Actual Causation

De Craecker, Tim
Ghent University, applied mathematics and computer science
Krijgslaan 281, S9
Ghent (9000),Belgium
E-mail: tim.decraecker@ugent.be

Introduction

The past two decades causal models have shed light on numerous problems in causal inference ([Pearl, 2000]). One area which is still somewhat shrouded in mystery is actual causation. Actual causation aims at identifying the causes of an event in a particular scenario. Or put more simply: actual causation is concerned with problems which in our ordinary language take the form "did 'x' cause 'y' ?", "Why did 'y' happen ?", etc. Traditionally this topic has been studied extensively in philosophy but has attracted far less attention in other fields. One exception though, is the law. Causation plays a significant role in the law ([Moore, 2009]). This inspired some recent developments ([Chockler et al., 2004] and [Halpern et al., 2010]) and much of the current paper.

The outline of the paper is as follows: in the first section I want to introduce the general topic of this paper and give a brief sketch of some common reasons which might motivate the search for actual causes. This, in order to have some idea for which a study of actual causation might be useful, but also to avoid philosophical confusion merely grounded in a different use of the same causal terminology. In the second section I want to give a short discussion on the topics which attracted most attention a decade since causal models were introduced to study actual causation ([Pearl, 2000], chapter 10). In the final section I want to consider the notion of proximate cause, to discuss one way how to distinguish which causes are more or less relevant and to make the connection with some older work in causality by I.J. Good ([Good, 1961/62]) .

Motivation

Actual causation plays a crucial role in constructing adequate causal explanations. In real life people often find little difficulty in answering questions of the form "Did 'x' cause 'y' ?" and 'Why did 'y' happen ?". However, since we use this same causal vocabulary to express inquiries which are quite different in nature, considerable confusion might arise if we want to study causal explanation within a formal framework. The kind of problems we raise, implicitly, when asking for a causal explanation can be very different, ranging from inquiries concerned with assigning responsibility to agents ([Good, 1999], [Moore, 2009]) to inquiries concerned with the design of effective strategies for intervening ([Hitchock et al., 2009]). In past, perhaps all too often, philosophers started a formal analysis of "causation" before specifying in much detail the kind of inquiry in which "causation" was supposed to be used. As a consequence even seemingly simple restrictions one could impose on causation lead to hot debate: "Is causation transitive ?" ([Hitchcock 2001]) "Can 'absent events' or 'omissions to act' be causes ?" ([Schaffer, 2000]) to name but a few popular topics in the philosophy of causation.

So, in what sense does the specific nature of the inquiry influence our causal judgments ? Usually there are a lot events that have made a difference, in some way, to the occurrence of the event for which we want a causal explanation. Nevertheless, in each case we pay attention to only a small fraction of these events. To quote a much cited example: "It is the strike of the match which caused the fire, not the oxygen in the surrounding air". How we distinguish relevant from irrelevant causes is

one of the topics which currently attracts attention ([Halpern et al., 2010]).What causes we consider more, or less relevant depends for a large part on the reasons that motivate our search for causes. To give an illustration (even if it is a bit a silly example) : In case an alarm goes off, a burglar might want to know how he could have avoided the alarm going off, the householder might be worried about what triggered the alarm, a neighbor might want to know who's responsible for that annoying sound nearby, etc. How large the differences between these various motivations might be, it's a worthy endeavor to look for general principles we use in distinguishing relevant from irrelevant causes and which enable us to construct adequate explanations.

First let me give a rough sketch of the two kind of inquiries previously mentioned:

1. Assigning responsibility: actual causation might be useful in inquiries which are concerned with who to hold responsible (and to what extent) for a certain event (e.g. damage to a plaintiff's property etc.) or in justifying the consequences of our actions (e.g. why did you arrive late ? etc.). Responsibility might be influenced by agent-relative factors, like whether an agent (e.g. the defendant in a trial) intended the event to happen or whether he foresaw his actions would bring the event about. The law explicitly makes a distinction [1] between a responsibility based on such agent-relative factors (mens rea) and a responsibility based on "objective" factors (actus reus). It's especially the latter in which causation has a significant role to play.

