution*, Paris: University of Notre Dame Press.
- Hacing (1990) Ian: *The Taming of Chance*, Cambridge: Cambridge University Press.
- Hume (1772 [2000]) David: *An Enquiry Concerning Human Understanding*, Ontario: Batoche Books.
- Kern (2004) Stephen: *A Cultural History of Causality: Science, Murder Novels, and Systems of Thought*, US: Princeton Unversity Press.
- Lewis (1973) David: "Causation", *The Journal of Philosophy*, Vol. 70, No 17 (Okt. 11, 1973), s556-567. Journal of Philosophy Inc.: US.
- Lie (2001) Einar og Hege Roll-Hansen: *Faktisk talt: Statistikkens historie i Norge*, Oslo: Universitetsforlaget.
- Lyon (1967) Ardon: "Causality", *The British Journal for the Philosophy of Science*, Vol. 18, No 1 (May, 1967), s1-20. Oxford University Press.

- Matilal (1975) Bimal Krishna: "Causality in the Nyāya-Vaiśeṣika School", in *Philosophy East and West*, Vol. 25, No. 1, pp. 41-48. US: University of Hawai`i Press.

- Mjøset (2001) Lars: "Theory: Conceptions in the Social Sciences" in *The International Encyclopedia of the Social and Behavioral Sciences*, pp. 15641-15647, Elsevier Science Ltd.

- O´ Donovan (2005)Ruth, Michael Burke, Aoife Cahill, Josef van Genebith & Andy Way: " Large-Scale Induction and Evaluation of Lexical Resources from the Penn-II and Penn-III Treebanks" in *Computational Lingustics*, Vol. 31, No3, pp 329-367.

- Porter (1995) Theodore M.: *Trust in numbers : the pursuit of objectivity in science and public life*, New Jersey: Princeton University Press.

- Sedgwick (2002): *Cultural Theory: Key Thinkers*, Routledge.

- Strand (2001) Anders: *The Fallacy of Epiphobia - A Discussion on the Prospects of Mental Causation*, Oslo: Universitetet i Oslo, Institutt for Filosofi.

- Sætnan (2005) Ann Rudinow: "For Whom the Bell Curves". Project home page: http://www.svt.ntnu.no/iss/ann.r.saetnan/bell/project.html

- Viner (1917) J.: "Some Problems of Logical Method in Political Economy", in *The Journal of Political Economy*, Vol.25, No 3, pp. 236-260.