Actual causes as changes in continuous time

Aurélien Fermo, ENS Ulm

February 7, 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.

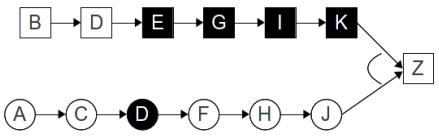

Figure 1

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 ($x(t) : 0 \to 1$). 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.

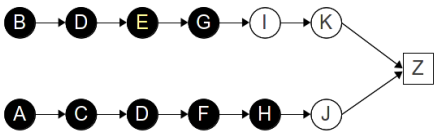

Figure 2

2 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 properly formalized in an algorithmic language using the semantics of graph theory]

Let's take an other example of graph that our model will have to account for as well. Fig.3 corresponds to the state of a graph at the initial time t_0. We observe that node Z is not activated. Then at a certain time t' node N is activated which leads, after some delay, to the final graph in Fig.4 where eventually node Z is activated. If we consider changes as causes, node N is clearly the actual cause of the activation of Z[2].

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 (for instance from the left − via Y − or from the right − via X − in Fig.3).

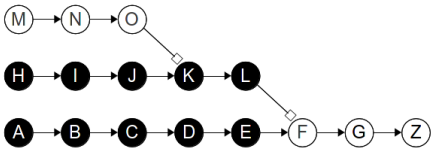

Figure 3

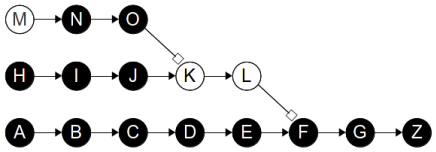

Figure 4

Accordingly our algorithm has to trace back the causal history of Z, that is the history of changes, by identifying for each change :

1. which path the previous change belongs to, if the current change is a common effect of several disjoint nodes (like O−∘K←J);
2. whether it has to stop at the current change or not, if the node considered belongs to a chain of nodes (like M→N→O). Of course it stops if the node is a root node.

This logics (from Z to N) is depicted in Fig.5. The idea is as follows : for each part of the graph (1, 2, 3, 4 and 5 here) the algorithm bases its decision on a FUNCTIONAL CAUSAL MODEL which needs to take into account the dimensionality of time (i.e. to take as one of its input the time at which each change occurred); it retains the result and continue with

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

[2]An other problem is to know whether the causal judgment as function of changes is symmetric, that is whether judgments based on the occurrence of a cause (change from 0 to 1) are substantially different or not from the ones based on the disappearance of a cause (change from 1 to 0). For instance in Fig.4 if node I rather than N is turned off at t', then Z will occur at t' as well. Here node I would be very likely considered the actual cause of Z, yet would this causal judgement be the same as the one based on the occurrence of N instead ?

the previous part and so on. At the end the algorithm returns the path of changes that has led to the occurrence of Z (N→O⊸K→L⊸F→G→Z in our example).

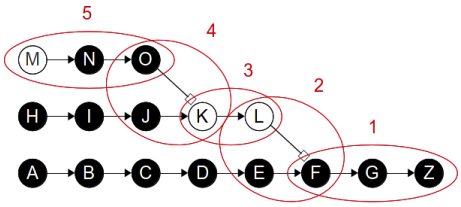

Figure 5

2.3 A functional causal model

[To be fully specified and completely formalized.]

2.3.1 Two parents and one child

Let's begin with point 1. above, that is let's see on which type of functional causal model (FCM) our algorithm will base its decision when it assesses the change that occurred in a common effect[3] of two nodes like in Fig.6 which corresponds to part 4 of the previous graph.

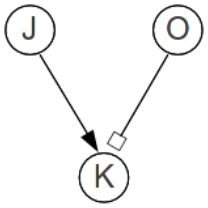

Figure 6

The model could be easily generalized to any number of parents but let focus here to the case above where there are only two parents. Importantly our FCM, regardless of the type of edges between parents and child (producing ones like J→K or preventing ones like O⊸K), will be always associated with a graph structurally identical to the one in Fig.7. Thus it is only the functions associated to that graph which translate the different type of edges, not the structure itself of the graph.

[3]Obviously here the meaning of the term *effect* is broad enough to take into consideration the outcome of a change in the value of either a preventing parent-node or a producing one.