   > Example: to give one simple illustration of how causation can influence our moral judgments, consider a soldier being killed after an officer had send him on a risky mission, in hope he would not survive. How morally degraded the officer's behavior might be, we still might consider his action less evil compared to an officer who would have killed the soldier in cold blood.

2. Designing effective strategies for intervening: actual causation might be useful in selecting suitable targets for intervention. For instance, in case of inquiries investigating how a certain event could have been avoided. This motivation is discussed at length in [Hitchock et al., 2009]. Good strategies for intervening are ones which are in general effective in avoiding the event, without having much undesirable side effects. Off course, knowing the causes of an event might also guide our subsequent behavior directly after that event.

Whatever the approach to causal explanation, for some inquiries it is required to incorporate information that goes beyond the information contained in the causal model that captures the effect of interventions on the actual world: statistical information and descriptions about how things ought to behave provide valuable information for the inquiries previously mentioned ([Hitchock et al., 2009]). One recent approach [Halpern et al., 2010] combines several of these factors into one unified concept, norms, which presumably impose an ordering on all possible worlds. On its turn, this ordering can be used to separate relevant from irrelevant causes.

Instead, in the last section, I will consider an older approach, pioneered by Good [Good, 1961/62]. The main idea behind this approach is, first, to develop a measure of the general causal tendency of an event to cause some other event and, afterwards, to use this concept to construct a quantity that measures the degree by which an event actually caused some other event. Such a quantitative approach can possibly capture nuances which the qualitative approach of [Halpern et al., 2010] can not, at least not in its current state. Moreover, this approach crosses the bridge between "general" and "singular" causation. Two subjects which are still treated rather separately in the causal literature, even though

[1] The distinction is also made in [Good, 1999] and [Chockler et al., 2004]

causal models provide a popular framework for both topics. The main difficulties stem from the fact that it is not straightforward at all to develop measures of causal tendency and actual causation which work in general.

Actual Causation in Causal Models

In this section I want to clarify briefly in what sense causal models have shed light on this topic. In short, a causal model consists of a set of variables and a set of "structural" equations. These equations govern how a change in one variable might affect the other variables in the system. A distinction is made between exogenous variables which behavior is determined by factors outside the system and endogenous variables which behavior is modeled by the structural equations. Each setting of the exogenous variables constitutes a possible world. Statements about actual causation are relative to one particular world (the actual world happened).

Considerable effort went into sharpening the intuïtion that a cause 'x' of an event 'y' should, at least, have made "*some* difference" to 'y'. For instance, in law the but-for-test asks whether 'x' was necessary for 'y' in the sense that if 'x' would not have happened neither would 'y' have happened, the more elaborate NESS test asks whether 'x' was a necessary part of a set of events sufficient for 'y'. [2] Causal models provide a clear semantics for counterfactuals [Pearl, 2000] and several proposals have been raised to give more precise formulations of the intuïtion mentioned ([Pearl, 2000], [Hitchcock, 2001], [Halpern et al., 2005], [Glymour et al., 2007] and [Halpern, 2008]). I'm not concerned with the details of these proposals, it's sufficient to capture the general idea:

> In order for 'x' to be a cause of 'y' at least one of the paths connecting 'x' to 'y' must have been "produced" by 'x', i.e. it's possible to change the values of all the variables on one of the paths linking 'x' to 'y' through an "experiment", on the *actual* world, in which we change the value of 'X'. When we change 'X' back to its original value we can witness how 'y' is being produced from 'x' in a slightly modified version of the actual world.

