Changes in state *vs* persistent state in continuous time

Aurélien Fermo, ENS - EHESS

February 5, 2020

1 Theory

Last meeting we considered the opportunity for addressing a problem that seems to challenge both the Physical Process (PP) and Counterfactual (CF) account of actual causation. Again let's take the same graph below (Fig.1) as the last time. In this graph the AND-Gate depicted by the acr of circle says that the effect Z has to be activated by both K and J. Let's say (contrarily to the last time) that at the beginning nodes E, G, I and K are already activated, that no node of the chain of circles is activated, and we observe no changes for a while. Let's say that after a certain time node D is activated and then all the descendant nodes up to Z which eventually fires. We put aside for now the distinction between circles and squares − that is between persistent and non-persistent node − which is not relevant for our current purpose.

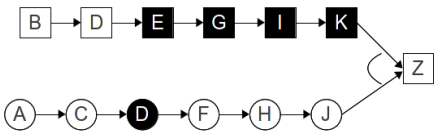

Figure 1

We wage that people would be intuitively more prone to judge D as the cause of the occurrence of Z, precisely because D is the one which initiates a change in the value of Z at a particular time. Yet both the PP and CF account of causation predict that both E and D are equally the causes of Z (see the document *Proposal_29.01*). This example potentially illustrates the need to find a model of actual causation where events are not defined in term of states ($X = x$) but in term of changes of state in continuous time ($x(t) : 0 \rightarrow 1$). In other words, we hypothesize that if through a chains of intermediary changes (from 0 to 1 or 1 to 0, for binary nodes), a change in the value of a primary node brings about the occurrence of the final effect Z (switching the node value from 0 to 1), then the former node will be said the actual cause of the latter one. This hypothesis states in substance that people, relying on a heuristic, find a way of tracing back the history (or path) of inter-connected (or dependant) changes from the final effect up to the actual cause.

However we firstly need to understand what this heuristic consists on. One candidate would be that consisting on picking out the node which has changed the last, among all the other parent-nodes of Z, and tracing back the history of this change. In the previous graph (Fig.1) this method seems intuitive. However this method wouldn't work in many other

cases different from that above where there is more than one initial change. Indeed our model has to deal with these cases to explain not only why people (by supposition) consider the unique changing node (like above) as the actual cause, but also why among different changing nodes they pick out one in particular rather than an other. So let's consider the OR-Gate graph in Fig.2 and say that A and B fire but A few milliseconds before B. If the time delay between the activation of each node is kept fixed, and if the delay between the activation of A and B is such that K eventually fires before the activation of J brings about its effect to Z, then the method doesn't apply : K is the last node whose value has changed but people will probably trace back the history of changes of Z through the path J $\rightarrow$ ... $\rightarrow$ A, not K $\rightarrow$ ... $\rightarrow$ B. Thus an other explanation would be to say that in case of AND-Gate effect, people find the actual cause in the root change (change in a root node) that, among all the other ones, occurred the last; whereas it would be the opposite for OR-Gate effect. But first this heuristic would suffer from lack of unity (especially as there are more than just two types of graph), and second it would be true only in case of fixed and equally represented − among the different paths − time delay between the activation of a cause and the occurrence of the effect. We think that the latter condition is not necessarily met in many realistic scenario. Rather we suggest that the heuristic borrows its main characteristic from the *Intrinsicness thesis* of the Physical Process account of actual causation. More specifically people rely probably on the intrinsic structure of the temporal process of changes along a path. In other words the heuristics tells : *find the path, from the effect Z to its alleged causes, which maximizes the homogeneity of the temporal process of changes.* Indeed for each node activation at time t, we have expectations, based on previous observations, about the time t' at which the child-node has to be activated.

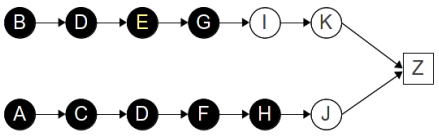

Figure 2

2 The model

2.1 Preliminaries

We can now put forward an idea of model that takes into account this heuristics for identifying some changing nodes as actual causes of an effect. But first we have to make some assumptions and preliminary remarks:

- First of all we will deal with fixed and unique time delay only. However we think it is important to consider a model general enough to account for cases where time delay is fixed but different for each path or not fixed at all. The latter case is interesting in itself because it induces uncertainty about the causes of a common effect. Thus the model presented here will be in its general expression.
- Second we assume for now that there are only two types of node : producing one and preventing one. We will see later how we can introduce the difference between circles (non-persistent nodes) and squares (persistent nodes). But in any case we posit that

both the effect of a producing node and the effect of a preventing one necessarily occur after a certain time. In other words the value of the child-node and the value of the parent-node, whatever it is (producing or preventing), cannot appear simultaneously.