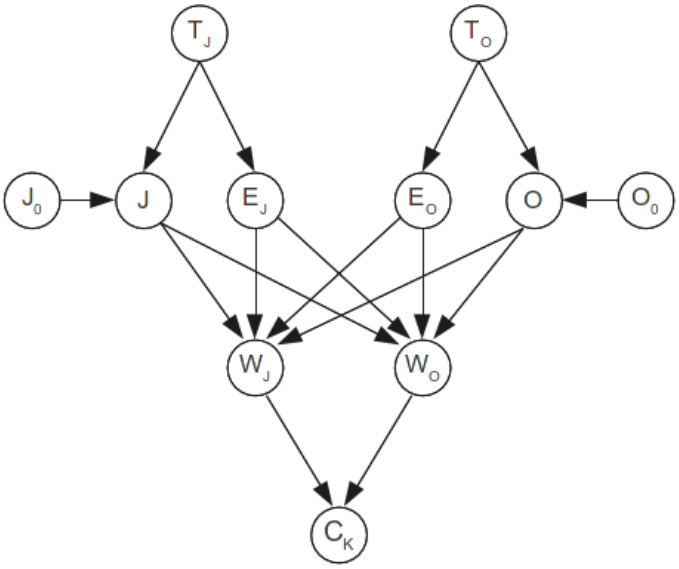

Figure 7

Let's explain the variables involved. J and O are the same nodes as before and take on value 0 if not activated and 1 if activated. C_K is the variable associated with the question *has there been a change in the value of node K ?* : $c_K = 1$ means that a change occurred ($K^0 \to K^1$ or $K^1 \to K^0$). T_J and T_O are the (absolute) time at which the value of J or the value of O or both have been changed, but if these nodes are set to 0 it means that no change has occurred. E_J corresponds to the subsequent time at which we expect that K change if the change in J is the cause thereof. It simply adds Δt_J to T_J when $T_J \neq 0$ and is equal to 0 otherwise. Similarly for E_O. Δt_J and Δt_O (not represented in the graph) correspond to the subjectively expected time delay between a change in a parent node and its effect in a child node. Typically in settings where time delay is not fixed, there is uncertainty around these two variables whose value can be inferred by Bayesian inference. We could perfectly deal with these uncertainty by feeding our causal model with a probability function. E_J and E_O are important because they represent the temporal heuristics we hypothesized above. J_0 and O_0 are the initial value (at t_0) of J and O before any change occurred to these nodes. And finally W_J and W_O answer the question *which way of changes is responsible for the change in K : through J or O ?* – for instance if $W_J = 1$ with probability 1 then the algorithm knows that causality comes from node J.

Moreover we have (for J, O and K binary variables):

$$
\begin{aligned}
&Val(T_J) = Val(T_O) = \mathbb{R}_+ \\
&\Delta t_J,\ \Delta t_O \in \mathbb{R}_+ \\
&Val(E_J) = Val(E_O) = \mathbb{R}_+ \\
&Val(J_0) = Val(O_0) = \{0, 1\} \\
&Val(J) = Val(O) = \{0, 1\} \\
&Val(W_J) = Val(W_O) = \{0, 1\} \\
&Val(C_K) = \{0, 1\}
\end{aligned}
$$

Now we turn more specifically to the functions associated to the graph in Fig.6. We didn't represent on Fig.7 the nodes that correspond to extraneous variables *[to be clearly specified]* but a functional causal model $\mathcal{M}$ over a set of variables $\mathcal{X}$ is a causal model which is always defined over the set $\mathcal{X}$ and a set $\mathcal{U}$ of extraneous variables U^X, with $X \in \mathcal{X}$.

Variable J

Let's say that $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are two tuples of all the variables $X \in \mathrm{Pa}_J$ with their assignments (Pa_J being the set of parents of J − here a singleton). For $\boldsymbol{x}_1 : \{T_J = 0\}$ and $\boldsymbol{x}_2 : \{T_J = i\}$, $i \in \mathbb{R}_+^*$, the assignment $\boldsymbol{u}$ to the set of extraneous variable U^J that affects the outcome of J is such that:

$$\begin{aligned} f_J(\boldsymbol{u}, \boldsymbol{x}_1) &= j_0 \\ f_J(\boldsymbol{u}, \boldsymbol{x}_2) &= 1 - j_0 \end{aligned}$$

Stated differently it simply means that in context $\boldsymbol{u}$:

$$j = \begin{cases} j_0 & \text{if} \quad T_J = 0 \\ 1 - j_0 & \text{otherwise} \end{cases}$$

Variable O

For $\boldsymbol{x}_1 : \{T_O = 0\}$ and $\boldsymbol{x}_2 : \{T_O = i\}$, $i \in \mathbb{R}_+^*$, we have in context $\boldsymbol{u}$:

$$\begin{aligned} f_O(\boldsymbol{u}, \boldsymbol{x}_1) &= o_0 \\ f_O(\boldsymbol{u}, \boldsymbol{x}_2) &= 1 - o_0 \end{aligned}$$

Variable E_J

For $\boldsymbol{x}_1 : \{T_J = i\}$ and $\boldsymbol{x}_2 : \{T_J = 0\}$, $i \in \mathbb{R}_+^*$, we have in context $\boldsymbol{u}$:

$$\begin{aligned} f_{E_J}(\boldsymbol{u}, \boldsymbol{x}_1) &= t_J + \Delta t_J \\ f_{E_J}(\boldsymbol{u}, \boldsymbol{x}_2) &= 0 \end{aligned}$$

Variable E_O

For $\boldsymbol{x}_1 : \{T_O = i\}$ and $\boldsymbol{x}_2 : \{T_O = 0\}$, $i \in \mathbb{R}_+^*$, we have in context $\boldsymbol{u}$:

$$\begin{aligned} f_{E_O}(\boldsymbol{u}, \boldsymbol{x}_1) &= t_O + \Delta t_O \\ f_{E_O}(\boldsymbol{u}, \boldsymbol{x}_2) &= 0 \end{aligned}$$

