Proposed graphs

Aurlien Fermo, ENS - EHESS

January 8, 2020

1 Introduction

In a system where an effect is due to one or several causes, the goal is to understand :

1. When and why people truly rely on the heuristics according to which the actual cause is the starting node of a chain (path) of events that directly lead to the activation of the last node (the effect); and when they rule out the possibility to consider other starting nodes as (counterfactual or not) causes;

2. The cases (if any) where people keep continuing to rely on this heuristics and consider as a cause the starting node of a chain which is not however the actual cause of the end node;

3. And finally the exact opposite behaviour, that is the tendency (to be shown) to consider as a true cause a node which is not the starting one in a path of events (and then which is not the actual one).

2 First node as cause

There is probably a quite strong heuristics which leads people to consider the starting node of a path of events (to an effect) as the cause of the effect. In Fig. 1 for example we represented a system where an effect is produced by two chains of events with different length, and where the actual causes are clearly the two starting nodes of the different chains. We see that at $t-1$ (i.e. just before the effect occurs), both paths are completely active. So in this case the heuristics leads to a consistent response.

But now we represented in Fig. 2-5 cases where at $t-1$ the shortest path (B) is totally active before the longest one (A). So the actual cause in these cases, is always the staring node of the shortest path (B). However, according to me, it is probable that particularly in Fig. 2 and Fig. 3 people keep continuing to rely on the heuristics and consider the starting node of the longest path (A) as a cause, even if at $t-1$ the path is not entirely activated.

I think there are at least two correlated factors that might contribute to that : the difference between the two paths at $t-1$ (here the number of white node for A), and the notion of time and simultaneity. About the first factor, in Fig. 2 the last node in A is activated just one time step after the last one in B that might lead people to consider that A contributes in a sense to support a little bit later the effect of B. About the second factor, in Fig. 2 A is activated the first, and this precedence might support the tendency to consider the starting node of A as a part cause of the system. In Fig. 3 this is maybe the simultaneity between the activation of the two starting nodes which might lead people to consider the starting node of A as also causally involved.

I think that we can find other cases (graphs) where people continue to rely on the heuristics even if it clearly implies to consider take into account causes that are not the actual ones.

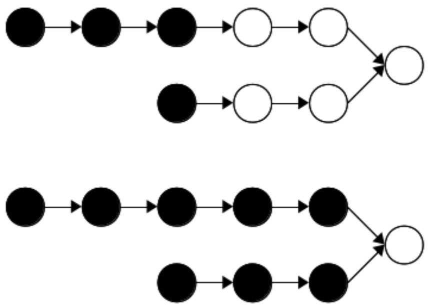

Figure 1

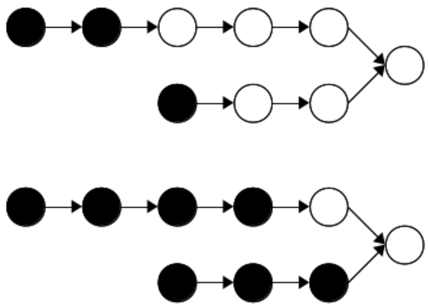

Figure 2

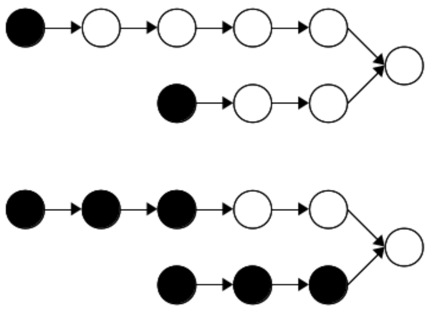

Figure 3

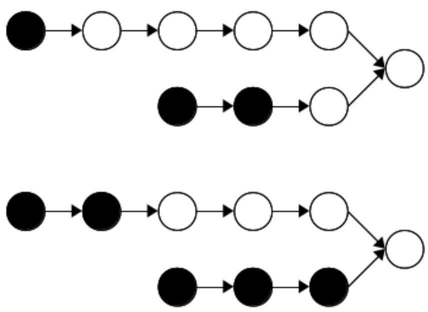

Figure 4

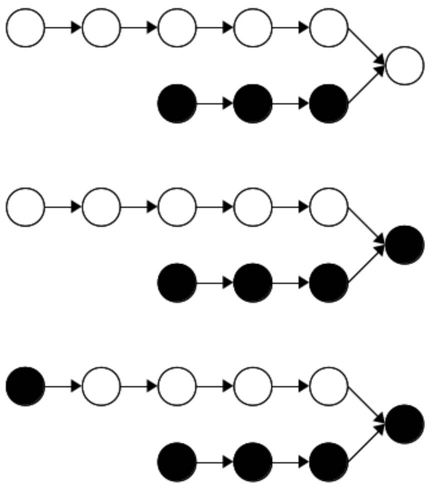

Figure 5

3 Cause and contrast

In some cases, I think that we can observe the exact opposite behaviour, that is the tendency to find a cause where it is actually not the beginning of completely activated path. The Fig. 6 is a proposal of such cases (at $t-1$) where people could potentially depart from the described heuristics and don't conceive of the first node as the actual cause. The red square indicates which node could be more strongly recognized as the direct cause.