Much additional effort might go into specifying further which "experiments" can be allowed. Usually it is necessary to fix some of the variables in the model on a value different from its actual value in order to witness that 'x' can produce 'y'. Furthermore, as noted in the previous section, we do not consider all "causes" equally relevant. Some counterfactual considerations are obviously too "far fetched": for instance, counterfactuals which make reference to "abnormal" or "exceptional" conditions or counterfactuals which make reference to very remote conditions. The former provide much of the intuïtion behind the approach of [Halpern et al., 2010] (roughly: in searching for actual causes we only consider modified versions of the actual world which are more "normal" then the actual world), the latter will be considered in the next section.

Degrees Of Causation

Whether the goal is to design effective ways of intervening or to assign responsibility to agents, a more quantitative theory of actual causation seems appropriate (this is sometimes expressed through the phrase "causation is not all or nothing but comes in degrees"). After all, in the former case some strategies of intervening might be better then others, in the latter case we do not punish all wrongdoers in the same way (even if the severity of the harm caused is similar). These differences must depend to some extent on the causal structure and on statistical information. In this section I want to explore this topic.

[2] Ideas similar to NESS are mentioned in philosophy [Mackie, 1965] and in epidemiology ([Rothman et al., 2005])

There might be several motivations to introduce degrees of causation. One motivation which I only briefly want to mention comes from cases where several agents bring about a result together. In such cases it might be relevant to apportion responsibility among these agents. The most simple cases are ones in which each agent contributed some definite part to the outcome. A typical example might involve a competition in which each jury member gives a score to each participant which contributes to the total score of that participant. A discussion on this topic can be found in [Chockler et al., 2004].

A second, more subtle way in which degrees of causation come into play might be reveled if one considers the notion of "proximate cause": in order for a defendant to be held liable for a certain harm, the harm should not be a "too remote consequence" of the defendant's wrongful conduct. The easiest way to grasp this intuïtion is perhaps to consider cases where a large number of causes intervened the process that lead from the defendant's act to the harm. For instance, consider the following court case cited by Moore:

> Example: Berry v. Borough of Sugar Notch ([Moore, 2009], p122)
> *A motorman recklessly speeds early on his route. No one is injured while he is speeding, and when he catches up his schedule he resumes his normal, non-reckless speed. Nonetheless, because he sped early on his route, he arrives at the last part of his route just in time to have a tree fall on his car, injuring a passenger.*

Even if it can be proved that the harm wouldn't have occurred if the defendant had traveled with normal speed along the whole route, he might still be relieved from all liability on grounds that the injury was a too remote consequence of his wrongful conduct (the speeding). The law contains various different kind of doctrines, loosely built around the intuïtion mentioned, which might limit a defendant's liability. For instance, in the former case it might be argued that there was simply too much time between the defendant's reckless speeding and the injury. Or one might argue that the defendant should be relieved from all the harm which happened after the tree fell on the car (using the legal doctrine of intervening cause). [3]

Although these various doctrines sharpen the intuïtion behind the phrase "proximate cause" to some extent, it's desirable to construct more clear and general principles. I believe that some early attempts by Good ([Good, 1961/62/91/99],) to define measures for the general tendency of x to cause y ($Q_{suf}(y:x)$ [Good, 1994]) and to quantify the degree in which x actually caused y can best be viewed in this light.

Later, Pearl ([Pearl, 2000]) constructed a pair of related concepts in the framework of causal models: the probability of sufficiency, or the probability of x to produce y in worlds were both x and y are absent, can replace Good's measure of Q_{suf}. And the notion of actual causation, which I already discussed in the second section can replace Good's measure of actual causation. However, a scalar nature of actual causation is not taken into account, although the notion of proximate cause in law seems to suggest that this would be appropriate. Nor is the probability of sufficiency always the most appropriate measure to quantify the general tendency of x to cause y. I hope the following example can illustrate both points:

> Example: an urn contains one red and one black ball, one ball is in the left part of the urn the other in the right part. Consider the following game: at the start of the game the player has to decide in which order the balls are put in the urn. After a number of n shakes $\hat{s_1}, \ldots, \hat{s_n}$ the ball in the left part of the urn is taken. Suppose that by shaking