- Third we assume more generally a *no coincidence principle* according to which two changes in the system occur necessarily at a different time[1].

2.2 An algorithm for tracing back the history of changes

[To be formalized. It will have to split a given graph into subgraphs (like in Fig.3) of intermediary histories of changes: for each intermediary change in a common effect of disjoint causes, the algorithm, based on a functional causal model (see below) will have to find its origin (from the left − via Y − or from the right − via X − in Fig.3 for instance).]

2.3 The functional causal model

[To be fully explained.]

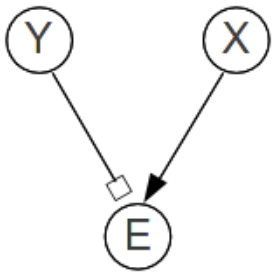

Figure 3

[1]But this principle can lead, in some very specific cases, to quite counter-intuitive judgments. We could add some restrictions to it later yet keep the model as general as possible for now.

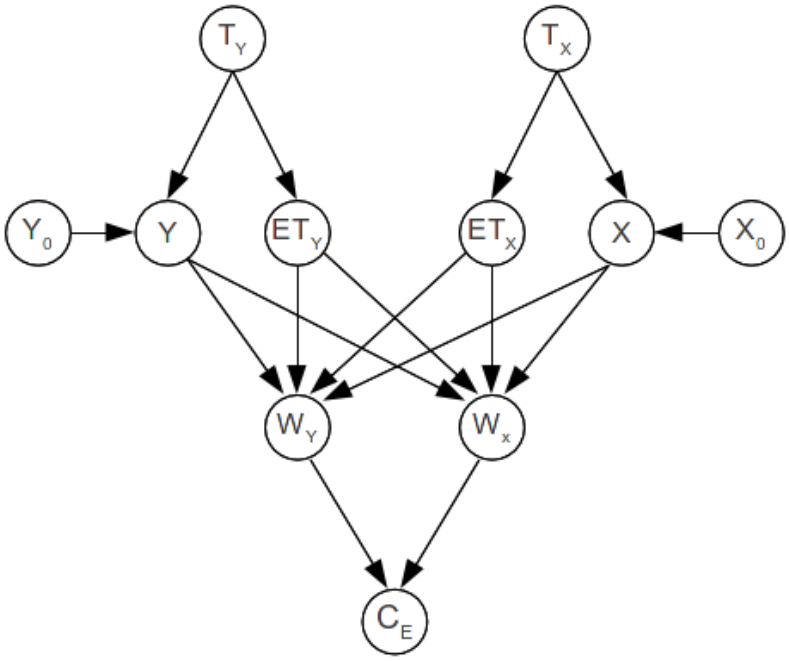

Figure 4

The values for T_Y, T_X, Y_0 and X_0 are already given with :

$$Val(T_Y) = Val(T_X) = \mathbb{R}^+$$
$$Val(Y_0) = Val(X_0) = \{0, 1\}$$

$$y = \begin{cases} y_0 & \text{if} \quad t_y = 0 \\ 1 - y_0 & \text{otherwise} \end{cases}$$

$$x = \begin{cases} x_0 & \text{if} \quad t_x = 0 \\ 1 - x_0 & \text{otherwise} \end{cases}$$

$$et_y = \begin{cases} t_y + \Delta t_y & \text{if} \quad t_y > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$et_x = \begin{cases} t_x + \Delta t_x & \text{if} \quad t_y > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$w_y = \begin{cases} \text{True} & \text{if} \quad x = 1 \wedge y = 0 \wedge et_x = 0 \wedge et_y > 0 \\ \text{True} & \text{if} \quad x = 0 \wedge y = 0 \wedge et_x > et_y > 0 \\ \text{False} & \text{otherwise} \end{cases}$$

$$w_x = \begin{cases} \text{True} & \text{if} \quad x = 0 \wedge y = 0 \wedge et_x > 0 \wedge et_y = 0 \\ \text{True} & \text{if} \quad x = 0 \wedge y = 1 \wedge 0 < et_x < et_y \\ \text{True} & \text{if} \quad x = 1 \wedge y = 0 \wedge et_x > 0 \wedge et_y = 0 \\ \text{True} & \text{if} \quad x = 1 \wedge y = 1 \wedge 0 < et_x < et_y \\ \text{False} & \text{otherwise} \end{cases}$$

$$c_e = w_y \vee w_x$$

2.4 Applying the interventionist account based on each functional causal model

[To be formalized. Once we have our FCM that shouldn't be difficult based on the Pearl method for finding CAUSAL BEAM *[1]]*

References

[1] Pearl J. *Causality: Models, Reasoning and Inference*. Cambridge University Press, USA, 2nd edition, 2009.