I propose the following explanation : the cause that people judged to be the direct one (in red square), which can be then different from the actual one (the starting node), is the one which creates the most important ***contrast***. On the top of the Fig. 6 the node in the red square contrasts with all of the preceding nodes being the node of a shortcut branch. In the middle now, the lonely node is the one which maybe creates the strongest contrast: it contrasts with all of the inactivated nodes which, for the effect to occur, should have been activated had the lonely node been inactivated. On bottom, it's maybe more ambiguous because the lonely node creates in a certain sense maybe a less strong contrast : indeed each of the inactivated nodes would have been likely to activate the lonely node, and then raised the probability for the effect to occur even if the node of the first short-cut branch hadn't activated the lonely node. On the other hand the lonely node highly contrasts in the sense that it is the one which almost all the edges converges on.

The notion of contrast would allow us to interpret the remaining graphs (Fig. 7 and 8). But I think that it goes together with the notions of *root sparsity* and *path length*. The contrast is maximal when either :

I. A node is a root of several different branches which lead to the same effect

Or:

II.a. The number of inactivated nodes in parallel branches is maximal;

II.b. The number of shortcut branches is minimal;

II.c. The length of shortcut branches is minimal.

According to me conditions IIa-c are precisely met in the bottom graph of Fig. 7 and top graph of Fig. 8. In the other cases it's the condition I which is met. However in the top graph of Fig. 7 it is maybe more ambiguous (maybe the number of different branches that come from the same root has to be taken into account).

The next step would be to see how we can model this notion of contrast by a functional causal model.

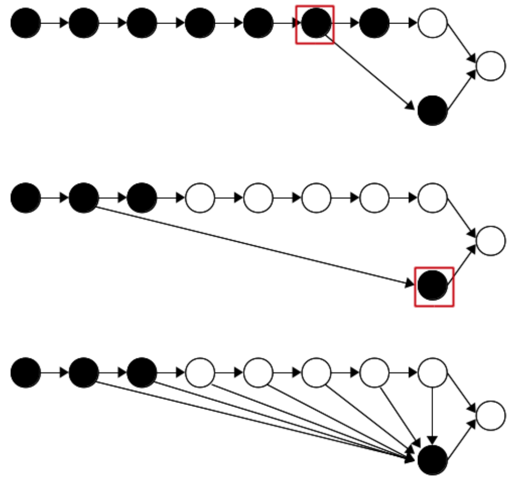

Figure 6

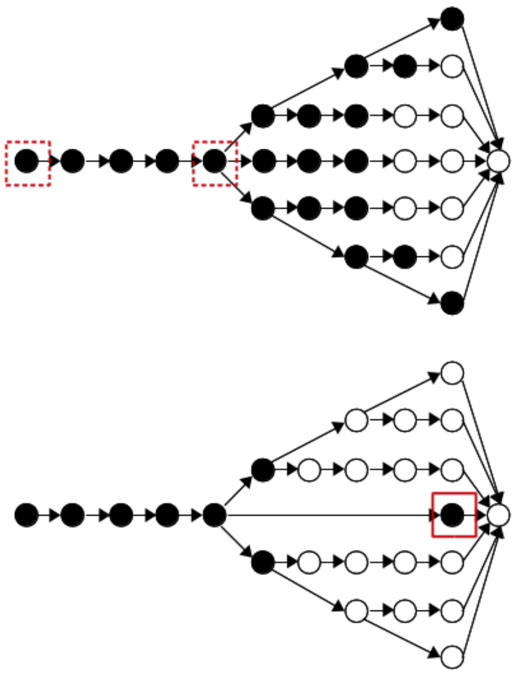

Figure 7

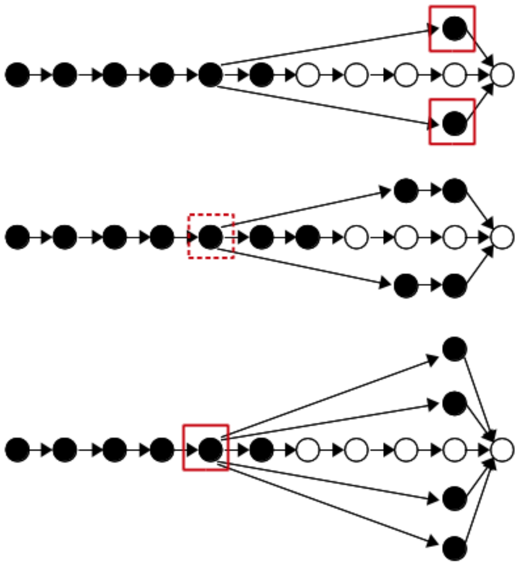

Figure 8