[3] [Moore, 2009] (chapter 4) contains a detailed discussion of the various kinds of proximate cause tests in the law.

the urn once there is a chance p that one causes a permutation of the balls. The player wins the game if the red ball is chosen. Let R_i denote the position of the red ball at time i ($R_i = 1$ if the red ball is in the left part of the urn), P_i denote whether a permutation takes place at time i ($P_i = 1$ if this is the case) and E whether the player wins ($E = 1$ if this is the case), and finally let A denote whether an even or odd number of permutations took place (A=1 if even). [4]

Causal Model:

$$\begin{array}{ccccc} & \hat{s_1} & \cdots & \hat{s_n} & \\ & \downarrow & & \downarrow & \\ & P_1 & & P_n & \\ & \downarrow \mathcal{P}(P_1=1)=p & & \downarrow \mathcal{P}(P_n=1)=p & \\ \hat{R_0} \rightarrow & R_1 & \rightarrow \cdots \rightarrow & R_n & \rightarrow E \end{array}$$

simplified model:

$$\begin{array}{c} R_0 \rightarrow E = AR_0 \vee (\neg A)(\neg R_0) \\ \uparrow \\ A \end{array}$$

Suppose that for every game we play $n = 4$ and $p = \frac{1}{4}$ (the probability of an even number of permutations is given by $\frac{17}{32}$) and that the actual course of events is given by:

$$R_0 = 1 \overset{\frac{1}{4}}{\rightarrow} R_1 = 0 \overset{\frac{1}{4}}{\rightarrow} R_2 = 1 \overset{\frac{3}{4}}{\rightarrow} R_3 = 1 \overset{\frac{3}{4}}{\rightarrow} R_4 = 1 \rightarrow E = 1$$

In the slightly modified version of the actual world below, we can witness that $\hat{R_0} = 1$ produces $E = 1$. Therefor $R_0 = 1$ is an actual cause of $E = 1$.

$$\begin{array}{ccccccccc} & & P_1 = 1 & & P_2 = 1 & & P_3 = 0 & & P_4 = 0 \\ & & \downarrow & & \downarrow & & \downarrow & & \downarrow \\ R_0 = 0 \rightarrow & R_1 = 1 & \rightarrow & R_2 = 0 & \rightarrow & R_3 = 0 & \rightarrow & R_4 = 0 & \rightarrow E = 0 \end{array}$$

However $R_0 = 1$ should be qualified as only a weak cause of $E = 1$. How to measure its strength ? Since we have no control on the details of the various processes that might lead from $R_0 = 1$ to $E = 1$ (this would be different, if for instance we were allowed to change the number of shakes), I believe it's appropriate to ignore the substructure of the processes leading from $R_0 = 1$ to $E = 1$ and to use the general tendency of $R_0 = 1$ to cause $E = 1$ to measure the strength of the chain connecting $R_0 = 1$ with $E = 1$ (in accordance with Good's measure of strength of a single causal link [Good, 1961/62]).

In this case the probability of sufficiency seems not an appropriate quantity to measure the tendency of $R_0 = 1$ to cause $E = 1$. $P(E[R_0 = 1] = 1 | R_0 = 0, E = 0) = P(E[R_0 = 1] = 1 | A = 1) = 1$ (Y[X] are counterfactuals). If we would use this quantity to measure tendency, we would conclude that $R_0 = 1$ strongly caused $E = 1$. In this case Good's measure of sufficiency seems more reasonable: the most straightforward extension of Good's measure of causal tendency to causal models is, in this case, $log(\frac{P(E[R_0=0]=0)}{P(E[R_0=1]=0)} = log(\frac{P(A=1)}{P(A=0)}) = log(\frac{17}{15})$ which is slightly positive.