Variable W_J

For $\boldsymbol{x}_1 : \{O = 1, J = 0, E_O = 0, E_J = i\}$, $\boldsymbol{x}_2 : \{O = 0, J = 0, E_O = k, E_J = l\}$, $i,\ l,\ k \in \mathbb{R}_+^*$ with $l > k$, and $\bar{\boldsymbol{x}}_{-1,-2}$ the tuple of all assignments $\boldsymbol{x}_n$ except for $n = 1$ and $n = 2$, we have in context $\boldsymbol{u}$:

$$\begin{aligned} f_{W_J}(\boldsymbol{u}, \boldsymbol{x}_1) &= True \\ f_{W_J}(\boldsymbol{u}, \boldsymbol{x}_2) &= True \\ f_{W_J}(\boldsymbol{u}, \bar{\boldsymbol{x}}_{-1,-2}) &= False \end{aligned}$$

In other words it means that in context $\boldsymbol{u}$:

$$w_J = \begin{cases} \text{True} & \text{if} \quad O = 1 \wedge J = 0 \wedge E_O = 0 \wedge E_J > 0 \\ \text{True} & \text{if} \quad O = 0 \wedge J = 0 \wedge E_O > E_J > 0 \\ \text{False} & \text{otherwise} \end{cases}$$

Variable W_O

For $\boldsymbol{x}_1 : \{O = 0, J = 0, E_J = 0, E_O = i\}$, $\boldsymbol{x}_2 : \{O = 1, J = 0, E_J = 0, E_O = k\}$, $\boldsymbol{x}_3 : \{O = 0, J = 1, E_J = l, E_O = m\}$, $\boldsymbol{x}_4 : \{O = 1, J = 1, E_J = n, E_O = p\}$, $i,\ k,\ l,\ m,\ n,\ p \in \mathbb{R}_+^*$ with $l > m$, and $n > p$, and for $\bar{\boldsymbol{x}}_{...,-4}$ the tuple of all assignments $\boldsymbol{x}_n$ except for $n = 1$, $n = 2$, $n = 3$ and $n = 4$, we have in context $\boldsymbol{u}$:

$$\begin{aligned} f_{W_O}(\boldsymbol{u}, \boldsymbol{x}_1) &= True \\ f_{W_O}(\boldsymbol{u}, \boldsymbol{x}_2) &= True \\ f_{W_O}(\boldsymbol{u}, \boldsymbol{x}_3) &= True \\ f_{W_O}(\boldsymbol{u}, \boldsymbol{x}_4) &= True \\ f_{W_O}(\boldsymbol{u}, \bar{\boldsymbol{x}}_{...,-4}) &= False \end{aligned}$$

In other words it means that in context $\boldsymbol{u}$:

$$W_O = \begin{cases} \text{True} & \text{if} \quad O = 0 \wedge J = 0 \wedge E_J = 0 \wedge E_O > 0 \\ \text{True} & \text{if} \quad O = 1 \wedge J = 0 \wedge E_J = 0 \wedge E_O > 0 \\ \text{True} & \text{if} \quad O = 0 \wedge J = 1 \wedge E_J > e_O > 0 \\ \text{True} & \text{if} \quad O = 1 \wedge J = 1 \wedge E_J > e_O > 0 \\ \text{False} & \text{otherwise} \end{cases}$$

Variable C_K

$$c_K = w_J \vee w_O$$

2.3.2 In a chain of nodes

[To be fully specified and completely formalized.]

Now let's look at a simple chain of nodes like parts 1, 3 and 5 in Fig.5. The functional causal model in this case will be based always in a graph identical to the one in Fig.8, which is very similar to the one in Fig.7 except that it is truncated : the left part has been removed. W_0 and W_N answers the question *is the change in O own to a change in the previous node N or to a change by itself ?*.

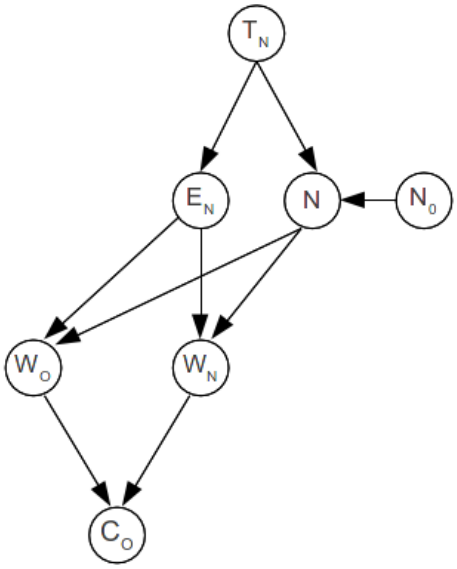

Figure 8

2.4 Applying the interventionist account based on each functional causal model

[To be formalized. Having a FCM, we shouldn't face here a big challenge if we take cue from the Pearl's method for finding CAUSAL BEAM *[1]]*

References

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