Remark, that in making the chain leading from R_0 to E longer, by doing additional shaking, we can weaken the strength of the chain. Furthermore, if we would decide whether a permutation happens or not by a coin toss, one single coin toss would be sufficient to diminish the strength of the link between $R_0 = 1$ and $E = 1$ to zero. Both the gradual diminishment of causation by adding more links between cause and effect, and the sudden "breaks" in a chain, because of a particular intervening cause, are qualitative aspects mentioned in the work of [Moore, 2009] in relation to the notion of proximate cause.

Off course this is only a special case. Often we do distinguish the various processes that lead from cause to effect. Much additional effort might go into quantifying the degrees of actual causation in more general settings (for instance [Good, 1999]).

[4] For a realistic, but complex, case which is similar in some respects: see for instance [Bayer et al., 1992]

Although the notions of "probability of sufficiency" and "susceptibility" are closely linked with actual causation, the topic of actual causation is not well known outside the philosophical literature on causation. I hope in this short paper I was able to clarify this topic and to bring the notion of degrees of actual causation back under attention.

REFERENCES (RÉFERENCES)

[Bayer et al., 1992] D. Bayer and P. Diaconis (1992): Trailing the Dovetail Shuffle to its Liar, The Annals of Applied Probability: vol 2, No.2, 294-313

[Chockler et al., 2004] H. Chockler, J.Y. Halpern (2004): Responsibility and Blame: A Structural-Model Approach, Journal of A.I. Research: 93-115

[Good, 1961/62] I.J. Good (1961) A Causal Calculus I/II. British Journal for the Philosophy of Science 11/12: 305-318/43-51.

[Good, 1994] I.J. Good (1994): Causal Tendency, Necessitivity and Sufficientivity: an updated review. In Patrick Suppes, Scientific Philosopher, P. Humphreys (Ed), Kluwer Dordrecht: Vol. 1 293-315.

[Good, 1999] I.J. Good (1999): Legal responsibility and causation. In Machine Intelligence 15, Koichi Furukawa, Donald Michie and Stephen Muggleton (Eds.), Clarendon Press Oxford: 25-60

[Glymour et al., 2007] C. Glymour and F. Wimberly (2007). Actual Causes and Thought Experiments. In Causation and Explanation, J. K. Campbell, M. O' Rourke, H. Silverstein (Eds.), MIT Press: 43-67.

[Halpern et al., 2005] J.Y. Halpern and J. Pearl (2005). Causes and Explanations: a structural-model approach. Part I: Causes. British Journal for Philosophy of Science 56: 843-887.

[Halpern, 2008] J.Y. Halpern (2008): Defaults and Normality in Causal Structures. In Principles of Knowledge Representation and Reasoning: Proc. Eleventh International Conference: 198-208.

[Halpern et al., 2010] J.Y. Halpern and C. Hitchcock (2010). Actual Causation and the Art of Modeling. In Heuristics, Probability and Causality, A Tribute to Judea Pearl, R. Dechter, H. Geffner, J.Y. Halpern (Eds.), College Publications, 383-406.

[Hitchcock, 2001] C. Hitchcock (2001). The Intransitivity of Causation Revealed in Equations and Graphs. Journal of Philosophy 98 (6): 273-299.

[Hitchock et al., 2009] C. Hitchcock and J. Knobe (2009). Cause and Norm. Journal of Philosophy 106 (11): 587-612.

[Mackie, 1965] JL. Mackie (1965) Causes and conditions. American Philosophical Quarterly.2:245-255.

[Moore, 2009] M. S. Moore (2009). Causation and Responsibility, An Essay in Law, Morals and Metaphysics. Oxford University Press.

[Pearl, 2000] J.Pearl (2000). Causality: Models, Reasoning, and Inference. Cambridge University Press.

[Rothman et al., 2005] (2005) K. J. Rothman, S. Greenland American Journal of Public Health Health, 95: S144-150.

[Schaffer, 2000] J. Schaffer (2000) Causation by Disconnection. Philosophy of Science, Vol. 67,No. 2.: 285-300.